Federal Reserve Bank of New York

September 2008
Volume 14 Number 2

Economic Policy Review

Special Issue: The Economics of Payments

The *Economic Policy Review* is published by the Research and Statistics Group of the Federal Reserve Bank of New York. Articles undergo a comprehensive refereeing process prior to their acceptance in the *Review*. The views expressed are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

www.newyorkfed.org/research

Federal Reserve Bank of New York Economic Policy Review

September 2008
Volume 14 Number 2

Special Issue: The Economics of Payments

Contents

3 Introduction
James McAndrews

Theoretical Models of Money and Payments
Articles:

7 Intraday Liquidity Management: A Tale of Games Banks Play
Morten L. Bech

Over the last few decades, most central banks, concerned about settlement risks inherent in payment netting systems, have implemented real-time gross settlement (RTGS) systems. Although RTGS systems can significantly reduce settlement risk, they require greater liquidity to smooth nonsynchronized payment flows. Thus, central banks typically provide intraday credit to member banks, either as collateralized credit or priced credit. Because intraday credit is costly for banks, how intraday liquidity is managed has become a competitive parameter in commercial banking and a policy concern of central banks. This article uses a game-theoretical framework to analyze the intraday liquidity management behavior of banks in an RTGS setting. The games played by banks depend on the intraday credit policy of the central bank and encompass two well-known paradigms in game theory: "the prisoner's dilemma" and "the stag hunt." The former strategy arises in a collateralized credit regime, where banks have an incentive to delay payments if intraday credit is expensive, an outcome that is socially inefficient. The latter strategy occurs in a priced credit regime, where postponement of payments can be socially efficient under certain circumstances. The author also discusses how several extensions of the framework affect the results, such as settlement risk, incomplete information, heterogeneity, and repeated play.

25 An Economic Analysis of Liquidity-Saving Mechanisms

Antoine Martin and James McAndrews

A recent innovation in large-value payments systems has been the design and implementation of liquidity-saving mechanisms (LSMs), tools used in conjunction with real-time gross settlement (RTGS) systems. LSMs give system participants, such as banks, an option not offered by RTGS alone: they can queue their outgoing payments. Queued payments are released if some prespecified event occurs. LSMs can reduce the amount of central bank balances necessary to operate a payments system as well as quicken settlement. This article analyzes the performance of RTGS systems with and without the addition of an LSM. The authors find that, in terms of settling payments early, these mechanisms typically outperform pure RTGS systems. However, there are times when RTGS systems can be preferable to LSMs, such as when many banks that send payments early in RTGS choose to queue their payments when an LSM is available. The authors also show that the design of a liquidity-saving mechanism has important implications for the welfare of system participants, even in the absence of payment netting. In particular, the parameters specified determine whether the addition of an LSM increases or decreases welfare.

41 Divorcing Money from Monetary Policy

Todd Keister, Antoine Martin, and James McAndrews

Many central banks implement monetary policy in a way that maintains a tight link between the stock of money and the short-term interest rate. In particular, their implementation procedures require that the supply of reserve balances be set precisely in order to implement the target interest rate. Because bank reserves play other key roles in the economy, this link can create tensions with other important objectives, especially in times of acute market stress. This article considers an alternative approach to monetary policy implementation—known as a “floor system”—that can reduce or even eliminate these tensions. The authors explain how this approach, in which the central bank pays interest on reserves at the target interest rate, “divorces” the supply of money from the conduct of monetary policy. The quantity of bank reserves can then be set according to the payment or liquidity needs of financial markets. By removing the opportunity cost of holding reserves, the floor system also encourages the efficient allocation of resources in the economy.

Empirical Analyses of Trends in Large-Value Payments
Articles:

59 Global Trends in Large-Value Payments
Morten L. Bech, Christine Preisig, and Kimmo Soramäki

Globalization and technological innovation are two major forces affecting the financial system and its infrastructure. Perhaps nowhere are these trends more apparent than in the internationalization and automation of payments. While the effects of globalization and technological innovation are most obvious on retail payments, the influence is equally impressive on wholesale, or interbank, payments. Given the importance of payments and settlement systems to the smooth operation and resiliency of the financial system, it is important to understand the potential consequences of these developments. This article presents ten major long-range trends in the settlement of large-value payments worldwide. The trends are driven by technological innovation, structural changes in banking, and the evolution of central bank policies. The authors observe that banks, to balance risks and costs more effectively, are increasingly making large-value payments in real-time systems with advanced liquidity-management and liquidity-saving mechanisms. Moreover, banks are settling a larger number of foreign currencies directly in their home country by using offshore systems and settling a greater number of foreign exchange transactions in Continuous Linked Settlement Bank or through payment-versus-payment mechanisms in other systems. The study also shows that the service level of systems is improving, through enhancements such as longer operating hours and standardized risk management practices that adhere to common standards, while transaction fees are decreasing. Payments settled in large-value payments systems are more numerous, but on average of smaller value. Furthermore, the overall nominal total value of large-value payments is increasing, although the real value is declining.

83 Changes in the Timing Distribution of Fedwire Funds Transfers
Olivier Armantier, Jeffrey Arnold, and James McAndrews

The Federal Reserve's Fedwire funds transfer service—the biggest large-value payments system in the United States—has long displayed a peak of activity in the late afternoon. Theory suggests that the concentration of late-afternoon Fedwire activity reflects coordination among participating banks to reduce liquidity costs, delay costs, and credit risk; as these costs and risk change over time, payment timing most likely will be affected. This article seeks to quantify how the changing environment in which Fedwire operates has affected the timing of payment value transferred within the system between 1998 and 2006. It finds that the peak of the timing distribution has become more concentrated, has shifted to later in the day, and has actually divided into two peaks. The authors suggest that these trends can be explained by a rise in the value of payments transferred over Fedwire, the settlement patterns of the private settlement institutions that use the system, and an increase in industry concentration. Although the study's results provide no specific evidence of heightened operational risk attributable to activity occurring later in the day, they point to a high level of interaction between Fedwire and private settlement institutions.

113 The Timing and Funding of CHAPS Sterling Payments
Christopher Becher, Marco Galbiati, and Merxe Tudela

Real-time gross settlement (RTGS) systems such as CHAPS Sterling require large amounts of liquidity to support payment activity. To meet their liquidity needs, RTGS participants borrow from the central bank or rely on incoming payments from other participants. Both options can prove costly—the latter in particular if participants delay outgoing payments until incoming ones arrive. This article presents an empirical analysis of the timing and funding of payments in CHAPS. The authors seek to identify the factors driving the intraday profile of payment activity and the extent to which incoming funds are used as a funding source, a process known as liquidity recycling. They show that the level of liquidity recycling in CHAPS is high and stable throughout the day, and attribute this result to several features of the system. First, the settlement of time-critical payments provides liquidity to the system early in the settlement day; this liquidity can be recycled for the funding of less urgent payments. Second, CHAPS throughput guidelines provide a centralized coordination mechanism, in effect limiting any tendency toward payment delay. Third, the relatively small direct membership of CHAPS facilitates coordination between members, for example, through the use of bilateral net sender limits. Coordination encourages banks to maintain a relatively constant flux of payments throughout the day. The authors also argue that the high level of recycling helps to reduce liquidity risk, and that the relatively smooth intraday distribution of payments serves to mitigate operational risk associated with highly concentrated payment activity. They note, however, that the benefits of liquidity recycling are not evenly distributed between members of CHAPS.

Risk Management in Payments Systems
Articles:

137 Understanding Risk Management in Emerging Retail Payments
Michele Braun, James McAndrews, William Roberds, and Richard Sullivan

New technologies used in payment methods can reduce risk, but they can also lead to new risks. Emerging retail payments are prone to operational and fraud risks, especially security breaches and potential use in illicit transactions. This article describes an economic framework for understanding risk control in retail payments. Risk control is a special type of good because it can protect one payment participant without diminishing the protection of other participants. As a result, the authors' economic framework emphasizes risk containment, primarily through the establishment and enforcement of risk management policies. Application of the framework to three types of emerging payments suggests that a payments system can successfully manage risk if it quickly recognizes problems, encourages commitment from all participants to control risk, and uses an appropriate mix of market and public policy mechanisms to align risk management incentives. The authors conclude that providers of emerging payment methods must mitigate risk effectively or face rejection in the payment market.

161 An Economic Perspective on the Enforcement of Credit Arrangements: The Case of Daylight Overdrafts in Fedwire

Antoine Martin and David C. Mills

A fundamental concern for any lender is credit risk—the risk that a borrower will fail to fully repay a loan as expected. Thus, lenders want credit arrangements that are designed to compensate them for—and help them effectively manage—this type of risk. In certain situations, central banks engage in credit arrangements as lenders to banks, so they must manage their exposure to credit risk. This article discusses how the Federal Reserve manages its credit risk exposure associated with daylight overdrafts. The authors first present a simple economic framework for thinking about the causes of credit risk and the possible tools that lenders have to help them manage it. They then apply this framework to the Federal Reserve's Payments System Risk policy, which specifies the use of a variety of tools to manage credit risk. The study also analyzes a possible increase in the use of collateral as a credit risk management tool, as presented in a recent proposal by the Federal Reserve concerning changes to the Payments System Risk policy.

Introduction

James McAndrews

James McAndrews

Introduction

Simply put, payments settle things. A payment is typically the final step in a sequence of activities that make up an economic or financial trade. Because the parties to the trade look forward to this final step, payments are the focus of their expectations about timing, risks, and costs. The more certain it is that this final step of a trade will take place, the easier it is for trades to occur and for people to undertake efficient economic actions.

As the world economy continues to grow, the share devoted to financial services has risen relative to the share representing other economic activities. Payments have increased in importance as that trend has manifested itself, and their contribution to the global economy is likely to increase as well.

The rising importance of payments to economic activity in general is a significant development—especially for banks and central banks as providers of payment services. Bank customers frequently rely on credit provided by their institutions to complete payments, whether they are using credit cards or other payment instruments. Similarly, banks often rely on very-short-term credit provided by central banks to make payments. Needless to say, whenever financial institutions provide credit, they must manage risk to prevent excessive risk taking. Managing payments is therefore part of a larger risk management process in the financial sector.

This special issue of the *Economic Policy Review* is devoted to the economics of payments. The wide-ranging articles in this collection investigate large-value payments systems, both theoretically and empirically; risk in retail payments systems; payments system development trends across countries; and the interaction of the provision of reserves by central banks and the operation of payments systems. They illustrate the diversity of interests and methods that economists have developed for studying payment activities.

The contributions to this volume center on three broad themes: theoretical models of money and payments, empirical analyses of trends in large-value payments, and risk management in payments systems.

The theoretical theme is examined in three articles—by Morten L. Bech; Antoine Martin and James McAndrews; and Todd Keister, Antoine Martin, and James McAndrews. The Bech study and the Martin-McAndrews study, both focusing on large-value payments systems, explore the strategic incentives for banks to submit payments on time in different economic environments. They consider how the incentives are affected by different central bank policies, such as the terms under which intraday credit is provided. Keister, Martin, and McAndrews consider alternative models of monetary policy implementation as well as the relationship between the demand for intraday balances to meet payment needs and the reserve balances used to implement the policy objectives of the monetary authorities.

Bringing an empirical perspective to the topic, Morten L. Bech, Christine Preisig, and Kimmo Soramäki conduct a global tour of developments in large-value payments over the last

James McAndrews is a senior vice president at the Federal Reserve Bank of New York.
<james.mcandrews@ny.frb.org>

The views expressed are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

decade. Relying largely on data compiled by the Bank for International Settlements, the authors discern a number of key trends in the growth and evolution of large-value payments. In addition, two articles focusing on the timing of large-value payments systems—the first by Olivier Armantier, Jeffrey Arnold, and James McAndrews and the second by Christopher Becher, Marco Galbiati, and Merxe Tudela—complement Bech's theoretical work by explaining how different central bank policies influence payment timing. Both empirical analyses use data obtained directly from the large-value payments systems they study. Armantier, Arnold, and McAndrews review the timing distribution of payments in Fedwire, the Federal Reserve's system, while Becher, Galbiati, and Tudela consider the timing of payments in CHAPS, the major U.K. system. The differences in timing between the two systems are found to be significant.

Two studies consider the theme of risk management—one by Michele Braun, James McAndrews, William Roberds, and Richard Sullivan, the other by Antoine Martin and David C. Mills. Both apply the economic reasoning associated with risks in credit arrangements to the specific case of payments. Braun et al. emphasize emerging retail systems while Martin and Mills focus on the risk associated with a central bank's intraday lending to banks.

The economics of payments is a rapidly developing field. These studies offer a variety of approaches that use theoretical as well as empirical techniques to explore different aspects of payments. Going forward, as researchers gain greater access to payment data, they will be better equipped to test other hypotheses about the determinants of behavior in payments systems. We hope that our findings will stimulate such initiatives and deepen interest in this dynamic field.

Theoretical Models of Money and Payments

Articles by:

Morten L. Bech

Antoine Martin and James McAndrews

Todd Keister, Antoine Martin, and James McAndrews

Morten L. Bech

Intraday Liquidity Management: A Tale of Games Banks Play

- To ensure smooth operation of real-time gross settlement systems, central banks extend intraday credit, either against collateral or for a fee.
- As intraday credit is costly—either explicitly as fees or implicitly as the opportunity cost of collateral—participating banks seek to minimize their use of liquidity by timing the release of payments.
- A game-theoretical study of intraday liquidity management behavior shows how the strategic incentives of banks depend on the intraday credit policy of central banks.
- Two classic games emerge: "the prisoner's dilemma" and "the stag hunt."
- The prisoner's dilemma arises in a collateralized credit scenario, where banks delay payments even though they would be better off if they all sent payments early; the stag hunt arises in a priced credit scenario, where banks seek to coordinate the timing of their payments to avoid overdraft fees.

1. Introduction

> *"[Banks] like to hang on to their cash and deliver it as late as possible at the end of the working day."*
> "The Long Shadow of Herstatt," *The Economist,* April 14, 2001

The value and volume of interbank payments increased dramatically throughout the 1980s and 1990s as a result of rapid financial innovation and the integration and globalization of financial markets. In the United States, settlement of interbank payments grew from $300 trillion in 1985, or forty-five times GDP, to almost $500 trillion in 1995, or seventy-five times GDP (Bech, Preisig, and Soramäki 2008).

Historically, interbank payments have been settled via deferred (end-of-day) netting systems. As the volume and value of transactions increased, however, central banks became worried about settlement risks inherent in netting systems. In particular, the banks were concerned about the potential for contagion, or "knock-on," effects attributable to the unwinding of net positions that would result if a participant failed to meet its end-of-day obligations. Consequently, over the last couple of decades, many countries have chosen to modify the settlement procedures employed by their interbank payments system.

Most central banks opted for the implementation of a real-time gross settlement (RTGS) system. By 1985, three central

Morten L. Bech is a senior economist at the Federal Reserve Bank of New York.
<morten.bech@ny.frb.org>

banks had implemented RTGS systems. A decade later, that number had increased to sixteen, and at the end of 2006 the use of RTGS systems had diffused to ninety-three central banks (Bech and Hobijn 2007).

RTGS systems eliminate the settlement risk from unwinding because payments are settled irrevocably, and with finality, on an individual gross basis and in real time. However, the elimination of settlement risk comes at the cost of an increased need for liquidity to smooth nonsynchronized payment flows.[1] Thus, central banks typically provide intraday credit.

Two types of intraday credit policies have emerged among central banks: *collateralized* credit and *priced* credit (Furfine and Stehm 1998). Collateralized credit, in one form or another, is the prevailing option in Europe and elsewhere outside the United States. Collateralized credit usually takes the form of pledging collateral to the central bank or entering into an intraday repurchase agreement with the central bank. Priced credit is the policy of choice in the United States. The Federal Reserve has been charging a fee for intraday overdrafts since 1994. Quantitative limits, or "caps," are used in combination with both types of credit extensions.

This article develops a stylized game-theoretical model to analyze banks' intraday liquidity management behavior in an RTGS [real-time gross settlement] environment.

Intraday credit is costly, whether explicitly in the form of a fee or implicitly as the opportunity cost of the pledged collateral. Consequently, banks try to economize on their use of liquidity throughout the day by carefully scheduling the settlement of payment requests received from customers and the banks' own proprietary operations. Intraday liquidity management has become an important competitive parameter in commercial banking and a policy concern of central banks (see, for example, Greenspan [1996] and Berger, Hancock, and Marquardt [1996]).

This article develops a stylized game-theoretical model to analyze banks' intraday liquidity management behavior in an RTGS environment. It analyzes the strategic incentives under different intraday credit policy regimes employed by central banks. We characterize how the Nash equilibria depend on the underlying cost parameters, and discuss the efficiency implications of the different outcomes. As it turns out, two classic paradigms in game theory emerge from the analysis: the "prisoner's dilemma" and the "stag hunt." Hence, many policy questions can be understood in terms of well-known conflicts and dilemmas in economics. This study uses the framework to conduct a comparative analysis of the relative desirability of different intraday credit regimes from the perspective of a benevolent central bank. In addition, it discusses in turn how several extensions of the model will affect the results. These extensions include settlement risk, incomplete information, heterogeneity, repeated play, multitudes of players, and more than just two actions. We conclude with general observations on the future of intraday liquidity management.

2. Intraday Liquidity Management Game

Envision an economy with two identical banks using an RTGS system operated by the central bank to settle interbank claims.[2] Bank A and Bank B seek to minimize the cost of making their payments. We look at one business day that consists of two periods: *morning* and *afternoon*.

At dawn, both banks receive a request from a customer to pay $1 to a customer of the other bank on the same business day.[3] Assume for simplicity that the banks can either process the request right away, or postpone it until the afternoon period. We abstract from reserve requirements and precautionary motives for banks to hold balances with the central bank, and thus each bank has a zero balance on its settlement account at dawn. Banks cannot send payments from their accounts in amounts that exceed their account balances. However, banks can borrow funds from the central bank. The cost of borrowing and how it is assessed depend on the intraday credit policy of the central bank. Here, overdrafts are assessed at noon and at the end of the day. Overnight overdrafts are penalized at a very high rate, making banks avoid them altogether.

If there were no adverse consequences, each bank would prefer to postpone making its payment and use the funds received via incoming payments from the other bank to provide the balances to cover its own outgoing payments.

[1] In most RTGS systems, payments are settled using reserve account balances at the central bank. For an individual bank, there are basically four different sources of liquidity to fund outgoing payments: 1) overnight reserve balances, 2) intraday credit extensions by the central bank, 3) borrowing from other banks via the interbank money market, and 4) incoming payments from other banks. The first two sources affect the aggregate level of liquidity available in the system, while the latter two redistribute the liquidity among banks. Moreover, liquidity from the first three sources generally comes at a price, whereas liquidity from incoming payments is free from the perspective of the receiver.

[2] In many countries, the interbank payments systems are neither owned nor operated by the central bank, but rather by a private company or a consortium of banks. However, payments are usually settled in liabilities on the central bank. For ease of exposition, we ignore these differences here.

[3] The customer could be internal to the bank, in which case the decision-making agent can be thought of as the payment manager of the bank.

Game 1
Intraday Liquidity Management Game

		Bank B	
		morning	*afternoon*
Bank A	*morning*	$c^A(m,m)$, $c^B(m,m)$	$c^A(m,a)$, $c^B(a,m)$
	afternoon	$c^A(a,m)$, $c^B(m,a)$	$c^A(a,a)$, $c^B(a,a)$

However, postponing is also costly, as customers might either demand compensation for late settlement or take their business elsewhere in the future, thereby imposing a reputation cost on the delaying bank.

For many payments, the cost of intraday delay is presumably small, as the underlying contractual obligation of the customer only specifies payment on a given business day. However, for an increasing number of financial transactions, the underlying contract stipulates payment prior to some specific time on a given business day, and the cost of delay could conceivably be high. Here, we simply assume the cost of delay to be a positive number *D* per dollar per period within.[4] Moreover, postponing payments until the next day is extremely expensive in terms of reputation effects or direct compensation to customers, so banks always submit any remaining payments in the afternoon.

A convenient way of arranging the possible actions of the banks and the associated costs is a 2 x 2 game, as shown in Game 1. Each bank can play one of two strategies: morning or afternoon. The first element in each cell denotes the settlement cost of Bank A, whereas the second element denotes that of Bank B. Following Bech and Garratt (2003), we label this game *the intraday liquidity management game*.

In the next three sections, we explore the games that emerge under different intraday credit regimes. Our solution concept is Nash equilibrium—that is, a set of strategies for which neither bank would wish to change its strategy on the assumption that the other bank will not change its strategy either. We focus on Nash equilibria in pure strategies—that is, strategies where a player chooses to take one action with probability 1—which is in contrast to a mixed strategy, where individual players choose a probability distribution over several actions. We evaluate the efficiency of different outcomes by comparing both individual and aggregate settlement costs. The regimes are free, collateralized, or priced intraday credit.

[4] Delay has several private and social costs associated with it. First, time is money (even intraday) and hence delay of settlement may displease customers and counterparties, which are left with higher costs and greater uncertainty. Second, delayed settlement increases operational risk insofar as the time span during which an incident may disrupt the settlement process increases and the time to recover after an incident decreases. Third, the process of delaying can be costly, and the resources devoted to managing intraday positions are a cost. Fourth, delay increases the length of time participants may be faced with credit risk exposures vis-à-vis each other.

Game 2
Free Intraday Credit Game

		Bank B	
		morning	*afternoon*
Bank A	*morning*	$\underline{0}$, $\underline{0}$	$\underline{0}$, *D*
	afternoon	*D*, $\underline{0}$	*D*, *D*

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

3. Free Intraday Credit Regime

The first adopters of RTGS systems provided intraday credit for free, and we use this intraday credit policy regime as a benchmark. With free credit within the day, there is no incentive to postpone payments. The free intraday credit game is shown in Game 2.

It is best for both banks to play the morning strategy because they incur no costs. Conversely, they incur the cost of delay if they postpone to the afternoon. The morning strategy dominates the afternoon strategy, and the strategy profile (*morning, morning*) is said to be an equilibrium in dominating strategies. A pair of dominating strategies is a unique Nash equilibrium.

In Game 2 and in other games, we adopt the convention of underlining the cost associated with the strategy that is the best response of a bank given the strategy played by the other. We summarize the results of the free intraday credit game in Result 1:

> *Result 1: Early settlement (*morning, morning*) is a unique equilibrium in the free intraday credit game. The outcome is efficient in that it ensures the lowest possible aggregate settlement cost across all pairs of strategies.*

In reality, payment flows are not perfectly symmetric as they are in the model, and imbalances frequently occur. Moreover, the zero price for intraday credit creates no incentives to economize on overdrafts.[5] In fact, the size of the overdrafts generated by banks (relative to their capital base) in an RTGS environment came as a surprise to many central banks. As guarantor of the finality of payments, the central bank is exposed to credit risk—as, ultimately, are taxpayers. Hence, central banks are almost unanimous in the opinion that the provision of free intraday liquidity is not a viable option.[6]

4. Collateralized Intraday Credit Regime

In most countries, central banks provide commercial banks with intraday credit against collateral. The practical implementation varies across countries, depending on the institutional infrastructure for the safekeeping and settlement of securities. For ease of exposition, we assume that credit is extended via intraday repurchase agreements (repos), as in the United Kingdom and Switzerland.

Under the intraday repo agreement, the central bank provides the bank with $1 in its account at the beginning of the period in return for eligible collateral worth the same amount plus a "haircut" to cover any market and credit risk associated with the collateral. At the end of the period, the transaction is reversed. The central bank does not charge explicit interest for this service, but the collateral subject to repo entails an opportunity cost for the banks, as this collateral cannot be used for other purposes. The opportunity cost of collateral is assumed to be C per period per dollar.

If Bank A and Bank B both decide to process their requests early, then they each have to engage in an intraday repo with the central bank in order to obtain liquidity, and consequently they will each incur the cost C. However, if, say, Bank A decides to delay while Bank B decides to process, then Bank A will incur the cost of delay D in the morning period. However, in the afternoon period, it can use the incoming liquidity from Bank B to fund its own outgoing payment in the next period. Conversely, Bank B receives no liquidity and has to roll over the repo with the central bank for an additional period and incur the cost C one more time for a total of $2C$. Finally, if both banks choose to postpone, they both incur the cost of delay D. Moreover, at noon they still have no liquidity available, and both have to engage in an intraday repo in the afternoon period for which they each will incur the opportunity cost of collateral C. The settlement costs are summarized in Game 3, hereafter referred to as the *collateralized credit game*.

In the collateralized credit game, the equilibrium depends solely on the relative size of the opportunity cost of collateral and the cost of postponing a payment request. If

[5] In 2006, the Federal Reserve eliminated the extension of free intraday credit to government-sponsored enterprises (GSEs) and certain international organizations for the purpose of securities-related interest and redemption payments. This action was taken in part because, for some issuers, the lag between the time the Federal Reserve credited the interest and redemption payments to the recipients' accounts (early in the morning) and the time the issuer covered the resulting overdraft extended, at times, until shortly before the close of the Fedwire system, hence exposing the Federal Reserve to credit risk for the duration of the day. Currently, interest and redemption payments have to be funded up front.

[6] In models without credit risk, Freeman (1996) and Martin (2004) find that free intraday credit is the socially optimal policy.

Game 3
Collateralized Credit Game

		Bank B	
		morning	*afternoon*
Bank A	*morning*	C, C	$2C, D$
	afternoon	$D, 2C$	$C + D, C + D$

the cost of delaying is greater than the cost of obtaining liquidity—that is, $D > C$—then banks have no incentive to delay and the strategy profile (*morning, morning*) is the only Nash equilibrium. If Bank B plays morning, the best strategy for Bank A is to play morning as well. Moreover, if Bank B chooses to postpone, the best strategy for Bank A is still morning. In other words, morning is a dominating strategy for Bank A and, by symmetry, for Bank B as well. However, if the cost of liquidity is higher than the cost of delaying—that is, $C > D$—then the strategy profile (*afternoon, afternoon*) is the only Nash equilibrium. It is a unique Nash equilibrium, since neither bank wishes to switch to morning if the other bank keeps playing afternoon because a switch would increase its settlement cost. However, it is also clear that the banks would be better off if they both chose to process payments in the morning. Unfortunately, (*morning, morning)* is not an equilibrium in this one-shot game. Starting from (*morning, morning)*, each bank would wish to postpone payment in order to lower its settlement cost. This strategic situation is a classic paradigm in game theory called the prisoner's dilemma.[7] We summarize the results of the collateralized credit game in Result 2:

> *Result 2: In the collateralized credit game, early settlement* (morning, morning*) is a unique equilibrium if the opportunity cost of collateral is less than the cost of delaying*

[7] The "prisoner's dilemma" is the most famous paradigm in game theory. Suppose that the police have arrested two former felons who they know have committed an armed robbery together. Unfortunately, they lack enough admissible evidence to get a jury to convict them of armed robbery. They do, however, have enough evidence to send each prisoner away for two years for theft of the getaway car.

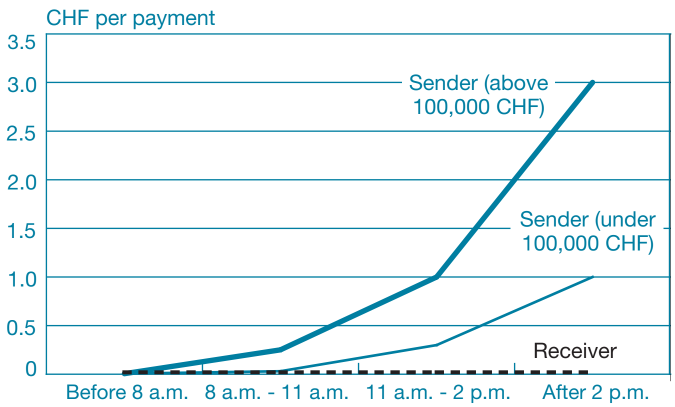
Prisoner's Dilemma

		Prisoner 2	
		Confess	Silence
Prisoner 1	Confess	5, 5	0, 10
	Silence	10, 0	2, 2

The chief inspector now makes the following offer to each prisoner: "If you will confess to the robbery, implicating your partner, and he does not also confess, then you'll go free and he will get ten years. If you both confess, you'll each get five years. If neither of you confesses, then you'll each get two years for the auto theft." It is a Nash equilibrium for each prisoner to confess; yet they would both be better off if they both chose to remain silent.

Chart 1

Pricing Structure for Swiss Interbank Clearing



Source: <http://www.sic.ch>.

*(*C < D*). This outcome is efficient. Conversely, late settlement (*afternoon, afternoon*) is a unique equilibrium if* C > D*, and the game is a prisoner's dilemma. Late settlement is inefficient.*

Central banks and other stakeholders in the interbank payments system are keenly aware that costly liquidity may lead to delays in processing payments or even to situations where the settlement of payments awaits the settlement of other payments. The latter situation is often referred to as *gridlock,* and the prisoner's dilemma above is a form of gridlock. Several different solutions to discourage banks from holding back payments have been employed around the world.

First, central banks seek to keep the opportunity cost of collateral low by accepting a wide range of different types and offering flexible arrangements for posting and using it. Recent examples include the European Central Bank, which recently expanded the pool of eligible collateral to include commercial loans, and the Scandinavian Cash Pool, which allows banks in Denmark, Norway, and Sweden to move collateral seamlessly across borders between the national RTGS systems.[8]

Second, some central banks and industry groups have put forward guidelines under which banks are to process certain percentages or types of traffic by predetermined times over the course of the business day. In the United Kingdom, members of the RTGS system are required to manage their payment flows in such a way that, on average, 50 percent of the value throughput is sent by noon and 75 percent is sent by 2:30 p.m. In Japan, banks are encouraged to return call money market loans within the first hour of operations.

[8] See Danmarks Nationalbank (2003).

Game 4

Offsetting in the Morning Game (*C* > *D*)

		Bank B	
		morning	*afternoon*
Bank A	*morning*	<u>0</u>, <u>0</u>	2*C*, *D*
	afternoon	*D*, 2*C*	<u>*C* + *D*</u>, <u>*C* + *D*</u>

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

Third, central banks can use pricing. For example, the Swiss National Bank charges higher prices for payments sent later in the day, thereby giving banks a direct incentive to process early. Moreover, the transaction fee increases more steeply for payments of larger value (Chart 1).

Finally, many systems place an upper limit on the value of payments, forcing larger payments to be split into smaller payments and thereby allowing balances to be used more efficiently. In Fedwire, the largest payments allowed are 1 cent short of $10 billion. In most cases, these solutions have been effective in securing smooth settlement of payments.

Nevertheless, in recent years, a number of RTGS systems with collateral requirements have introduced mechanisms that allow queued payments to be offset bilaterally or multilaterally. These enhancements were introduced with a view to reducing the amount of liquidity or collateral required for smooth settlement. An offsetting mechanism or gridlock resolution reduces the need to post collateral.

In the context of our model, an offsetting mechanism allows payments to be processed in a given period without the need to post collateral if an offsetting payment is submitted to the system in the same period. However, if no offsetting payments arrive, the system processes the payment at the end of the period and collateral needs to be posted. The situation where an offsetting mechanism is running only in the morning period is illustrated in Game 4. The prisoner's dilemma changes into a coordination game. Coordination games are a class of games with multiple (pure strategy) Nash equilibria in which players choose the same or corresponding strategies.

If Bank A submits in the morning, then the best response of Bank B is to do the same; if Bank A postpones to the afternoon, then the best response of Bank B is, again, to do the same.

A fundamental question in coordination games is which equilibrium the players will choose. In this case, it is fairly obvious. The game is a so-called pure coordination game, or game of common interest, in which both banks prefer the (*morning, morning*) equilibrium to the (*afternoon, afternoon*) equilibrium. In other words, early submission Pareto

Game 5
Priced Credit Game

		Bank B	
		morning	*afternoon*
Bank A	*morning*	0, 0	*F, D*
	afternoon	*D, F*	*D, D*

dominates late submission, and one would expect banks to choose the cost-efficient strategy. In sum, the introduction of a gridlock-resolution mechanism may change submission behavior. We offer the following conjecture:

> *Conjecture: Gridlock resolution and offsetting mechanisms may eliminate the potential prisoner's dilemma.*

The issue of liquidity-saving mechanisms is discussed further in Martin and McAndrews (2008).

5. Priced Intraday Credit Regime

Under the priced credit regime, banks are charged the fee *F* per dollar if their settlement account is overdrawn at the end of a period. This implies that no overdraft fee is incurred if the banks manage to synchronize their payments. The settlement costs associated with the different possible pairs of strategies are shown in Game 5, the priced credit game.

If both banks play morning, then payments net out and banks incur no costs. The payments also net out if both banks play afternoon, but each will incur the cost of delay. If one bank pays and the other delays, then the paying bank will incur an overdraft at noon while the other can use the incoming payment from the morning period to fund its outgoing payment in the afternoon. However, the bank that delays will incur the cost *D*.

As in the collateralized credit regime, the outcome depends on the relative size of the cost of liquidity and the cost of postponing the processing of a request. Again, the strategy profile (*morning, morning*) is a unique Nash equilibrium if the cost of liquidity is less than the cost of delaying the payment request—that is, $F < D$.

However, if $F > D$, then the strategy profiles (*morning, morning*) and (*afternoon, afternoon*) are both Nash equilibria. To see this, assume that $F = 5$ cents and $D = 2$ cents, as in Game 6. If both banks choose to process payments early, then neither bank would want to change and postpone because that would increase its settlement cost from 0 cents to 5 cents. Likewise,

Game 6
Priced Credit as the Stag Hunt Game

		Bank B	
		morning	*afternoon*
Bank A	*morning*	0, 0	5, 2
	afternoon	2, 5	2, 2

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

if both banks choose to postpone, neither bank would unilaterally want to deviate and process because that would increase its settlement cost from 2 cents to 5 cents.

Here, the priced credit game has the structure of a classic coordination game called the stag hunt.[9] The key feature of the stag hunt game is that while the (*morning, morning*) equilibrium is preferred by both players in terms of lowest cost, the other is preferred in terms of strategic risk. In the early-settlement equilibrium, the settlement cost of one bank depends on the action of the other. One bank's deviation from morning, for whatever reason, will impose increased settlement costs on the other bank. In contrast, the strategy to postpone payments carries no risk in the sense that the settlement cost is the same regardless of which action the other bank takes. A cautious bank may reasonably choose to postpone, ensuring the 2 cents with certainty rather than risking the cost of 5 cents. This is especially true if there are concerns regarding the other bank's ability to coordinate (for example, because of operational risk). We recap the results of the priced credit game in Result 3:

> *Result 3: In the priced credit game, early settlement* (morning, morning) *is a unique equilibrium if the overdraft fee is less than the cost of delaying* ($F < D$)*. The outcome is efficient. In contrast, both* (morning, morning) *and* (afternoon, afternoon) *are feasible equilibria if* $F > D$ *and the game is a stag hunt. Late settlement is inefficient.*

In the prisoner's dilemma, there is a conflict between individual rationality and mutual benefit. In the stag hunt, rational players are pulled in one direction by consideration of mutual benefit and in the other by individual risk concerns (Skyrms 2004). In the stag hunt game, the outcome depends on the player's appetite for strategic risk—that is, the uncertainty

[9] The "stag hunt" is a story that became a game. The game is a prototype of the social contract. The story is briefly told by the eighteenth-century philosopher Jean-Jacques Rousseau in *A Discourse on Inequality* (Skyrms 2004): "If it was a matter of hunting a deer, everyone well realized that he must remain faithful to his post; but if a hare happened to pass within reach of one of them, we cannot doubt that he would have gone off in pursuit of it without scruple."

that arises from the interaction between the players of the game. The conflict in the priced credit game is a trade-off between lower settlement costs and strategic risk.

One way of pinning down a unique equilibrium is by using Harsanyi and Selten's (1988) concept of *risk dominance*.[10] In a symmetric 2 x 2 game, risk dominance asserts that players will choose the strategy that gives the highest expected payoff under the assumption that the opponent randomizes with equal probability over the two available strategies. Fixing D to be 2 cents in Game 6 implies that (*morning, morning*) is the risk-dominant equilibrium if $F < 4$ cents and (*afternoon, afternoon*) is the outcome if $F > 4$ cents.

> *Result 4: In the priced credit game, the risk-dominant equilibrium is early settlement (*morning, morning*) if* $F < 2D$. *Otherwise, late settlement (*afternoon, afternoon*) is the risk-dominant equilibrium.*

Using the analysis above, we now turn to a comparison of the aggregate settlement costs to the economy under collateralized and priced intraday credit policy regimes.

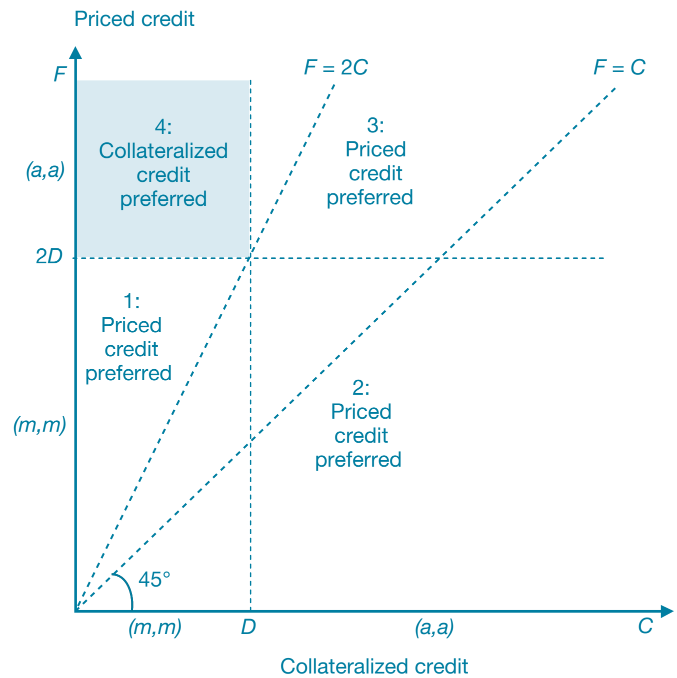
6. Choice of Intraday Credit Policy

An omnipresent question for central banks is the choice of intraday credit policy. For example, the Board of Governors of the Federal Reserve System is currently reviewing its Payment System Risk Policy with a view to reducing liquidity risk, credit risk, and operational risk while maintaining or improving payments system efficiency. For this reason, the Federal Reserve Board published a consultation paper in June 2006 to elicit information from financial institutions and other interested parties on their experiences managing intraday risk associated with Fedwire funds transfers. Our model can provide insight into the desirability of different payments system policies and highlight some of the difficulties facing policymakers.

Assume that a central bank is a benevolent provider of the RTGS system insofar as it seeks to secure the lowest possible aggregate settlement costs for the economy. The central bank can choose between a collateralized credit and a priced credit regime. Additionally, for the purposes of this analysis, assume that the central bank cannot (further) influence the cost of liquidity under either regime or the cost of delay. The preferred regime then depends on the equilibrium outcome under the two regimes, which in turn depends on the relative magnitudes of F, C, and D.

[10] In 1994, John C. Harsanyi and Reinhard Selten were awarded the Nobel Prize in Economics, together with John F. Nash Jr., for their pioneering analysis of equilibria in the theory of noncooperative games.

Chart 2
Comparative Analysis of Intraday Credit Regimes



The aggregate settlement costs under the equilibria of the two intraday credit regimes are easily calculated by summing the entries in each cell in Game 3 and Game 5, respectively. The aggregate settlement costs when one bank is playing morning and the other afternoon are $2C + D$ and $F + D$, respectively, under the two regimes. With two possible (risk-dominant) Nash equilibria under each regime, there are four different scenarios to consider.

The comparative analysis is summarized in Chart 2. The X-axis shows the Nash equilibrium in a collateralized credit regime as a function of the opportunity cost of collateral. The Y-axis shows the risk-dominant Nash equilibrium under a priced credit regime as a function of the overdraft fee. In the priced credit regime, aggregate settlement costs are zero if the equilibrium is (*morning, morning*). In contrast, a collateralized credit regime always implies positive settlement costs. Consequently, priced credit is the preferred regime if $F < 2D$—that is, in scenarios 1 and 2. In other words, take the parameters of the model as exogenously given.

If payments are delayed under both regimes—that is, $2D < F$ and $D <$ C—then aggregate settlement costs are $2D + 2C$ and $2D$, respectively. Hence, priced credit is the preferred regime C in scenario 3 in Chart 2. Conversely, collateralized credit is the preferred regime if banks do not delay payments under such a regime but they do under a priced credit regime—that is, $C < D < 2D < F$ (scenario 4). We summarize this as:

> *Result 5: Priced credit is preferred to collateralized credit except when collateralized credit leads to quicker settlement of payments.*

The model provides very clear results in terms of the desirability of the two regimes, but in reality the analysis is more involved. Moreover, the analysis does not take into account default risk, against which the collateral protects. A challenge for comparative analysis in practice is that the cost of delay is not observable. In fact, little is known about the costs banks face if they delay settlement of payments. Without knowledge of the cost of delay, the comparative analysis becomes less informative, but the simple analysis presented in Chart 2 does yield the following necessary, but not a sufficient, condition for collateralized credit to be the preferred regime:

> *Result 6: For collateralized credit to be the preferred regime, a necessary condition is that the opportunity cost of collateral be lower than (literally half) the overdraft fee charged under priced credit.*

The opportunity cost of collateral is not directly observable either, but the rate differential between federal funds loans, which are uncollateralized, and loans through repurchase agreements, which are collateralized, suggests that the opportunity cost is in the range of 12 to 15 basis points per annum (see Board of Governors of the Federal Reserve System [2006]).[11] However, the overdraft fee is readily observable because it is set by the central bank with a view to managing credit exposure from overdrafts. Currently, daylight overdraft fees in the Fedwire Funds Service are calculated using an annual rate of 36 basis points, quoted on the basis of a 21.5-hour day. This simple "back-of-the-envelope" comparison suggests that there may be scope for investigating an increased role for collateral in the Fedwire system.

In the following sections, we investigate how the conclusions from the model are likely to change as more realism is added. We start by considering settlement risk, followed by incomplete information, repeated play, and more than two banks and periods. In Box 1 and Box 2, we analyze, respectively, the strategic interaction between banks when there is no intraday credit available and when banks are heterogeneous.

[11] The opportunity cost of collateral would, in all likelihood, increase if the Federal Reserve implemented a collateralized credit regime because the demand for collateral would increase.

7. Settlement Risk

Settlement risk is an important concern in all payment arrangements (see, for example, Kahn, McAndrews, and Roberds [2003] and Mills and Nesmith [2008]). Fundamentally, it is the risk that settlement does not take place as expected. As such, settlement risk comprises both liquidity and credit risks. Liquidity risk is the risk that a counterparty will not settle an obligation for full value when due, but at some unspecified time thereafter. Credit risk is the risk that a counterparty will not settle an obligation for full value either when due or anytime thereafter. The presence of settlement risk affects the strategic interaction between banks and hence their intraday liquidity management behavior.

To illustrate how settlement risk affects strategic interaction, we assume that the banks have entered into trades with each other yielding obligations to pay the other \$1, to be settled gross. On settlement day, a bank might experience an operational incident or default altogether, which leads to either a temporary inability or permanent failure to pay. Because payment flows are symmetric, neither bank starts out with an exposure vis-à-vis the other at dawn. If a bank defaults before the opening of the RTGS system, the other bank can just withhold its payment. However, by paying early, a bank exposes itself to the inability or failure of the other to pay. Everything else being equal, one would expect banks to be more cautious in their behavior when facing settlement risk. In essence, settlement risk reduces the effective cost of delaying.

The presence of settlement risk affects the strategic interaction between banks and hence their intraday liquidity management behavior.

Specifically, we model liquidity risk by assuming that, with probability α, banks will not be able to submit payments to the RTGS system in the morning because of, say, a telecommunications outage. However, the telecommunications links to stricken banks are reestablished at noon and banks can then make payments in the afternoon period. The expected costs for banks are derived in the appendix. These costs are used in the intraday liquidity management games with liquidity risk shown in Game 7 and Game 8 for the two policy regimes (collateralized credit and priced credit). For convenience, we show only the costs for Bank A, but by symmetry the costs are the same for Bank B.

Box 1

The No Intraday Credit Game

The exception proving the rule that early adopters provided intraday credit for free is Switzerland. The Swiss National Bank implemented real-time gross settlement (RTGS) systems in 1987, but did not provide intraday credit until the autumn of 1999. The change in policy was motivated by an increase in time-critical payments and, in particular, the future introduction of the Continuous Linked Settlement system for foreign exchange transactions. According to Heller, Nellen, and Sturm (2000), the amount of payments settled by noon rose from one-third to one-half of the daily turnover as a result.

Going against conventional wisdom, the Reserve Bank of New Zealand implemented a new liquidity management regime in 2006 that discontinued its intraday automatic reverse repurchase facility (autorepo). Instead, the Reserve Bank chose to supply a significantly higher level of cash (overnight monies) sufficient to enable participants to settle payments efficiently. The change was necessitated by a growing scarcity of New Zealand government securities (see Reserve Bank of New Zealand [2006]).

If the central bank does not provide intraday credit, then payments have to be funded by balances held with the central bank, interbank money market borrowings, or incoming payments from other banks. The first two sources are costly, whereas the last is free from the perspective of the receiver. Let ρ denote the (marginal) opportunity cost of balances held at the central bank. The opportunity cost of reserves is closely linked to the central bank's policy with respect to remunerating reserves. If the central bank does not pay interest on reserves, then the opportunity cost is close to the overnight money market rate, whereas if the central bank does pay interest on reserves, it depends on the difference between the money market rate and the administrative rate paid on reserves.

The no intraday credit game is given below for the interesting case where $\rho > D$.

No Intraday Credit Game ($\rho > D$)

		Bank B	
		morning	*afternoon*
Bank A	*morning*	ρ, ρ	$\underline{\rho}, \underline{D}$
	afternoon	$\underline{D}, \underline{\rho}$	$D + \rho, D + \rho$

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

If the opportunity cost of reserve balances is less than the cost of delay, then (*morning, morning*) is the equilibrium in dominating strategies. However, if the opportunity cost of reserves is larger than the cost of delay, the game is an *anti-coordination game*, so named because it is mutually beneficial for the players to play different strategies. If Bank B plays morning, then the best response of Bank A is to play afternoon. Conversely, if Bank B plays afternoon, then the best strategy for Bank A is to play morning. The underlying conflict in the game is that both banks want to benefit from free liquidity, but liquidity is rivalrous—that is, banks cannot benefit from it at the same time. Hence, both (*morning, afternoon*) and (*afternoon, morning*) are possible Nash equilibria, but neither Pareto dominates the other or is focal in any sense. The mixed-strategy Nash equilibrium implies that banks play morning with probability $p = D/\rho$ and afternoon with the complementary probability. The expected settlement cost for a bank is ρ (see appendix). Hence, the mixed strategy does not Pareto dominate either of the pure Nash equilibria, and a bank might as well play morning and save itself the trouble of randomizing.

It is not obvious how banks can solve the conundrum of who gets the benefit of free liquidity. One solution in these types of games is for banks to engage in pre-play communication. In pre-play communication, each player announces the action it intends to take (or, alternatively, the action it would like the other to take). In game theory, pre-play communication that carries no cost is referred to as cheap talk (Farrell and Rabin 1996). Interestingly, in some experimental settings, cheap talk has been found to be effective. Another form of pre-play communication is for one bank to signal convincingly that it will play afternoon. One way to do this would be to open late, but that would probably be bad for business in general and thus costly.

Aumann (1974) provides a generalization of Nash equilibrium known as *correlated equilibrium*, which allows for possible dependencies in strategic choices. A perfectly correlated equilibrium would be for banks to use a fair coin to determine which bank gets to play afternoon. In a repeated setting, a convention for banks to alternate sending early could conceivably evolve.

Above and beyond the potential instability of the equilibrium outcome, a key insight of the no intraday liquidity management game is that the monetary policy stance may directly affect the settlement of payments intraday owing to the close link between the opportunity cost of holding reserves and the overnight interest rate. Any movement in the monetary policy stance will affect the opportunity cost and may shift the equilibrium around. Interestingly, Heller, Nellen, and Sturm (2000) claim that a less restrictive monetary policy stance from 1993 to 1999 can explain a large part of the reduced congestion observed in Swiss Interbank Clearing, as this led banks to hold increased account balances.

Box 2

Heterogeneity

In the analysis, we focus on the interaction between two identical banks. Obviously, participants in a real-time gross settlement system are not a homogenous group. Here, we explore the implications of introducing heterogeneity among participants. For ease of exposition, we consider only two cases. First, we look at the case where participants face different liquidity and delay costs. We do this in the context of a recent policy change in the Fedwire system. Second, we consider the case where payment flows are not balanced and then we try to gauge the extent to which that affects the strategic interaction between banks.

In 2006, the Federal Reserve eliminated the extension of free intraday credit to government-sponsored enterprises (GSEs) and certain international organizations for the purpose of securities-related principal and interest payments. This action was taken in part because, for some issuers, the lag between the time the Federal Reserve credited the interest and redemption payments to the recipients' accounts (early in the morning) and the time the issuer covered the resulting overdraft extended, at times, until shortly before the close of the Fedwire system. As a result, the Federal Reserve was exposed to credit risk for the duration of the day. Currently, principal and interest payments have to be funded up front.

To see how the simple framework can account for this observation, assume that one player is now an issuer of securities and needs to pay \$1 in principal and interest to the other player—a bank. Assume that issuer had the necessary cash to pay out principal and interest on hand the previous day and chose to lend it out in the overnight money market to earn a return. For simplicity, also assume that the borrower was the bank that henceforth has to return the \$1 plus interest (ρ) to the issuer.

The central bank is granting free intraday credit to the issuer but charges the bank for overdrafts. Owing to market conventions, the cost of delaying the payout of principal and interest is high (H), whereas the cost of delaying the return of a money market loan to a participant that has access to free intraday credit is virtually nil. The resulting *principal and interest game* is shown below.

Principal and Interest Game

		Issuer	
		morning	*afternoon*
Bank	*morning*	$\underline{\rho F}$, 0	$(1+\rho)F$, H
	afternoon	$\underline{\rho F}$, 0	$\underline{\rho F}$, H

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

Clearly, it is a dominating strategy for the issuer to pay out early. If the issuer plays morning, then the bank is indifferent between returning early or late. However, returning the overnight loan late is a weakly dominating strategy because if the issuer for some reason should delay then it would be best for the bank to delay as well. A small intraday opportunity cost for a bank using overdraft capacity to cover the interest on the loan would eliminate the (*morning, morning*) equilibrium. The (*morning, afternoon*) equilibrium leaves the issuer with an overdraft at the central bank for the entire day. In sum, different cost structures for participants can lead to interesting games with asymmetric equilibria. We now turn to payment flow imbalances.

On any given day, payment flows are never balanced because banks receive different amounts of payment requests from their customers. Banks manage their projected end-of-day balances throughout the day. Liquidity is redistributed via the interbank money market from the "haves" to the "have nots." The question is the extent to which such differences in payment flows can affect the strategic interaction among banks. To provide insight, we assume that Bank B has two \$1 payments to send to Bank A whereas Bank A still has only \$1 to send to Bank B. The strategy set expands for Bank B, which can choose to send them both early, delay them both, or send one early while holding back the other. The resulting games are shown below.

Collateralized Credit Game with Payment Flow Imbalance ($C > D$)

		Bank B		
		m,m	*m,a*	*a,a*
Bank A	*m*	C, $3C$	C, $D+2C$	$2C$, $\underline{2D+C}$
	a	$\underline{D}$, $4C$	$\underline{D}$, $D+3C$	$\underline{D+C}$, $\underline{2D+2C}$

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

The games become slightly more complicated, but the fundamental issues remain. In the case of collateralized credit, banks may still end up delaying payments even though it is more efficient to process early in terms of minimizing aggregate settlement cost. In the case of priced credit, it is possible only to offset two payments against each other, and thus it turns out that Bank B will always hold back one payment. The stag hunt is played with the remaining payment.

Priced Credit Game with Payment Flow Imbalance ($F > D$)

		Bank B		
		m,m	*m,a*	*a,a*
Bank A	*m*	$\underline{0}$, $2F$	$\underline{0}$, $\underline{D+F}$	F, $F+2D$
	a	D, $3F$	D, $D+2F$	$\underline{D}$, $\underline{2D+F}$

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

Game 7

Collateralized Credit Game with Liquidity Risk

		Bank B	
		morning	*afternoon*
Bank A	*morning*	$(1+\omega)C+\omega D$	$2C+\omega D$
	afternoon	$D+\omega C$	$D+C$

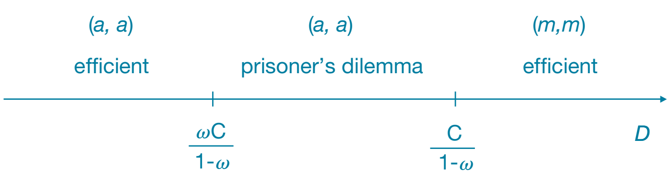
Game 8

Priced Credit Game with Liquidity Risk

		Bank B	
		morning	*afternoon*
Bank A	*morning*	$\omega(1-\omega)F+\omega D$	$(1-\omega)F + \omega D$
	afternoon	D	D

Exhibit 1

Equilibrium and Efficiency in the Collateralized Credit Game with Liquidity Risk



The introduction of liquidity risk implies that a bank risks incurring the cost of delay even if it is playing the morning strategy. On the flip side, the other bank incurs additional liquidity costs due to the lack of an incoming payment. As such, liquidity risk affects both the equilibrium outcomes and the efficiency thereof. In a collateralized credit regime, the (*afternoon, afternoon*) equilibrium becomes more likely (Exhibit 1).

Without liquidity risk, the condition for late settlement is $D < C$, whereas with liquidity risk it is $D < C/(1-\omega)$. Increasing the exposure to liquidity risk—that is, $\omega \to 1$ — raises the likelihood that banks hold back payments as $C/(1-\omega) \to \infty$. Interestingly, holding back payments is the efficient outcome if the cost of delay is sufficiently low or the risk of a temporary failure to pay is high. The prisoner's dilemma disappears if $D < \omega C/(1-\omega)$.

With priced credit, banks will still play (*morning, morning*) if the cost of liquidity is less than the cost of delay; and if D drops below F, the stag hunt emerges, as in Result 3 (Exhibit 2). However, (*afternoon, afternoon*) is now a feasible unique Nash equilibrium. This equilibrium is efficient and will be the outcome if the exposure to liquidity risk is sufficiently high—that is, $\omega > D/F$.

To model credit risk, we assume that a bank with probability δ will be closed by its regulator at noon. Should that occur, the bank will not be making any further payments. Thus, the other bank has to borrow from the discount window at rate R in order to square its account at the end of the day. Furthermore, assume that a surviving bank eventually will recover $(1-\alpha)$ of the dollars that it is owed. Hence, the total cost of default is $\theta = \alpha + R$.

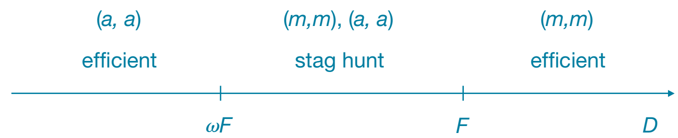
We assume that if a bank defaults then there is no reputation cost of delaying. Hence, the expected cost of delay is $(1-\delta)D$. The resulting games for the collateralized and priced credit regimes are shown in Game 9 and Game 10, respectively. Again, the settlement costs are derived in the appendix, and only the expected settlement costs for Bank A are shown.

For collateralized credit, it turns out that the results are identical to those for liquidity risk. The only difference is that the probability of a default, δ, replaces the probability of a temporary failure to pay, ω, in Exhibit 1. With priced credit, banks will play the (*morning, morning*) equilibrium whenever $D > F + \delta\theta$ compared with $D > F$, when there is no risk of default, as in Result 3. Otherwise, the game is a stag hunt. We sum up the results from introducing settlement risk as:

> *Result 7: Settlement risk makes (other things being equal) late settlement* (afternoon, afternoon) *a more likely outcome of the intraday liquidity management game. Late settlement may be efficient.*

Exhibit 2

Equilibrium and Efficiency in the Priced Credit Game with Liquidity Risk



Game 9

Collateralized Credit Game with Credit Risk

		Bank B	
		morning	*afternoon*
Bank A	*morning*	C	$(2-\delta)C+\delta(1-\delta)\theta$
	afternoon	$(1-\delta)D$	$(1-\delta)(C+D)$

Game 10

Priced Credit Game with Credit Risk

		Bank B	
		morning	*afternoon*
Bank A	*morning*	0	$(1-\delta)F+\delta(1-\delta)\theta$
	afternoon	$(1-\delta)D$	$(1-\delta)D$

8. Incomplete Information

The analysis so far has assumed that banks have complete information with regard to the payments to be settled. In reality, banks have only an incomplete picture during the day. In fact, there can be substantial uncertainty about both incoming payments and requests that customers will submit over the remainder of the day. This ambiguity further complicates the task of managing the liquidity position of a bank within the day.

Bech and Garratt (2003) develop a Bayesian game in which banks have private knowledge about their own pending payment requests but only imperfect information about those of the opponent. Moreover, banks face uncertainty (fundamental) about the arrival of new payment requests and uncertainty (strategic) in terms of the opponent's action. In the model, payment requests arrive from customers at dawn and at noon with probabilities p and q, respectively. Banks seek to minimize expected settlement costs. It turns out that the strategies of banks are determined by the action they take when they do receive a request at dawn. This simplifies the analysis and allows us to stay within a 2 x 2 framework for the purposes of determining equilibria. We construct games where the payoffs are conditional on having received a request from the perspective of each bank. For example, the expected settlement costs of sending early against an opponent that also sends early (if it has a request) are $C + (1–p)C$ and $(1–p)F$, respectively. The extra component relative to collateralized and priced credit games described earlier reflects the chance that the opponent might not have received a request and hence the bank would have to borrow additional liquidity from the central bank. The outcomes under the two intraday credit regimes—now Bayes-Nash equilibria—are determined by Game 11 and Game 12, respectively. However, the full Bayesian game is needed to evaluate the efficiency implications of different strategy profiles as banks individually do not take into account the positive externality of liquidity to the other bank. We cite the results on efficiency here and refer the reader to the original paper (Bech and Garratt 2003) for the details.

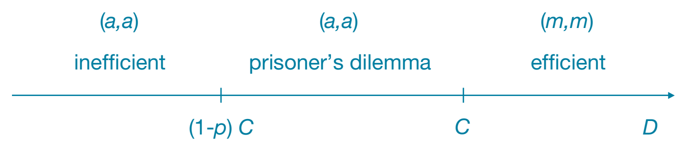
Game 11
Collateralized Credit Game with Incomplete Information

		Bank B	
		morning	*afternoon*
Bank A	*morning*	$(2\text{-}p)C$	$2C$
	afternoon	$D + (1\text{-}p)C$	$C + D$

Game 12
Priced Credit Game with Incomplete Information

		Bank B	
		morning	*afternoon*
Bank A	*morning*	$(1\text{-}p)F$	F
	afternoon	D	D

Exhibit 3
Equilibrium and Efficiency in the Bayesian Collateralized Credit Game



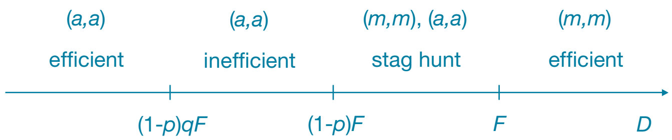
In the case of collateralized credit, it is still true that (*morning, morning*) is the Nash equilibrium whenever $D > C$ because the additional cost of $(1\text{-}p)C$ is incurred regardless of whether the bank in question is playing morning or afternoon. Otherwise (*afternoon, afternoon*) is the equilibrium (Exhibit 3). While Game 11 is only a prisoner's dilemma when the additional cost of delaying is larger than the expected cost of processing the payment—that is, $(1–p)C < D < C$—it is still inefficient, from an aggregate expected settlement cost perspective, to delay. Early settlement is the only efficient outcome.

In the case of priced credit, (*morning, morning*) is again the equilibrium whenever $D > F$. Conversely, if $D < (1 - p)F$, then the strategy profile (*afternoon, afternoon*) is the Bayes-Nash equilibrium (Exhibit 4). In the intermediate case, the stag hunt emerges once again. In contrast to the case of collateralized credit, it is possible for postponement of payments to be efficient in the case of priced credit. This occurs if $D < (1\text{-}p)qF$—that is, if the cost of delay is low (relative to the overdraft fee) and the arrival of payment requests is sufficiently skewed toward the afternoon (low p and high q). In that case, the expected benefit from being able to offset payments in the afternoon outweighs the cost of delay.

We reiterate the outcome of introducing incomplete information and random arrivals of payment requests in the following result:

> *Result 8: Incomplete information about payment flows increases (other things being equal) the likelihood of late settlement* (afternoon, afternoon)*. In the case of priced credit, late-day requests may make late settlement efficient.*

Exhibit 4
Equilibrium and Efficiency in the Bayesian Collateralized Credit Game



9. Repeated Interaction

In most payments systems, participating banks interact repeatedly with each other, both within and across days. It is well known that playing the same game (such as the prisoner's dilemma) numerous times might yield a different outcome. Unlike a game played once, a repeated game allows for a strategy to be contingent on past moves, thus allowing for reputation effects and retribution. The key is that cooperation now can be rewarded by cooperation later, and cheating can be punished by not cooperating later. It is thus not always wise to pursue a short-run gain in a repeated game.

A *trigger* strategy in which cheating is punished in subsequent periods can encourage cooperation. One harsh example is for a player to begin by cooperating in the first period and to continue cooperating until a single defection by the opponent, after which the player never cooperates again. A less harsh trigger strategy is "tit for tat," where a player responds in one period with the same action the opponent used in the last period. The repetition of a game may solve some of the single-play issues discussed above. However, by offering more complex strategies, a repeated game can also result in more equilibrium outcomes. In other words, the repetition of a game itself does not necessarily solve the quandaries faced by players in single-play games. Additional structure is often needed.

Here, we assume an infinite play setting where banks discount the future. The daily discount factor is given by $0 < \beta < 1$. Banks can choose between two possible strategies. One strategy is to always delay. The other is a trigger strategy whereby a bank will send early as long as the other does, but will delay afterward if the other bank deviates. Using the formula for infinite geometric series, we can compute the future discounted settlement cost under the two strategies for each of the two intraday credit regimes. For example, in a collateralized credit regime where both banks are playing trigger, the future discounted settlement costs are:

$$c^i(t,\ t) = C + \beta C + \beta^2 C + \beta^3 C + \ldots = \frac{C}{1-\beta},$$

Game 13
Repeated Collateralized Credit Game

		Bank B	
		trigger	*always delay*
Bank A	*trigger*	$\frac{C}{1-\beta}$	$2C + \frac{\beta(C+D)}{1-\beta}$
	always delay	$D + \frac{\beta(C+D)}{1-\beta}$	$\frac{C+D}{1-\beta}$

Game 14
Repeated Priced Credit Game

		Bank B	
		trigger	*always delay*
Bank A	*trigger*	0	$F + \beta D/(1\text{-}\beta)$
	always delay	$D/(1\text{-}\beta)$	$D/(1\text{-}\beta)$

where *t* denotes the trigger strategy and bank $i \in \{A, B\}$. In the case of priced credit, the future discounted settlement costs for a bank playing trigger strategy against an opponent that always delays are the overdraft fee in the first period and then the cost of delay for any subsequent days in which they interact. That is:

$$c^i(\ t, a\ ) = F + \beta D + \beta^2 D + \beta^3 D + \ldots = F + \frac{\beta}{1-\beta} D,$$

where *a* denotes the "always delay" strategy. The settlement cost for the remaining strategy profiles can be derived in a similar fashion. The resulting games are shown in Game 13 and Game 14, respectively. We show only the discounted future settlement cost for Bank A. The prisoner's dilemma remains in the collateralized credit regime if the future matters little to banks, as illustrated in Exhibit 5. In fact, with $\beta = 0$, we get the one-stage collateralized credit game described earlier. However, if the discount factor is significantly large—that is, $\beta > 1 - D / C$—then repeated play transforms the prisoner's dilemma into a stag hunt. Moreover, as shown in the appendix, if the discount factor is even higher—that is, $\beta > 2(C-D) / (2C-D)$—then the risk-dominant equilibrium is (*trigger*, *trigger*).

In the case of priced credit, the infinitely repeated version of the game remains a stag hunt game if $F > D$. However, early processing is the risk-dominant equilibrium if $F < (2-\beta) D / (1-\beta)$ compared with $F < 2D$ in the one-stage game (see appendix). Hence, the more the future matters for banks, the more likely it becomes that banks will coordinate toward early processing. We summarize the results of introducing repeated play in Result 9:

Exhibit 5

Equilibria and Efficiency in the Repeated Collateralized Credit Game ($C > D$)

Result 9: In a repeated game setting with a trigger strategy, the prisoner's dilemma in the case of collateralized credit may turn into a stag hunt if the discount factor is sufficiently high. In the priced credit regime, the stag hunt game remains a stag hunt. Under both regimes, the likelihood of early processing is increasing in the value placed on future costs.

10. More Players

The number of participants in RTGS systems around the world varies significantly. In the United Kingdom, the CHAPS Sterling system has fifteen direct participants, whereas the Federal Reserve's Fedwire Funds Service has more than 7,000. Obviously, our two-player framework is a simplification of reality. Adding additional banks to the mix increases the dimensionality of the game. With three banks each having a dollar to send to one another, the number of different strategy profiles increases from four to sixty-four as banks now can delay to one bank while sending early to the other. With four banks, the same number is 4,096.

Here, we focus on the three-player game, where each bank has a dollar to send to the other two banks. The settlement costs of Bank $i \in \{A, B, C\}$, given its own strategy and the number of other banks sending payments early to the bank in question, are shown in Game 15 and Game 16, respectively.

In the collateralized credit game, it is still a dominating strategy to delay. Hence, $((a, a), (a, a), (a, a))$ is a unique equilibrium, but the outcome is inefficient. In the priced credit game, the best response of Bank i is to do the same if the two opponents either send or delay all their payments. If one bank is playing morning and the other is playing afternoon, then it is the best response of Bank i also to send one early and delay the other. However, such strategy profiles are only equilibria if the payment flow somehow miraculously forms a cycle—for example, Bank A sends to Bank B, which sends to Bank C, which sends to Bank A. In other words, the underlying payment flows become intertwined with the strategic interaction. Understanding the network structure of payment flows is important when analyzing behavior in interbank payments systems (see, for example, Soramäki et al. [2007] and Bech and Garratt [2006]). We summarize the effects of expanding the number of players as:

Game 15

Three-Player Game with Collateralized Credit ($C > D$)

		Number of Banks Playing Morning vis-à-vis Bank i		
		0	1	2
Bank i	(m,m)	$4C$	$3C$	$2C$
	(m,a) or (a,m)	$3C+D$	$2C+D$	$C+D$
	(a,a)	$2C+2D$	$2D+C$	$2D$

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

Game 16

Three-Player Game with Priced Credit ($F > D$)

		Number of Banks Playing Morning vis-à-vis Bank i		
		0	1	2
Bank i	(m,m)	$2F$	F	0
	(m,a) or (a,m)	$D+F$	D	D
	(a,a)	$2D$	$2D$	$2D$

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

Result 10: Adding more players does not fundamentally change the strategic interaction of the intraday liquidity management game.

11. More Actions

A trend among RTGS systems has been to extend operating hours. Since 2001, the Fedwire Funds Service has opened at 9:00 p.m. ET on the preceding calendar day and closed at 6:30 p.m. ET.[12] In comparison, the Swiss Interbank Payment system opens at 5:00 p.m. on the preceding calendar day and closes at 4:15 p.m., thus approaching around-the-clock processing. This trend, coupled with the fact that RTGS systems operate in continuous

[12] For example, on a Sunday, the Fedwire Funds Service will open at 9:00 p.m. ET with a cycle date of Monday, although transfers sent from 9:00 p.m. to midnight ET on Sunday will settle in real time on Sunday.

Game 17
Collateralized Credit Game with Three Periods
(C > D)

		Bank B		
		m	*a*	*e*
Bank A	*m*	*C, C*	2*C*, *D*	3*C*, 2*D*
	a	*D*, 2*C*	*C*+*D*, *C*+*D*	*D*+2*C*, 2*D*
	e	2*D*, 3*C*	2*D*, *D*+2*C*	2*D*+*C*, 2*D*+*C*

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

Game 18
Priced Credit Game with Three Periods
(F > D)

		Bank B		
		m	*a*	*e*
Bank A	*m*	0, 0	*F, D*	2*F*, 2*D*
	a	*D, F*	*D, D*	*F* + *D*, 2*D*
	e	2*D*, 2*F*	2*D*, *F* + *D*	2*D*, 2*D*

Note: We underline the cost associated with the strategy that is the best response of a bank given the strategy played by the other.

time, suggests that a model with only a morning and an afternoon period is perhaps too coarse a representation.

As a first step, we extend our model with an extra period denoted *evening*. In the case of collateralized credit, the prisoner's dilemma remains if the opportunity cost of collateral is larger than the cost of delay. Banks end up playing (*evening, evening*), even though it would be better for them to play either (*morning, morning*) or (*afternoon, afternoon*), as shown in Game 17.

In the case of priced credit, adding an extra period yields the game shown in Game 18. The game remains a coordination game if the cost of liquidity is larger than the cost of delay. The strategy profile (*evening, evening*) is now an additional feasible Nash equilibrium. As the number of periods increases, it might become increasingly difficult for banks to coordinate. In such cases, focal points (often referred to as *Schelling points*), which are solutions that for some reason seem natural or special, may offer guidance in terms of equilibrium selection.[13] As discussed by McAndrews and Rajan (2000), focal points may, in the context of RTGS systems, include times at which ancillary payments or securities settlement systems settle their final positions.

Result 11: Expanding the number of periods within the day does not fundamentally change the strategic interaction of the intraday liquidity management game, but it does make equilibrium selection more difficult.

12. Conclusion

This article presents a simple game-theoretical framework that can be used to address both positive and normative economic questions associated with intraday liquidity management. The simplicity of the framework is both its strength and its weakness. The strength is that it clearly exposes the fundamental trade-offs associated with strategic interaction in an RTGS environment. However, the extensions discussed highlight the complexity faced by banks in managing intraday liquidity, the challenges faced by policymakers, and consequently the difficulty in devising an all-encompassing framework. Nonetheless, our analysis shows the commonality of issues faced by all stakeholders in the world's interbank payments systems.

The ongoing relevance of the issues discussed in this article is exemplified by the Federal Reserve Board's February 28, 2008, request for public comments on proposed changes to its Payments System Risk policy that are intended to loosen intraday liquidity constraints and reduce operational risks in financial markets and the payments system.[14] The Board is proposing a new strategy for providing intraday credit to depository institutions and would encourage these institutions to collateralize their daylight overdrafts. Specifically, the Board proposes to adopt a policy of supplying intraday balances to depository institutions predominantly through voluntarily collateralized daylight overdrafts. The proposed policy would encourage the voluntary pledging of collateral to cover daylight overdrafts by providing collateralized daylight overdrafts at a zero fee and by raising the fee for uncollateralized daylight overdrafts to 50 basis points (annual rate) from the current 36 basis points. The Board expects that a revised Payments System Risk policy could be implemented approximately two years from the adoption of a final rule.

[13] Thomas C. Schelling was awarded the Nobel Prize in Economics in 2005 "for having enhanced our understanding of conflict and cooperation through game-theory analysis." One contribution was the notion of focal points. Schelling found that coordinative solutions—which he called *focal points*—could be arrived at more often than predicted by theory. The ability to coordinate appears to be related to the parties' common frames of reference. Social conventions and norms are integral parts of this common ground (see <http://nobelprize.org/nobel_prizes/>).

[14] See <http://www.federalreserve.gov/newsevents/press/other/20080228a.htm>.

Appendix

Expected Settlement of Mixed Strategy

$$E[c] = p^2 c(m,m) + p(1-p)c(m,a) + (1-p)pc(a,m) + (1-p)^2 c(a,a)$$
$$= \left(\frac{D}{\rho}\right)^2 \rho + \frac{D}{\rho}\left(1-\frac{D}{\rho}\right)\rho + \left(1-\frac{D}{\rho}\right)\frac{D}{\rho}D + \left(1-\frac{D}{\rho}\right)^2(\rho + D)$$
$$= \rho$$

Settlement Risk

Collateralized credit liquidity risk:

$$c^A(m,m) = \omega^2(2C+D) + \omega(1-\omega)(C+D) + (1-\omega)\omega 2C + (1-\omega)^2 C = (1+\omega)C + \omega D$$
$$c^A(m,a) = \omega^2(2C+D) + \omega(1-\omega)(2C+D) + (1-\omega)\omega 2C + (1-\omega)^2 2C = 2C + \omega D$$
$$c^A(a,m) = \omega^2(C+D) + \omega(1-\omega)D + (1-\omega)\omega(C+D) + (1-\omega)^2 D = \omega C + D$$
$$c^A(a,a) = (\omega^2 + \omega(1-\omega)C + (1-\omega)\omega C + (1-\omega)^2)(C+D)$$

Credit risk:

$$c^A(m,m) = \delta^2 C + \delta(1-\delta)C + (1-\delta)\delta C + (1-\delta)^2 C = C$$
$$c^A(m,a) = \delta^2 C + \delta(1-\delta)C + (1-\delta)\delta(2C+\theta) + (1-\delta)^2 2C = (2-\delta)C + \delta(1-\delta)\theta$$
$$c^A(a,m) = \delta^2 0 + \delta(1-\delta)0 + (1-\delta)\delta D + (1-\delta)^2 D = (1-\delta)D$$
$$c^A(a,a) = \delta^2 0 + \delta(1-\delta)0 + (1-\delta)\delta(C+D) + (1-\delta)^2(C+D) = (1-\delta)(C+D)$$

Priced credit liquidity risk:

$$c^A(m,m) = \omega^2 D + \omega(1-\omega)D + (1-\omega)\omega F + (1-\omega)^2 0 = \omega(1-\omega)F + \omega D$$
$$c^A(m,a) = \omega^2 D + \omega(1-\omega)D + (1-\omega)\omega F + (1-\omega)^2 F = (1-\omega)F + \omega D$$
$$c^A(a,m) = \omega^2 D + \omega(1-\omega)D + (1-\omega)\omega D + (1-\omega)^2 D = D$$
$$c^A(a,a) = \omega^2 D + \omega(1-\omega)D + (1-\omega)\omega D + (1-\omega)^2 D = D$$

Credit risk:

$$c^A(m,m) = \delta^2 0 + \delta(1-\delta)0 + (1-\delta)\delta 0 + (1-\delta)^2 0 = 0$$
$$c^A(m,a) = \delta^2 0 + \delta(1-\delta)0 + (1-\delta)\delta(F+\theta) + (1-\delta)^2 F = (1-\delta)F + \delta(1-\delta)\theta$$
$$c^A(a,m) = \delta^2 0 + \delta(1-\delta)0 + (1-\delta)\delta D + (1-\delta)^2 D = (1-\delta)D$$
$$c^A(a,a) = \delta^2 0 + \delta(1-\delta)0 + (1-\delta)\delta D + (1-\delta)^2 D = (1-\delta)D$$

Repeated Collateralized Credit Game

When is it better to play "always delay" given that the opponent plays "trigger"?

$$\frac{C}{1-\beta} > D + \frac{\beta(C+D)}{1-\beta} \Rightarrow C > \frac{1}{1-\beta}D \Rightarrow \beta < 1 - \frac{D}{C}$$

When is it better to play "always delay" given that the opponent also plays "always delay"?

$$2C + \frac{\beta(C+D)}{1-\beta} > \frac{(C+D)}{1-\beta} \Rightarrow 2C + \frac{\beta-1}{1-\beta}(C+D) > 0 \Rightarrow C > D$$

When is (*trigger, trigger*) the risk-dominant equilibrium?

$$\frac{C}{1-\beta} + 2C + \frac{\beta(C+D)}{1-\beta} < D + \frac{\beta(C+D)}{1-\beta} + \frac{C+D}{1-\beta} \Rightarrow$$
$$2C < D + \frac{D}{1-\beta} = \frac{2-\beta}{1-\beta}D \Rightarrow \beta > \frac{2(C-D)}{2C-D}$$
$$\frac{1}{2}\left(F + \frac{\beta D}{1-\beta}\right) < \frac{D}{1-\beta} \Rightarrow F < \frac{2-\beta}{1-\beta}D$$

References

Aumann, R. 1974. "Subjectivity and Correlation in Randomized Strategies." Journal of Mathematical Economics 1, no. 1 (March): 67-96.

Bech, M., and R. Garratt. 2003. "The Intraday Liquidity Management Game." Journal of Economic Theory 109, no. 2 (April): 198-219.

———. 2006. "Illiquidity in the Interbank Payment System Following Wide-Scale Disruptions." Federal Reserve Bank of New York Staff Reports, no. 239, March.

Bech, M., and B. Hobijn. 2007. "Technology Diffusion within Central Banking: The Case of Real-Time Gross Settlement." International Journal of Central Banking 3, no. 3 (September): 147-81.

Bech, M. L., C. Preisig, and K. Soramäki. 2008. "Global Trends in Large-Value Payments." Federal Reserve Bank of New York Economic Policy Review 14, no.2 (September): 59-81.

Berger, A. N., D. Hancock, and J. C. Marquardt. 1996. "A Framework for Analyzing Efficiency, Risks, Costs, and Innovations in the Payments System." Journal of Money, Credit, and Banking 28, no. 4, part 2 (November): 696-732.

Board of Governors of the Federal Reserve System. 2006. "Consultation Paper on Intraday Liquidity Management and Payments System Risk Policy." Federal Register 71, no. 119: 35679-87. Docket no. OP-1257.

Danmarks Nationalbank. 2003. "SCP: Scandinavian Cash Pool." Monetary Review, fourth quarter: 23-5.

Farrell, J., and M. Rabin. 1996. "Cheap Talk." Journal of Economic Perspectives 10, no. 3 (summer): 103-18.

Freeman, S. 1996. "The Payment System, Liquidity, and Rediscounting." American Economic Review 86, no. 5 (December): 1126-38.

Furfine, C. H., and J. Stehm. 1998. "Analyzing Alternative Intraday Credit Policies in Real-Time Gross Settlement Systems." Journal of Money, Credit, and Banking 30, no. 4 (November): 832-48.

Greenspan, A. 1996. "Remarks on Evolving Payment System Issues." Journal of Money, Credit, and Banking 28, no. 4, part 2 (November): 689-95.

Heller, D., T. Nellen, and A. Sturm. 2000. "The Swiss Interbank Clearing System." Swiss National Bank, Payment System Subsection, June. Available at <http://www.sic.ch/dl_tkicch_sicartikelsnb.pdf>.

Kahn, C. M., J. McAndrews, and W. Roberds. 2003. "Settlement Risk under Gross and Net Settlement." Journal of Money, Credit, and Banking 35, no. 4 (August): 591-608.

Martin, A. 2004. "Optimal Pricing of Intra-Day Liquidity." Journal of Monetary Economics 51, no. 2 (March): 401-24.

Martin, A., and J. McAndrews. 2008. "An Economic Analysis of Liquidity-Saving Mechanisms." Federal Reserve Bank of New York Economic Policy Review 14, no. 2, (September): 25-39.

McAndrews, J., and S. Rajan. 2000. "The Timing and Funding of Fedwire Funds Transfers." Federal Reserve Bank of New York Economic Policy Review 6, no. 2 (July): 17-32.

Mills, D. C., and T. D. Nesmith. 2008. "Risk and Concentration in Payment and Securities Settlement Systems." Journal of Monetary Economics 55, no. 3 (April): 542-53.

Reserve Bank of New Zealand. 2006. "Reform of the Reserve Bank of New Zealand's Liquidity Management Operations." June.

Skyrms, B. 2004. The Stag Hunt and the Evolution of Social Structure. Cambridge: Cambridge University Press.

Soramäki, K., M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler. 2007. "The Topology of Interbank Payment Flows." Physica A 379, no. 1 (June): 317-33.

Antoine Martin and James McAndrews

An Economic Analysis of Liquidity-Saving Mechanisms

- Liquidity-saving mechanisms (LSMs) are queuing arrangements for payments that operate alongside traditional real-time gross settlement (RTGS) systems.
- LSMs allow banks to condition the release of queued payments on the receipt of offsetting or partially offsetting payments; as a result, banks are less inclined to delay the sending of payments.
- An analysis of LSMs finds that these mechanisms typically perform better than pure RTGS systems when it comes to settling payments early.
- RTGS systems can sometimes be preferable to LSMs, such as when many banks that send payments early in RTGS choose to queue their payments when an LSM is available.

1. Introduction

Large-value payments systems, used by banks to settle financial and commercial transactions, play a key role in the financial system. The importance of these payments systems can be illustrated by the large amounts they settle. Every year in the United States, the systems process value equal to approximately 100 times GDP.

Innovations in the design of large-value payments systems have led to many improvements in their operations. For example, over the last twenty years, many countries have adopted real-time gross settlement (RTGS) systems for their large-value payments. In an RTGS system, each payment is settled individually, on a gross basis, at the time the payment is sent. RTGS systems offer many advantages—for instance, they limit the risk exposure of payments system participants and allow for rapid final settlement of payments during the day. However, RTGS systems require large amounts of central bank balances to function smoothly.

More recent innovations have occurred in the design and implementation of various liquidity-saving mechanisms (LSMs) that are used in conjunction with RTGS systems.[1] An LSM gives participants in the payments system an additional option not offered by RTGS alone: A payment can be put into a queue and then released from the queue if some prespecified event occurs. Such mechanisms can reduce the amount of central bank balances necessary to operate the system smoothly as well as quicken the settlement of payments.

Antoine Martin is a research officer and James McAndrews a senior vice president at the Federal Reserve Bank of New York.
<antoine.martin@ny.frb.org>
<james.mcandrews@ny.frb.org>

The authors thank Enghin Atalay for excellent research assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

As we describe in detail below, an LSM allows banks to send payments conditional on the receipt of payments, and it can accommodate some netting of payments.

Over the past decade, researchers have been able to simulate the performance of various LSMs. In most of these simulations, the researcher makes assumptions about the behavior of the parties in the system and measures various consequences of the assumed behavior. This approach has great potential to yield useful answers to a number of questions.

This article outlines a different approach to the study of LSMs in a payments system. It examines a theoretical model of the behavior of parties, which for simplicity we refer to as *banks*. Each bank has particular motivations and constraints; as a result, its behavior can be determined as an equilibrium outcome in response to the incentives it faces. The theoretical approach has the advantage of allowing banks' reactions to alternative payments system designs to be determined within a theoretical model, rather than be assumed by the researcher. This approach also allows outcomes to be compared consistently across a number of designs.

Every year in the United States, [large-value payments] systems process value equal to approximately 100 times GDP.

Innovations in LSMs are numerous and, in some cases, quite complex. A 2005 report by the Bank for International Settlements, "New Developments in Large-Value Payment Systems," is an authoritative source on the many alternative systems introduced until that time. The research presented here generalizes the various types of LSMs by placing their essential characteristics into one of two categories: *balance reactive* or *receipt reactive.*

A simplified description of an LSM's operation is as follows. A bank wishes to make a payment and has a choice of when to submit it to the payments system. Upon submitting the payment, the bank has a second choice to make: It can either submit the payment to a central queue (the LSM part of the payments system, which will be called *the LSM channel*), or attempt to settle the payment at the time of submission (the RTGS part of the payments system, which will be called *the RTGS channel*). If the bank submits the payment to the LSM channel, that payment will settle only when certain conditions have been met. If the bank attempts to settle via the RTGS channel and it has sufficient funds available, the payment will settle immediately.

One condition that might trigger the settlement of a payment (and is common to both types of LSMs) occurs when the request in the LSM channel is made in the presence of an offsetting payment in the queue of the bank to which the payment is to be made. If the two payments offset, then both can be released by the LSM.

A condition of another sort is determined by the type of LSM used. In a balance-reactive LSM, a bank has the choice of reserving some level of its account balances for the settlement of payments submitted via the RTGS channel. If the bank's balances exceed the predetermined reserve level, then that bank's payments that were previously placed in the LSM can be released. In a receipt-reactive LSM, a queued payment can be released against recent receipts of the bank (rather than against its accumulated balances) and at any level of balances for the bank.

The research reported here reveals two sources of potential value to implementing an LSM. First, an LSM gives a bank a new tool not available in an RTGS system: It gives the bank the option of making a payment *conditional* on the receipt of another payment. In RTGS, banks can find themselves in the positions of those two overly polite gentlemen in the old vaudeville routine, who repeatedly say to each another: "After you, Alphonse." "No, after you, Gaston." [2] That is, so long as central bank balances are costly, each bank would prefer to have its counterparty make a payment first.

In an RTGS system, there is no way to condition the settlement of a payment upon the future receipt of a payment; however, this course of action is possible with an LSM. The LSM essentially allows the two banks to solve the precedence problem that bedevils them in RTGS.[3] The ability to condition payments on the receipt of offsetting payments provides banks with some insurance against the risk of having to borrow funds from the central bank (see, for example, Mills and Nesmith [2008]). In turn, this means that banks are more willing to submit payments to the LSM earlier than they would to the RTGS, which usually, but not always, has beneficial effects.

[1] In this article, *liquidity-saving mechanism* refers to a mechanism intended to economize on the use of central bank reserves. These reserves can typically be obtained intraday from the central bank either against collateral or for a small fee. A liquidity-saving mechanism can allow payments to be settled with fewer central bank reserves. As we discuss, under some circumstances, an LSM can indeed economize on the use of central bank balances as well as lead banks to submit payments earlier to the payments system. In general, this outcome can be defined as making the payments system more "liquid."

[2] The American version of this routine is based on a comic strip by Frederick Burr Opper, "Alphonse and Gaston," which was popular in the early 1900s and pokes fun at exaggerated politeness.

[3] Note that an internal queue, rather than an LSM, could allow a bank to send a payment conditional on receiving another payment. However, investment in an internal queue has benefits to other banks that cannot be appropriated by the investing bank. In addition, an internal queue does not solve the precedence problem; the bank with the internal queue arranges for its payments to follow the receipt from other banks. As a result, RTGS systems that have internal queues can still suffer from excessive delay in payments.

The second source of potential value to implementing an LSM comes from the offsetting of payments within the queue, which reduces the need for central bank balances. The benefit arising from payments offsetting in the queue is well understood. An important contribution of our research is to show that even when no such offsetting occurs, the first potential benefit described above means that an LSM can improve welfare.

Several authors have examined theoretical behavior in RTGS systems. Angelini (1998, 2000) considers the behavior of banks in an RTGS system in which they face delayed costs for payments as well as costly borrowing of funds. He shows that the natural payment timing equilibria of RTGS systems (in the absence of LSMs) involve excessive delay of payments, as banks do not properly internalize the benefits to other banks from the receipt of funds. Bech and Garratt (2003) carefully specify a game-theoretic environment in which they find that RTGS systems can be characterized by multiple equilibria, some of which can involve excessive delay. Roberds (1999) compares gross and net payments systems with systems offering an LSM. Examining the incentives banks have to engage in more risk-taking behavior in the different systems, Roberds finds that, under certain circumstances, the risk profiles of LSMs and net systems are identical. Mills and Nesmith (2008) also study the impact of incentives on banks' payment patterns. McAndrews and Trundle (2001) and the Bank for International Settlements (2005) provide extensive descriptive material on LSMs.

The research presented here generalizes the various types of [liquidity-saving mechanisms] by placing their essential characteristics into one of two categories: balance reactive or receipt reactive.

Willison (2005) examines the behavior of banks in an LSM, and this work is most similar to ours. He models agents as having an ordering of payment priority, which is similar in spirit to our assumption that some banks' payments are time sensitive. Willison models the extension of credit from the central bank as an ex ante amount to be borrowed by banks, while in our study the credit is tapped ex post, depending on a bank's per-period balance. Our model extends Willison's analysis in two dimensions that prove to be important: We consider a wider array of LSMs and, crucially, we allow for liquidity shocks, which we define as shocks to the level of a bank's balances on account.

Our study proceeds as follows. In the next section, we describe the environment in place when banks decide on their payment submission strategies. In subsequent sections, we consider various scenarios, including banks' behavior in an RTGS system, a balance-reactive LSM, and a receipt-reactive LSM.

2. The Environment

This section describes the economic environment in which banks operate. It specifies the economic agents, which we refer to as *banks*; the banks' objectives; and other factors that influence banks' decisions. We study a simple payments system. While some features of our model are unrealistic because of their simplicity, we believe that the model captures essential economic frictions that affect banks' behavior in the payments system. The simplicity of the model allows us to obtain explicit results and to provide transparent intuition for these results.[4]

In our model, the essential features of an LSM are as follows. Payment orders are put in a queue, and the release of a payment order from the queue will occur when some conditions have been satisfied. Here, the conditions are that a bank has sufficient funds and that an offsetting payment has been received. While actual LSMs can be considerably more complicated, they all share those basic, essential features. By capturing the essential features shared by LSMs, our model can describe the basic economic mechanisms associated with these systems in a tractable way.

To incorporate additional features of queues, we would require a more complicated model that may be difficult to solve explicitly. Those additional features could include: 1) limits on the amount of central bank balances that can be committed to a particular bank or set of banks, 2) a time before or after which a payment should be sent, and 3) different payment priorities that can change the ordering of payments submitted to the queue away from first in, first out. Also, queues may or may not be transparent. If a queue is transparent, banks can see pending payments in their favor.[5] In addition to the tractability issue, it is not clear that adding these features would modify our key insights into the potential benefits of the LSM's essential feature: to allow conditionality in the settlement of some payments.

[4] Our analysis summarizes our more technical paper (Martin and McAndrews 2008).

[5] In our model, queues are opaque—that is, banks cannot see particular payments in other banks' queues. However, banks are forward looking and, in equilibrium, they will expect banks of different types to have (or have not) submitted some payments to the LSM, and these expectations will be fulfilled.

Each day, the payments system operates during two periods: the morning and the afternoon. A large number of banks are involved, and each bank must send payments to other banks and will receive payments from other banks. For simplicity, all payments are of the same size.

We assume that banks are risk neutral. Our definition of risk neutrality is that a bank would be willing to pay up to 50 cents for the opportunity to participate in a lottery that promises one dollar with probability 1/2 and nothing with probability 1/2. We believe that our results would extend to more complicated economies in which there are many periods and payments of different sizes. Banks have rational expectations about the probability of receiving a morning payment. In other words, banks are able to calculate the correct probability of receiving a payment in the morning.

Banks must choose whether to send their payments to another bank in the morning or in the afternoon. Three factors influence this decision. First, banks face a cost if they must borrow funds at the central bank. Second, some banks must make time-sensitive payments. Third, banks may receive a positive liquidity shock, a negative liquidity shock, or no shock. Each factor is explained in more detail below.

If a time-sensitive payment is not settled until the afternoon, the sending bank incurs a delay cost. This cost could arise because delay creates a bad reputation for the bank in the eyes of its customers or counterparties.

Banks face a cost if they must borrow funds from the central bank. All banks start the day with zero reserves at the central bank and, because of the symmetry assumed by the model, end the day with zero reserves as well. Banks are not allowed to hold negative reserves at the end of the morning period, but they can borrow reserves from the central bank at a cost.[6] For example, a bank that sends a payment but does not receive a payment in the morning period would have to borrow reserves. The cost of borrowing can represent either an explicit fee imposed by the central bank, as in the United States, or the implicit cost of collateral, as in many countries where central bank balances are available for free but only against collateral. Evidence provided by the Bank for International Settlements (2005) suggests that the cost of central bank balances can influence banks' payment behavior.

Some banks must make time-sensitive payments. If a time-sensitive payment is not settled until the afternoon, the sending bank incurs a delay cost. This cost could arise because delay creates a bad reputation for the bank in the eyes of its customers or counterparties. The behavior of banks suggests that they perceive some payments as being more time sensitive than others. For example, payments made to close a real estate transaction may be more time sensitive than payments for previous deliveries of supplies to a manufacturer.[7] However, delaying non-time-critical payments until the afternoon is costless.

Banks may receive a positive liquidity shock, a negative liquidity shock, or no liquidity shock. The liquidity shock is modeled by assuming that each bank must send a second payment to (and will receive a payment from) a settlement institution. For the purpose of this article, U.S. settlement institutions are represented jointly by CLS Bank, CHIPS, and DTC.[8] The settlement institution is intended to capture other payment or settlement systems whose pay-ins and payouts are made using the central bank's payments system. This rearrangement of balances, which we refer to as a *shock*, introduces another source of heterogeneity between banks and leads to a more diverse set of strategic interactions.

In the model, banks find out when they must send a payment to the settlement institutions and when they will receive an offsetting payment from the settlement institutions *before* they decide whether to send their payment to other banks. We assume that payments to settlement systems cannot be delayed. We say that a bank that receives a payment from the settlement institutions in the morning and does not have to make an offsetting payment until the afternoon has received a positive liquidity shock, as its account balances are boosted throughout the day prior to its late offsetting payment. Conversely, a bank that receives a payment from the settlement institutions in the afternoon but must make a payment in the

[6] We assume that excess reserves have zero return. In our model, banks cannot trade with one another the funds needed to make payments. This assumption is a common one in the literature and reflects the empirical observation that markets for intraday funds are either nonexistent or operate only in extraordinary circumstances.

[7] Armantier, Arnold, and McAndrews (2008) find that Fedwire payments tend to settle earlier in the day on days with higher values of customer payments. This behavior could reflect higher costs of delay for customer payments relative to interbank payments.

[8] CLS Bank is a payment-versus-payment settlement system that settles foreign exchange transactions in fifteen currencies. CLS Bank is operated by CLS Bank International, a bank-owned Edge Act corporation incorporated in the United States. CHIPS (the Clearing House Interbank Payments System) is a private, large-value U.S. dollar payments system owned and operated by the Clearing House Payments Company. DTC (the Depository Trust Company) is a securities settlement system that settles the majority of U.S. corporate securities and commercial paper. DTC is a wholly owned subsidiary of Depository Trust and Clearing Corporation.

morning receives a negative shock, as its balances are depleted early. When the payment from the settlement institutions is received in the same period in which the bank's outgoing payment to the settlement institutions must be sent, the bank experiences no liquidity shock, as the offsetting payments have a neutral effect on its balances.

We assume that payments sent among banks are larger than payments sent to and received from settlement institutions. This implies that while banks may face large liquidity shocks, these shocks are relatively small compared with the banks' expected payment activity. In addition, payments made through settlement institutions occur either at the very beginning or very end of the business day. Thus, a bank that begins its business day with a negative liquidity shock, then receives a payment from another bank in the morning, and then delays sending a payment to another bank will have a positive balance at the end of the morning period. In contrast, a bank that receives a positive liquidity shock at the start of the business day, then proceeds to send a payment to another bank in the morning, and then does not receive a payment from another bank during the morning period will end the morning with a negative balance. The latter bank must borrow from the central bank, incurring a borrowing cost in the process.

Each bank can have one of six possible profiles: It may or may not have to make a time-sensitive payment, and it may experience a positive, negative, or no liquidity shock.

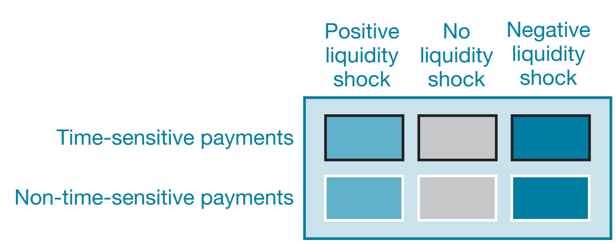
With all those factors explained, we can now trace out the intuition of the model. By itself, the cost of borrowing gives banks an incentive to send payments at the same time as other banks (either in the morning or in the afternoon). For example, absent liquidity shocks and delay costs, banks would be willing to send their payments in the morning (or afternoon) if they knew that all other banks would send theirs in the morning (or afternoon). That way, they would never have to borrow from the central bank. However, the cost of delay provides an incentive for banks to make payments early. For example, if the cost of delay is sufficiently high, a bank will prefer to send a time-critical payment early even if all other banks delay their payments. This will happen if the cost of delay is greater than the cost of borrowing. A cost of borrowing and a high cost of delay will therefore result in banks preferring to send payments all at the same time in the morning. Introducing the liquidity shock, together with the borrowing cost, provides an incentive for banks to delay payments. For example, a bank that experiences a negative liquidity shock and that must make a non-time-critical payment prefers to send it in the afternoon even if all other banks pay early. Indeed, in that case, the bank incurs no cost if it delays its payment, but it receives the benefit of not having to borrow from the central bank.

In summary, each bank can have one of six possible profiles: It may or may not have to make a time-sensitive payment, and may experience a positive, negative, or no liquidity shock. Exhibit 1 displays the six profiles in the form of a color key that will describe the equilibria of our economy. We assign the color light blue to banks that receive a positive liquidity shock, gray to banks that receive no liquidity shock, and dark blue to banks that receive a negative liquidity shock. A black border indicates banks that must make time-sensitive payments, while a white border indicates banks that make non-time-sensitive payments.

The remainder of this article examines different settlement systems for the payments system described above. We consider real-time gross settlement alone and an RTGS system supplemented by two types of liquidity-saving mechanisms. In each case, we describe the timing of payments predicted by the model in equilibrium. In equilibrium, every bank chooses a submission time for its direct payments that minimizes its delay and borrowing costs given the payment submission strategies of all other banks in the payments system. We also compare the desirability of different settlement systems according to our model.

Exhibit 1

Banks Using Large-Value Payments Systems



Note: The key applies to Charts 1-10. It describes six types of banks: those sending time-sensitive payments that experience a positive liquidity shock, no liquidity shock, or a negative liquidity shock; and those sending *non*-time-sensitive payments that experience a positive liquidity shock, no liquidity shock, or a negative liquidity shock.

3. Real-Time Gross Settlement

This section studies our model economy in terms of a real-time gross settlement system. With RTGS, payments are settled on an individual basis at the time they are sent. We assume that banks have sufficient borrowing capacity at the central bank to make payments, even if they do not receive a prior payment and/or have experienced a negative liquidity shock. Banks have the choice of sending payments in the morning or delaying them until the afternoon. They compare the expected cost of each option and choose the least expensive timing strategy. As shown in Exhibit 2, there is a pattern to how events unfold during the day. First, banks learn if they must send a time-critical payment and if they have a positive, negative, or no liquidity shock. Next, early payments to settlement institutions are made. Afterward, banks decide whether to send their direct payments in the morning period or to delay sending them until the afternoon period. At the end of the morning period, banks incur a delay cost if they have delayed a time-sensitive payment or a borrowing cost if they must borrow from the central bank. In the afternoon, delayed payments to other banks are made, as are late payments to settlement institutions.

Liquidity shocks force some banks to start the day with negative balances. If those banks do not have time-sensitive payments to make, then they can be counted on to delay their direct payments, since otherwise even the receipt of an offsetting payment would mean that they must borrow. So banks with negative liquidity shocks and non-time-sensitive payments will surely delay their payments until the afternoon period. Following domino-type logic, banks—even those that enjoyed a positive liquidity shock—will delay all other non-time-sensitive payments until the afternoon, as they cannot count on offsetting payments during the morning period. In sum, in a world with liquidity shocks, we expect that all non-time-sensitive payments will always be delayed.

We present the equilibria associated with RTGS in Charts 1-4. The dark gray bars indicate payments sent to settlement institutions. These payments are made either early in the morning or late in the afternoon. Payments to other banks are indicated in colors that correspond to a specific type of sender as described in Exhibit 1. Payments sent by banks that have experienced a negative liquidity shock are indicated in dark blue, those sent by banks that have received no liquidity shock are in gray, and those sent by banks that have experienced a positive liquidity shock are in light blue. A dark border corresponds to time-sensitive payments while a white border corresponds to non-time-sensitive payments. Given this color scheme, the behavior of each of the six types of banks can be seen within the bar, indicating whether a bank of a certain type sent its payments in the morning or the afternoon.

Charts 1-4 show how the pattern of payments changes as the cost of delay decreases. In Chart 1, the cost of delay is high, so all time-sensitive payments are sent early. If the cost of delay decreases a little (Chart 2), banks that have experienced a negative liquidity shock will choose to delay their time-sensitive payments. These banks have the highest expected cost of borrowing from the central bank, so if the cost of delay is not too high, banks will aim to reduce their need to borrow from the central bank by delaying payments. If the cost of delay decreases still further (Chart 3), only banks that have experienced a positive liquidity shock will send their time-critical payments early. Finally, if the cost of delay is sufficiently low (Chart 4), all banks will delay payments regardless of their liquidity profiles.

Exhibit 2
Real-Time Gross Settlement Timeline

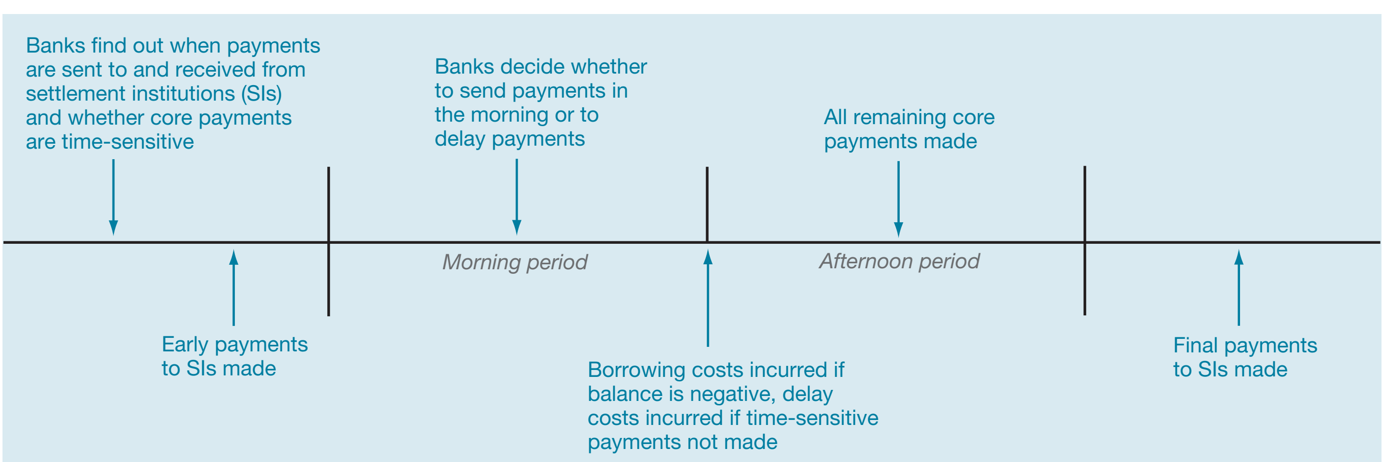
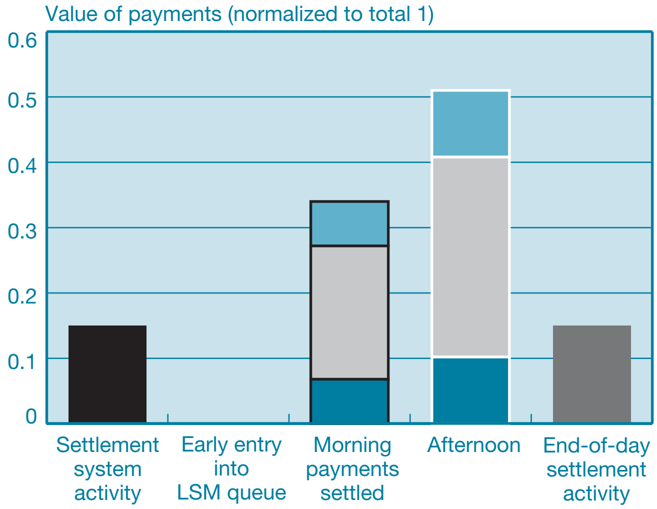


Chart 1

Real-Time Gross Settlement with High Costs of Delay

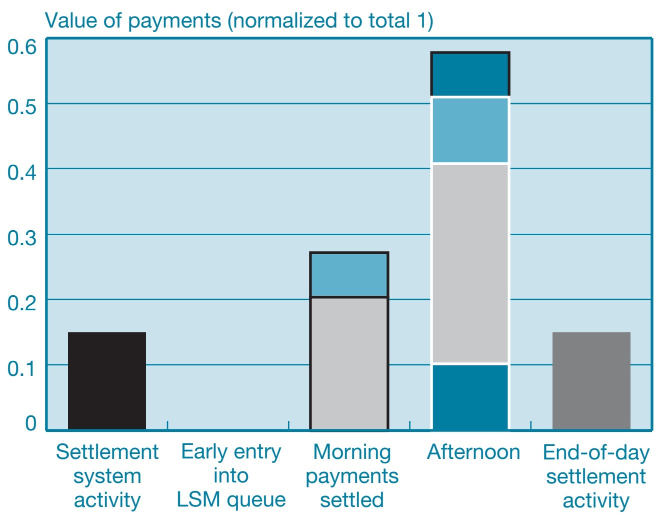


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

Chart 2

Real-Time Gross Settlement with Medium Costs of Delay

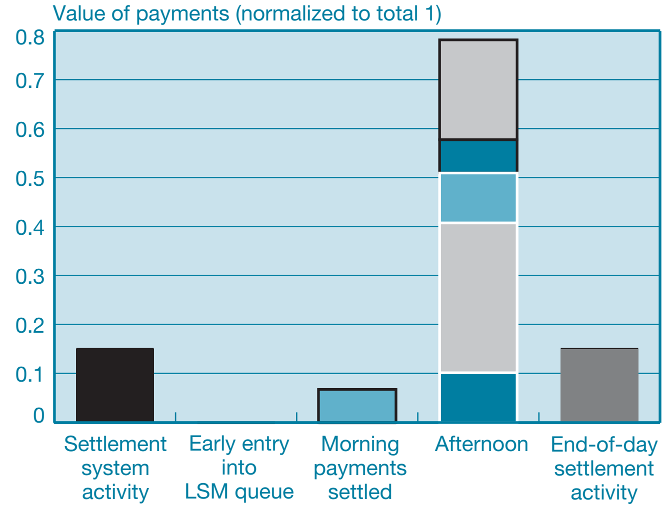


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

Chart 3

Real-Time Gross Settlement with Lower Costs of Delay

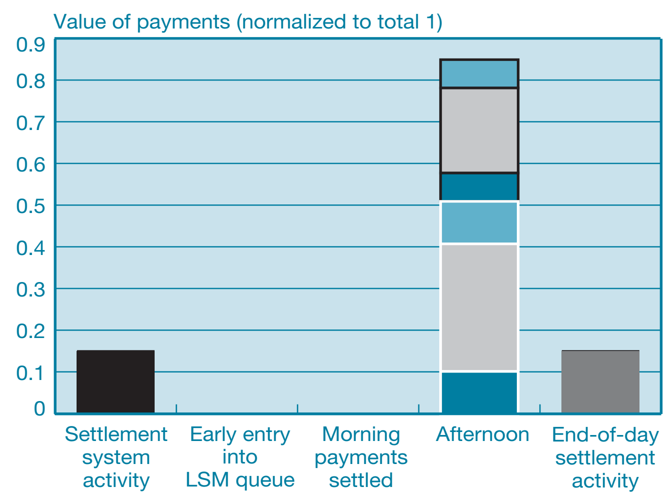


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

Chart 4

Real-Time Gross Settlement with Lowest Costs of Delay



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. LSM is liquidity-saving mechanism.

In general, multiple equilibria can exist. Each equilibrium described in Charts 1-4 may arise for the same set of parameters. There can be multiple equilibria because a bank's incentives to pay early depend on the behavior of other banks.[9] If all time-sensitive payments are sent early, then the probability of receiving a payment in the morning is high, implying that the probability of having to borrow from the central bank is low. If the expected cost of borrowing from the central bank is low, banks have a strong incentive to make their time-sensitive payments early to avoid the cost of delay.

[9] By multiple equilibria, we mean that, for the same set of parameters of the model, different sets of beliefs are consistent with an equilibrium.

Conversely, if all time-sensitive payments are delayed, then the probability of receiving a payment in the morning is low (in this case, zero). It follows that banks will have to borrow from the central bank if they make payments early. With a high expected cost of borrowing, agents have only weak incentives to send payments early, even if this would allow them to avoid the delay cost.

We can compare the welfare associated with different equilibria. Our measure of welfare is the expected utility of the average of the six possible types of banks. To measure this average payoff, we consider what would happen if all banks could meet before the beginning of the day. Before determining their respective liquidity shocks and whether they have to make a time-sensitive payment, all banks are identical and thus have the same preferences for payment patterns.

Banks have the choice of sending payments in the morning or delaying them until the afternoon. They compare the expected cost of each option and choose the least expensive timing strategy.

When several equilibria coexist, welfare is highest in the equilibrium in which the greatest number of payments is sent in the early period.[10] Prior to knowing its particular situation—that is, whether it has a positive or negative balance in the morning and whether it has a time-sensitive payment—the average bank would prefer to make its payments at the same time as all other banks. This simultaneous payment pattern reduces banks' expected borrowing needs. Given the level of coordination, the average bank would prefer to make simultaneous payments in the morning, thereby reducing delay costs. For all equilibria, our measure of welfare decreases with the cost of delay and with the cost of borrowing.

We can illustrate the role liquidity shocks play in our analysis by looking at how equilibrium outcomes change when these shocks are suppressed. This would correspond to the case where all banks make their pay-ins to the settlement institutions and receive their payouts from the settlement institutions early. Absent liquidity shocks, all banks start the morning period with zero balances and the only difference among them is whether or not their payments are time sensitive. Three types of equilibria are possible: 1) all payments are sent early, 2) only time-sensitive payments are sent early, or 3) all payments are delayed.

Consider the equilibrium in which all payments are sent in the morning period. This is an equilibrium, regardless of the magnitude of borrowing and delay costs, because if all other banks are sending their payments in the morning, an individual bank is at least as well off sending its payments in the morning as delaying them until the afternoon. If the bank has a time-sensitive payment, the bank is strictly better off sending it in the morning; if the bank does not, it is equally well off sending its payments in the morning to coincide with its receipts. This type of equilibrium does not exist when some banks have experienced a negative liquidity shock.

The costs of delay and borrowing matter for the other two equilibria. If the cost of delaying time-sensitive payments is low, if the borrowing cost is high, and if banks without a time-sensitive payment plan to pay in the afternoon, then banks that must make a time-sensitive payment would prefer to delay it. Indeed, banks want to avoid high borrowing costs, even if they must incur relatively low delay costs. Likewise, if the cost of delay is high relative to the cost of borrowing, then banks with time-sensitive payments would prefer to make them early, even if banks without time-sensitive payments choose to delay making theirs.

4. Balance-Reactive Liquidity-Saving Mechanism

This section studies our model economy under a particular design of a liquidity-saving mechanism. An LSM can be thought of as a queue, into which banks enter their payments as an alternative to sending or delaying payments outright. A queued payment is released in the morning if the bank that queued the payment receives an offsetting payment or if a group of queued payments offset multilaterally. Otherwise, the payment is released from the queue in the afternoon.

The probability of a payment being released from the queue depends on the underlying pattern of payments. Exhibits 3 and 4 illustrate two cases. In the exhibits, the Xs denote banks and the arrows denote a payment that must be made from one bank to another. In Exhibit 3, all payments form a unique cycle that links all banks. Here, payments in the queue cannot offset multilaterally unless all payments are queued. In Exhibit 4, payments form cycles of length 2 that link banks in pairs. This pattern of payments maximizes the probability that payments in the queue offset bilaterally. The case of a unique cycle, illustrated in Exhibit 3, is particularly interesting because it allows us to disentangle two different roles played by LSMs. One is that the LSM allows agents to condition the release of

[10] Note that the pattern of payments such that all payments, including time-sensitive ones, are made early may not provide maximum welfare. This does not contradict the result stated above, as this pattern of payments is not an equilibrium.

Exhibit 3
A Unique Cycle

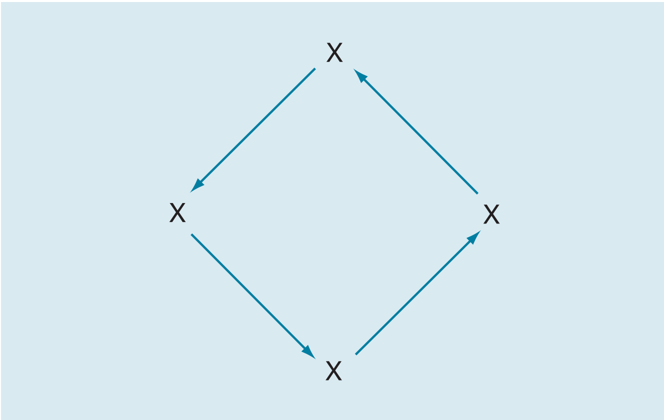
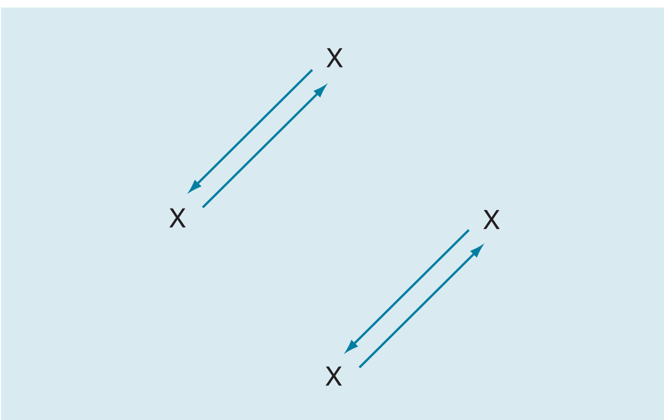


Exhibit 4
Cycles of Length 2

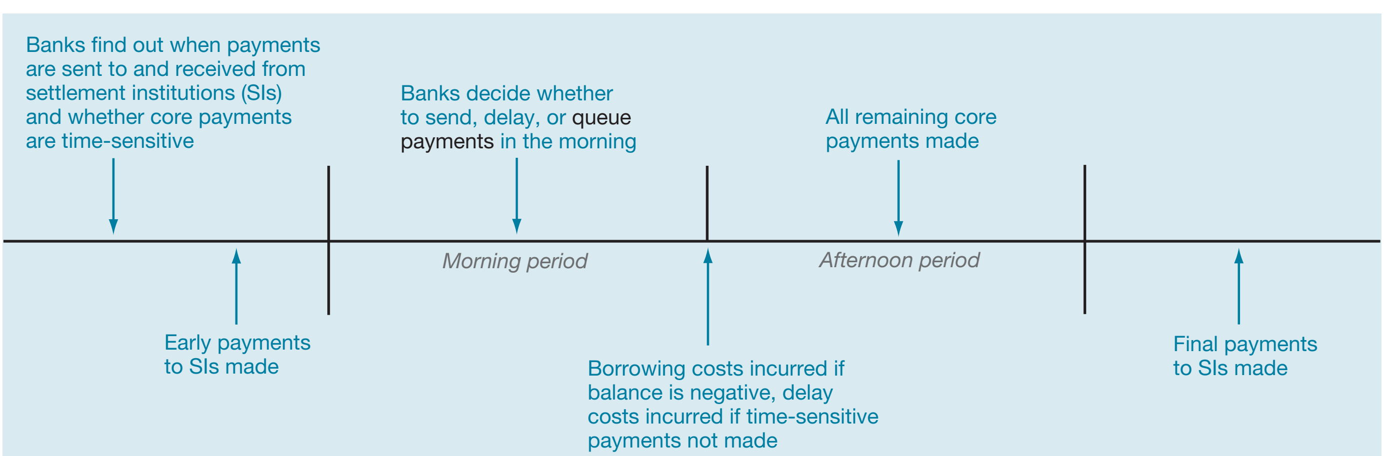


a payment on the receipt of an offsetting payment. A different, though complementary, role is that an LSM can permit bilateral or multilateral netting. In the case of a unique cycle, however, no netting occurs unless all payments are in the queue. Here we can study, in isolation, the role of the LSM in allowing banks to make payments conditional on the receipt of offsetting payments.

Exhibit 5 shows the timing of payments in a system in which banks are able to submit payments to the LSM. The difference between this and pure RTGS is that now banks have the option of queuing their payments. In other words, a bank can condition its sending of a queued payment on the receipt of an offsetting payment, in addition to sending or delaying the payments outright. The decision to queue a payment is made after the bank's liquidity shock is known, allowing the bank to take into account the shock's effect on the level of the balance before deciding whether to queue a payment. Thus, we call this a *balance-reactive* LSM. Later, we will consider an LSM in which the release of a payment from the queue does not depend on a liquidity shock.

Charts 5-8 illustrate the equilibria arising with a balance-reactive LSM. The visual coding in these charts is the same as in Charts 1-4. That is, payments from banks that experience a positive liquidity shock are indicated in light blue, those from banks that experience no liquidity shock are in gray, and those from banks that experience a negative liquidity shock are in dark blue. A black border indicates time-sensitive payments while a white border indicates payments that are not time sensitive. In addition, dark shading indicates payments entered in the queue, which we emphasize by adding a queue-submission stage. Payments released from the queue have a light shading, while payments sent outright or delayed have no

Exhibit 5
Balance-Reactive Liquidity-Saving Mechanism Timeline



shading. Some payments from the queue will be released in the afternoon unless all payments are queued.

In Chart 5, all payments are queued. This happens if both the cost of delay and the liquidity shocks are not too high. In this case, all payments are released in the morning and no delay cost is incurred. Banks with a negative liquidity shock, however, must borrow from the central bank.

If the cost of delay is higher and the liquidity shocks remain moderate, the equilibrium depicted in Chart 6 can occur. In this equilibrium, banks that experience positive liquidity shocks choose to make time-sensitive payments outright during the early period. Banks that experience negative liquidity shocks delay non-time-sensitive payments. All other payments—time-sensitive payments from banks that experience negative or zero liquidity shocks and non-time-sensitive payments from banks that experience positive or zero liquidity shocks—are queued. Some payments are released from the queue in the morning and some are released in the afternoon.

The intuition is as follows. Because the cost of delay is sufficiently high, banks that have experienced a positive liquidity shock prefer to insure themselves against the risk of having to suffer the cost of delay. For this reason, they send payments outright. Banks that receive a negative liquidity shock prefer to delay non-time-sensitive payments. By delaying their own outgoing payments, these banks avoid the need to borrow from the central bank if they receive a payment from another bank in the morning. Note that for parameters in which the equilibria in Charts 5 and 6 coexist, the latter equilibrium is not robust.

If the cost of delay is higher still, the equilibrium presented in Chart 7 can occur. In this equilibrium, all banks that must make time-sensitive payments choose to send them outright. Banks that experience a negative liquidity shock delay non-time-sensitive payments, while those that experience a positive or no liquidity shock queue non-time-sensitive payments.

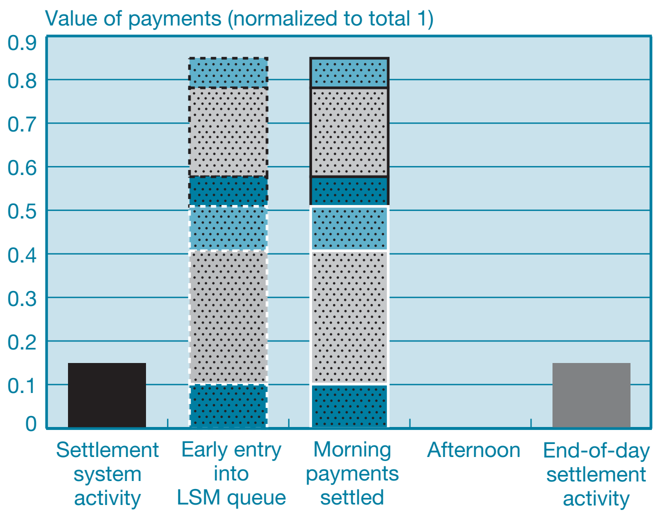
If the cost of delay is neither very high nor very low and the liquidity shocks are large, then the equilibrium in Chart 8 will occur. Because the liquidity shock is so large, banks that receive a negative liquidity shock at the beginning of the day delay both time-sensitive and non-time-sensitive payments. Only banks that receive a positive liquidity shock at the beginning of the day choose to send time-sensitive payments early; all other payments are queued.

The same types of equilibria exist in a unique cycle and in cycles of length 2. However, the parameters under which these equilibria exist in each case can be different. With cycles of length 2, there can be multiple equilibria, as is true with pure RTGS. In contrast, with a unique cycle, only one equilibrium exists for any configuration of parameters.

To understand this difference, note that an LSM allows banks to condition the release of a payment on other banks'

Chart 5

Balance-Reactive Liquidity-Saving Mechanism (LSM) with Low Costs of Delay

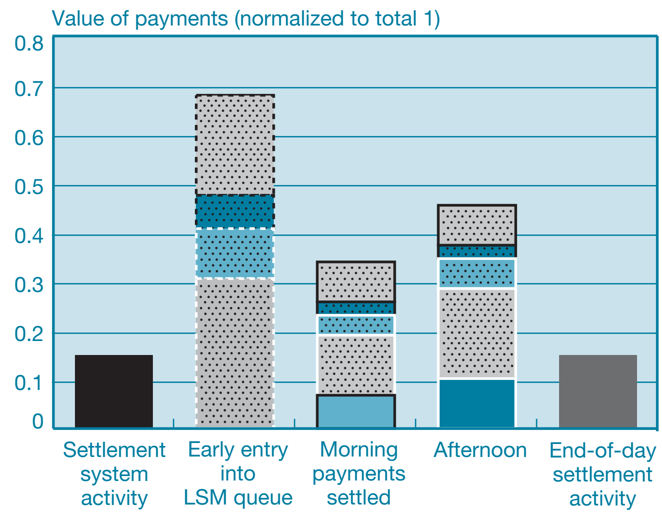


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

Chart 6

Balance-Reactive Liquidity-Saving Mechanism (LSM) with Medium Costs of Delay



Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

actions, which reduces strategic interaction between banks. In RTGS, by contrast, the level of strategic interaction is high: Whether or not a bank chooses to send a payment during the early period depends in large part on that bank's beliefs about the strategy other banks have adopted for the timing of their payments. With the LSM, the bank need not concern itself with such complicated reasoning because it can simply submit a payment to settle automatically if an offsetting payment is received, decoupling the bank's decision to submit the payment to the LSM from the bank's beliefs about the plans of other banks.

Whenever multilateral netting is expected to occur in the queue, a new set of strategic interactions emerges. A bank's incentives to submit payments to the queue depend on the bank's belief about what other banks will do. If many other banks submit their payments to the queue, then more netting will occur and the incentive to queue will be high. In contrast, if only a few other banks submit their payments to the queue, less netting will occur and the incentive to queue will be low. As a result, the possibility of multilateral netting of queued payments reintroduces strategic interaction into banks' submission behavior, leading to multiple equilibria. In contrast, since there is no netting in a unique cycle, multiple equilibria do not occur.

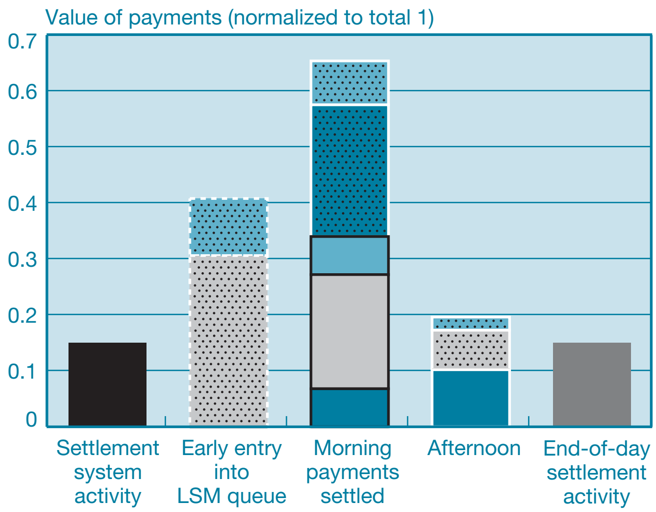
We can compare the welfare of banks if settlement occurs according to RTGS or to a balance-reactive LSM. Welfare is higher with the balance-reactive LSM if the liquidity shocks are small, if the cost of delay is either sufficiently high or sufficiently low, and if the probability of a liquidity shock occurring is large compared with the fraction of time-sensitive payments. In one specific case, a balance-reactive LSM can lead to a loss of some beneficial coordination and result in lower welfare.

When liquidity shocks are small, banks incur a large borrowing cost only if they send a payment early and do not receive an offsetting payment. An LSM increases welfare because it allows banks to insure themselves against this risk because the bank's payment is released from the queue only when an offsetting payment is received. Also, because banks can queue their payments rather than delay them outright, more payments are released in the morning and fewer banks incur a delay cost.

If the cost of delay is sufficiently high, banks send all time-sensitive payments early, regardless of the settlement mechanism. However, banks that do not receive a negative liquidity shock queue their payments when an LSM is available, rather than delay them as they would in RTGS. Hence, more payments are released early and the expected cost of borrowing is reduced for banks that send payments outright. If the cost of delay is sufficiently low, all banks will find it beneficial to queue payments when an LSM is available so that all payments are released early.

Chart 7

Balance-Reactive Liquidity-Saving Mechanism (LSM) with High Costs of Delay

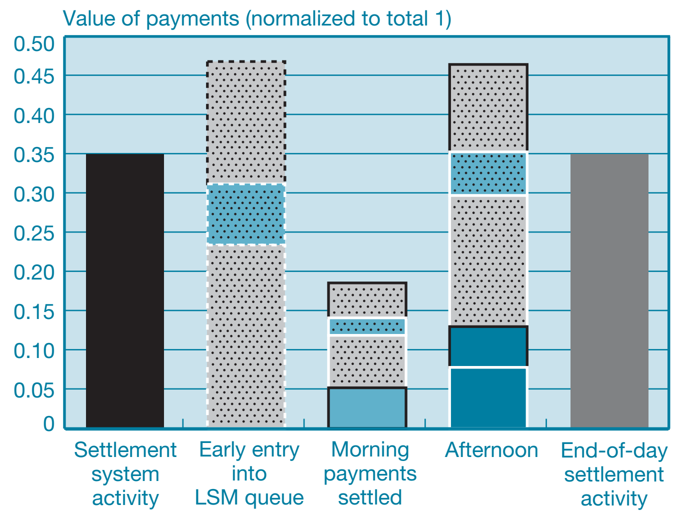


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

Chart 8

Balance-Reactive Liquidity-Saving Mechanism (LSM) with Large Liquidity Shock

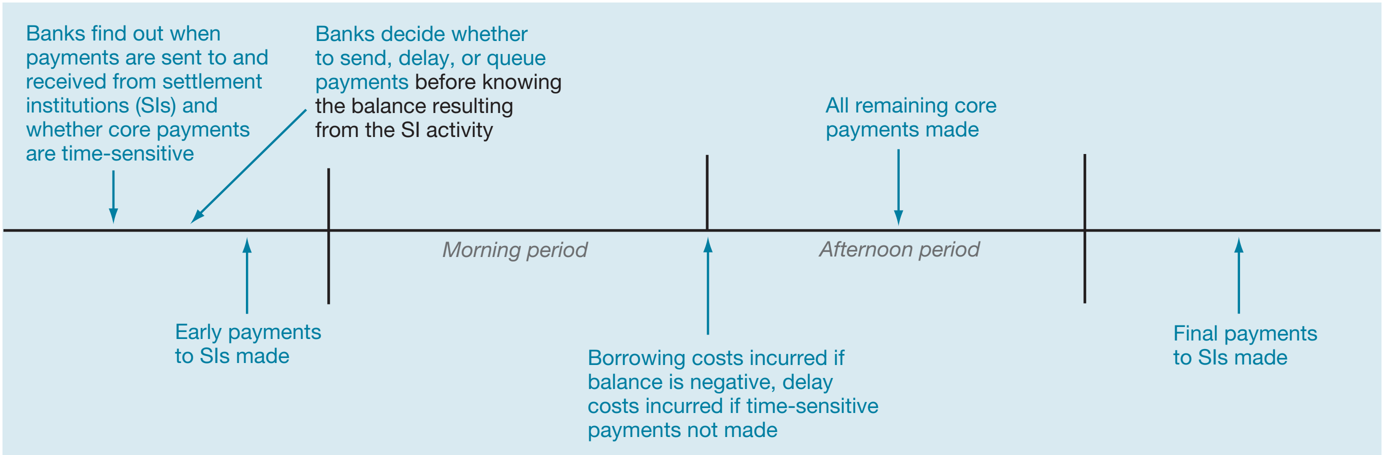


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

Exhibit 6

Receipt-Reactive Liquidity-Saving Mechanism Timeline



If the fraction of time-sensitive payments is relatively small, then the benefit of queuing non-time-sensitive payments under a balance-reactive LSM—rather than delaying them, which would occur in RTGS—is large. Here too a balance-reactive LSM provides higher welfare than would RTGS.

If none of these conditions is satisfied, an instance may arise in which RTGS provides higher welfare than does a balance-reactive LSM. There exists a set of parameters for which the equilibrium described in Chart 2 is the unique RTGS equilibrium. In this equilibrium, banks that do not receive a liquidity shock send a time-critical payment early. For this set of parameters, the best LSM equilibrium is described in Chart 8. In this equilibrium, banks that do not receive a liquidity shock prefer to queue a time-critical payment, rather than send it early, because it provides them with some insurance against the risk of having to borrow from the central bank. However, this action lowers the number of payments that are released in the morning and reduces welfare. This example shows that RTGS can create beneficial coordination among banks to send some payments early. In this case, the presence of a queue unravels that coordination.

We can illustrate the role of liquidity shocks by considering what happens when they are suppressed. Absent liquidity shocks, all payments are released early.

Again, we can illustrate the role of liquidity shocks by considering what happens when they are suppressed. Absent liquidity shocks, all payments are released early. This could happen if all the payments are queued, if they are all sent in the morning outright, or if some are queued and the others are sent in the morning outright.

5. Receipt-Reactive Liquidity-Saving Mechanism

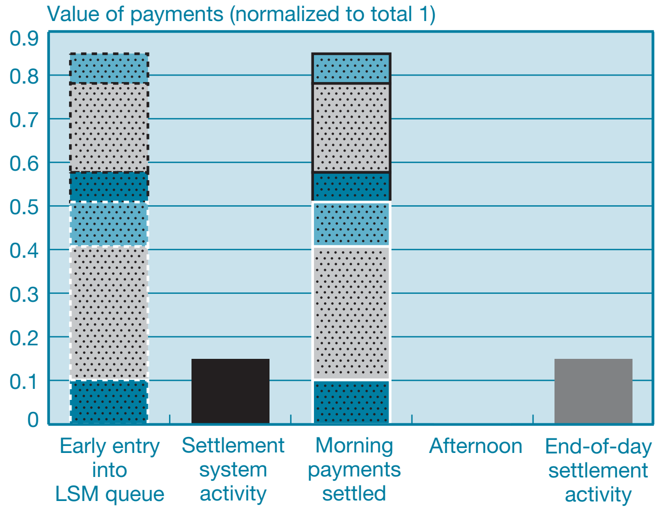
We now consider a different LSM design. Previously, we assumed that banks could make their decision to queue a payment conditional on their liquidity shock or, equivalently, on their balance. Here we assume that banks do not know their liquidity shock when they decide either to queue payments or to pay in the morning period. We also assume that the decision to queue is irrevocable. Since banks can condition their behavior only on the receipt of other payments, but not on their balance, we call this a *receipt-reactive* LSM.[11]

This case illustrates another possible feature of an LSM that can affect bank behavior. In the receipt-reactive LSM, banks are given a tool that enables them to commit to making a payment at a particular time. This ability is valuable to banks. The new timing of events is shown in Exhibit 6.

[11] Receipt-reactive LSMs were first discussed and studied in Johnson, McAndrews, and Soramäki (2004).

Chart 9

Receipt-Reactive Liquidity-Saving Mechanism (LSM) with Low Costs of Delay

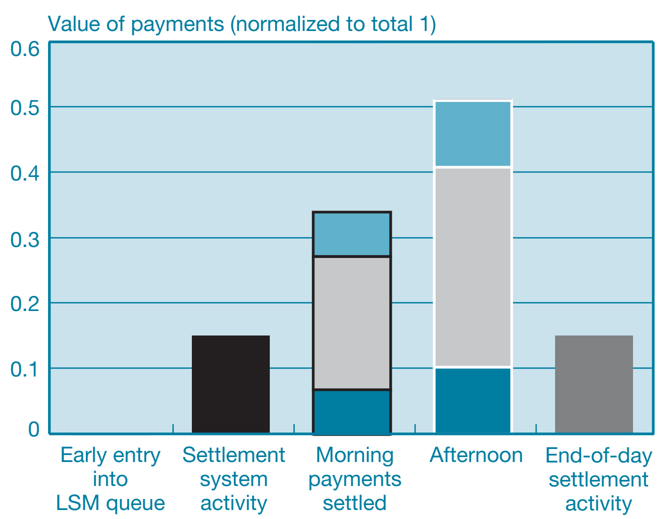


Source: Authors' calculations.

Notes: The six types of banks/colors are described in Exhibit 1. Dashed borders indicate payments submitted to the queue. Shading indicates payments released from the queue. An absence of shading indicates payments either sent outright or delayed (not queued).

Chart 10

Receipt-Reactive Liquidity-Saving Mechanism (LSM) with High Costs of Delay



Source: Authors' calculations.

Note: The six types of banks/colors are described in Exhibit 1.

Charts 9 and 10 illustrate the two types of equilibria that can occur with a receipt-reactive LSM. In the equilibrium described in Chart 9, all payments are queued and hence released in the morning. Note that early payments to settlement institutions, which represent the liquidity shock, occur after payments have been sent to the queue. In the equilibrium described in Chart 10, all time-sensitive payments are made early and non-time-sensitive payments are delayed.

We can now compare welfare among RTGS, a balance-reactive LSM, and a receipt-reactive LSM. In our model, a receipt-reactive LSM always provides welfare at least as high as RTGS. This outcome stands in contrast to a balance-reactive LSM, which provides lower welfare than RTGS under the circumstances outlined above. A receipt-reactive LSM may or may not provide higher welfare than a balance-reactive LSM does. If the cost of delay is sufficiently large, for example, a balance-reactive LSM will provide higher welfare. A balance-reactive LSM will also provide higher welfare when the cost of delay and the probability of a liquidity shock are small.

6. Conclusion

This article studies a model in which banks settle daily payments while seeking to minimize the costs associated with payment delays and intraday borrowing. The novel feature of our model is that banks are subject to two types of shocks. First, banks are randomly assigned to have either time-critical payments, whose late-period settlement imposes a cost on the bank, or non-time-critical payments. Second, banks are subject to liquidity shocks at the start of the day because of the nature of settlement institutions' operations. Together, these two types of shocks yield a rich array of strategic situations. The important parameters in our model are the cost of delay, the cost of borrowing intraday funds from the central bank, the relative size of the payments made to the settlement system with respect to bank-to-bank payments, and the proportion of time-critical payments.

To model the working of a balance-reactive liquidity-saving mechanism, we study two extreme cases. In the first, there is no possibility of netting any strict subset of payments. In the second, payments offset each other bilaterally. These two models provide different motives for using the LSM. In the first case, banks do not assign payments to the LSM queue in the hope of offsetting them within the queue, but rather to have them settle only *conditional* on receiving another payment. In the second, banks can also anticipate that, some of the time, their queued LSM payments will be offset and will settle inside the queue.

In most cases, the presence of a balance-reactive LSM increases welfare compared with a real-time gross settlement system alone, but, perhaps surprisingly, welfare may also be reduced. A balance-reactive LSM provides higher welfare if the cost of delay is high enough or low enough, and if the size of the outside settlement system and the proportion of time-critical payments are relatively low. When this is not true, RTGS can achieve higher welfare. The intuition is that RTGS creates some beneficial coordination of payments that can be undone by the presence of a queue. In our example, some banks that send payments early under the unique RTGS equilibrium choose to put the payments in the queue when they are available. The resulting reduction in the number of payments settled early leads to lower welfare.

With a receipt-reactive LSM, the level of welfare achieved is always at least as high as the level achieved in RTGS alone. Here the intuition is simpler. As banks cannot condition the sending of payments on balances, they either submit all their payments to the queue or they simply make all the time-critical payments in the early period.

In comparing balance-reactive and receipt-reactive LSMs, we find that when delay costs are high and the payments to settlement systems are not too large, the balance-reactive LSM yields a better outcome than its receipt-reactive counterpart. As a result, while our results point to LSMs being at least weakly preferred to RTGS for all parameter configurations, the practical choice can present more of a dilemma to the operator of the large-value payments system. The dilemma is that our results show that the LSM design matters. If the wrong LSM is implemented, it can yield either lower welfare than RTGS or lower welfare than a competing LSM design. The challenge for a payments system operator is to know the sizes of the four parameters of interest. Here we have considered basic design elements in choosing the LSMs to model; more complex designs would introduce other behavioral considerations that are beyond the scope of this article.

Future research in this area would thus benefit from focusing on the empirical magnitudes of the parameters of interest. The cost of delaying payments and the proportion of payments that are time critical are especially important to measure and difficult to observe. Research employing alternative distributions of these parameters will be important, as will be the extension of the current model to include several periods.

References

Angelini, P. 1998. "An Analysis of Competitive Externalities in Gross Settlement Systems." Journal of Banking and Finance 22, no. 1 (January): 1-18.

———. 2000. "Are Banks Risk Averse? Intraday Timing of Operations in the Interbank Market." Journal of Money, Credit, and Banking 32, no. 1 (February): 54-73.

Armantier, O., J. Arnold, and J. McAndrews. 2008. "Changes in the Timing Distribution of Fedwire Funds Transfers." Federal Reserve Bank of New York Economic Policy Review 14, no. 2 (September): 83-112.

Bank for International Settlements. 2005. "New Developments in Large-Value Payment Systems." Committee on Payment and Settlement Systems, Publication no. 67. May. Basel, Switzerland.

Bech, M. L., and R. Garratt. 2003. "The Intraday Liquidity Management Game." Journal of Economic Theory 109, no. 2 (April): 198-219.

Johnson, K., J. J. McAndrews, and K. Soramäki. 2004. "Economizing on Liquidity with Deferred Settlement Mechanisms." Federal Reserve Bank of New York Economic Policy Review 10, no. 3 (December): 51-72.

Martin, A., and J. McAndrews. 2008. "Liquidity-Saving Mechanisms." Journal of Monetary Economics 55, no. 3 (April): 554-67.

McAndrews, J., and J. Trundle. 2001. "New Payment System Designs: Causes and Consequences." Bank of England Financial Stability Review 11, December: 127-36.

Mills, D. C., and T. D. Nesmith. 2008. "Risk and Concentration in Payment and Securities Settlement Systems." Journal of Monetary Economics 55, no. 3 (April): 542-53.

Roberds, W. 1999. "The Incentive Effects of Settlement Systems: A Comparison of Gross Settlement, Net Settlement, and Gross Settlement with Queuing." Bank of Japan Institute for Monetary and Economic Studies Discussion Paper Series, no. 99-E-25, September.

Willison, M. 2005. "Real-Time Gross Settlement and Hybrid Payments Systems: A Comparison." Bank of England Working Paper no. 252.

Todd Keister, Antoine Martin, and James McAndrews

Divorcing Money from Monetary Policy

- Many central banks operate in a way that creates a tight link between money and monetary policy, as the supply of reserves must be set precisely in order to implement the target interest rate.
- Because reserves play other key roles in the economy, this link can generate tensions with central banks' other objectives, particularly in periods of acute market stress.
- An alternative approach to monetary policy implementation can eliminate the tension between money and monetary policy by "divorcing" the quantity of reserves from the interest rate target.
- By paying interest on reserve balances at its target interest rate, a central bank can increase the supply of reserves without driving market interest rates below the target.
- This "floor-system" approach allows the central bank to set the supply of reserve balances according to the payment or liquidity needs of financial markets while simultaneously encouraging the efficient allocation of resources.

1. Introduction

Monetary policy has traditionally been viewed as the process by which a central bank uses its influence over the supply of money to promote its economic objectives. For example, Milton Friedman (1959, p. 24) defined the tools of monetary policy to be those "powers that enable the [Federal Reserve] System to determine the total amount of money in existence or to alter that amount." In fact, the very term *monetary policy* suggests a central bank's policy toward the supply of money or the level of some monetary aggregate.

In recent decades, however, central banks have moved away from a direct focus on measures of the money supply. The primary focus of monetary policy has instead become the value of a short-term interest rate. In the United States, for example, the Federal Reserve's Federal Open Market Committee (FOMC) announces a rate that it wishes to prevail in the federal funds market, where overnight loans are made among commercial banks. The tools of monetary policy are then used to guide the market interest rate toward the chosen target. For this reason, we follow the common practice of using the term *monetary policy* to refer to a central bank's interest rate policy.

It is important to realize, however, that the quantity of money and monetary policy remain fundamentally linked under this approach. Commercial banks hold money in the form of reserve balances at the central bank; these balances are used to meet reserve requirements and make interbank

Todd Keister and Antoine Martin are research officers and James McAndrews a senior vice president at the Federal Reserve Bank of New York.
<todd.keister@ny.frb.org>
<antoine.martin@ny.frb.org>
<james.mcandrews@ny.frb.org>

payments. The quantity of reserve balances demanded by banks varies inversely with the short-term interest rate because this rate represents the opportunity cost of holding reserves. The central bank aims to manipulate the supply of reserve balances—for example, through open market operations that exchange reserve balances for bonds—so that the marginal

[The] link between money and monetary policy can generate tension with central banks' other objectives because bank reserves play other important roles in the economy.

value of a unit of reserves to the banking sector equals the target interest rate. The interbank market for short-term funds will then clear with most trades taking place at or near the target rate. In other words, the quantity of money (especially reserve balances) is chosen by the central bank in order to achieve its interest rate target.

This link between money and monetary policy can generate tension with central banks' other objectives because bank reserves play other important roles in the economy. In particular, reserve balances are used to make interbank payments; thus, they serve as the final form of settlement for a vast array of transactions. The quantity of reserves needed for payment purposes typically far exceeds the quantity consistent with the central bank's desired interest rate. As a result, central banks must perform a balancing act, drastically increasing the supply of reserves during the day for payment purposes through the provision of *daylight reserves* (also called *daylight credit*) and then shrinking the supply back at the end of the day to be consistent with the desired market interest rate.

Recent experience has shown that central banks perform this balancing act well most of the time. Nevertheless, it is important to understand the tension between the daylight and overnight need for reserves and the potential problems that may arise. One concern is that central banks typically provide daylight reserves by lending directly to banks, which may expose the central bank to substantial credit risk. Such lending may also generate moral hazard problems and exacerbate the too-big-to-fail problem, whereby regulators would be reluctant to close a financially troubled bank.

The tension is clearest during times of acute stress in financial markets. In the days following September 11, 2001, for example, the Federal Reserve provided an unusually large quantity of reserves in order to promote the efficient functioning of the payments system and financial markets more generally. As a result of this action, the fed funds rate fell substantially below the target level for several days.[1]

During the financial turmoil that began in August 2007, the tension was much longer lasting. Sharp increases in spreads between the yields on liquid and illiquid assets indicated a classic *liquidity shortage*: an increased demand for liquid assets relative to their illiquid counterparts. By increasing the supply of the most liquid asset in the economy—bank reserves—the Federal Reserve could likely have eased the shortage and helped push spreads back toward more normal levels. Doing so, however, would have driven the market interest rate below the FOMC's target rate and thus interfered with monetary policy objectives. Instead, the Federal Reserve developed new, indirect methods of supplying liquid assets to the private sector, such as providing loans of Treasury securities against less liquid collateral through the Term Securities Lending Facility.

Recently, attention has turned to an alternative approach to monetary policy implementation that has the potential to eliminate the basic tension between money and monetary

It is important to understand the tension between the daylight and overnight need for reserves and the potential problems that may arise.

policy by effectively "divorcing" the quantity of reserves from the interest rate target. The basic idea behind this approach is to remove the opportunity cost to commercial banks of holding reserve balances by paying interest on these balances at the prevailing target rate. Under this system, the interest rate paid on reserves forms a *floor* below which the market rate cannot fall. The supply of reserves could therefore be increased substantially without moving the short-term interest rate away from its target. Such an increase could be used to provide liquidity during times of stress or to reduce the need for daylight credit on a regular basis.[2] A particular version of the "floor-system" approach has recently been adopted by the Reserve Bank of New Zealand.

It should be noted that adopting a floor-system approach requires the central bank to pay interest on reserves, something

[1] Intraday volatility of the fed funds rate remained high, with trades being executed far from the target rate, for several weeks. See McAndrews and Potter (2002) and Martin (forthcoming) for detailed discussions.

[2] This approach has been advocated in various forms by Woodford (2000), Goodfriend (2002), Lacker (2006), and Whitesell (2006b).

the Federal Reserve has historically lacked authorization to do. However, the Financial Services Regulatory Relief Act of 2006 will give the Federal Reserve, for the first time, explicit authority to pay interest on reserve balances, beginning on October 1, 2011. A floor system will therefore soon be a feasible option for monetary policy implementation in the United States.

In this article, we present a simple, graphical model of the monetary policy implementation process to show how the floor system divorces money from monetary policy. Our aim is to present the fundamental ideas in a way that is accessible to a broad audience. Section 2 describes the process by which monetary policy is currently implemented in the United States and in other countries. Section 3 discusses the tensions that can arise in this framework between monetary policy and payments/liquidity policy. Section 4 illustrates how the floor system works; it also discusses potential issues associated with adopting this type of system in a large economy such as the United States. Section 5 concludes.

2. An Overview of Monetary Policy Implementation

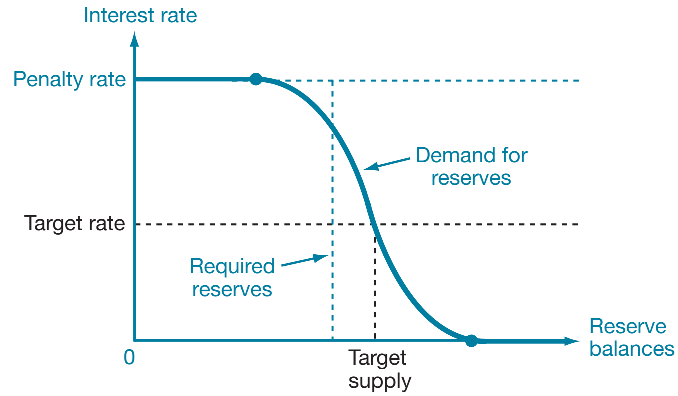
In this section, we describe a stylized model of the process through which many of the world's central banks implement monetary policy. Our model focuses on the relationship between the demand for reserve balances and the interest rate in the interbank market for overnight loans. Following Poole (1968), a variety of papers have developed formal models of portfolio choice by individual banks and derived the resulting aggregate demand for reserves.[3] Our graphical model of aggregate reserve demand is consistent with these more formal approaches. We first discuss the system currently used in the United States and then describe a symmetric channel system, as used by a number of other central banks.

2.1 Monetary Policy Implementation in the United States

We begin by examining the total demand for reserve balances by the U.S. banking system. In our stylized framework, this demand is generated by a combination of two factors. First, banks face reserve requirements. If a bank's final balance is

[3] Recent contributions include Furfine (2000), Guthrie and Wright (2000), Bartolini, Bertola, and Prati (2002), Clouse and Dow (2002), Whitesell (2006a, b), and Ennis and Weinberg (2007).

Exhibit 1
Monetary Policy Implementation in the United States



smaller than its requirement, it pays a penalty that is proportional to the shortfall. Second, banks experience unanticipated late-day payment flows into and out of their reserve account after the interbank market has closed. A bank's final reserve balance, therefore, may be either higher or lower than the quantity of reserves it chooses to hold in the interbank market. This uncertainty makes it difficult for a bank to satisfy its requirement exactly and generates a "precautionary" demand for reserves.

For simplicity, we abstract from a number of features of reality that, while important, are not essential to understanding the basic framework. For example, we assume that reserve requirements must be met on a daily basis, rather than on average over a two-week reserve maintenance period. Alternatively, one can interpret our model as applying to *average* reserve balances (and the average overnight interest rate) over a maintenance period. In addition, we do not explicitly include vault cash in the analysis, using the terms *reserve balances* and *reserves* interchangeably.[4]

Exhibit 1 presents the aggregate demand for reserves in our framework. The horizontal axis measures the total quantity of reserve balances held by banks while the vertical axis measures the market interest rate for overnight loans of these balances. The *penalty rate* labeled on the vertical axis represents the interest rate a bank pays if it must borrow funds at the end of

[4] Required reserves should therefore be interpreted as a bank's requirement net of its vault cash holdings. To the extent that vault cash holdings are independent of the overnight rate, at least over short horizons, including them in our model would have no effect. We also abstract from the Contractual Clearing Balance program, which allows banks to earn credit for priced services at the Federal Reserve by holding a contractually agreed amount of reserves in excess of their requirement; these contractual arrangements, once set, act much like reserves requirements.

the period to meet its requirement. One can interpret this penalty rate as the interest rate charged at the Federal Reserve's primary credit facility (the discount window), adjusted by any "stigma" costs that banks perceive to be associated with borrowing at this facility. The important feature of the penalty rate is that it lies above the FOMC's target interest rate.[5]

The . . . demand for reserve balances will vary inversely with the market interest rate, since this rate represents the opportunity cost of holding reserves.

To explain the shape of the demand curve in the exhibit, we ask: given a particular value for the interest rate, what quantity of reserve balances would banks demand to hold if that rate prevailed in the interbank market? First, note that if the market interest rate were above the penalty rate, there would be an arbitrage opportunity: banks could borrow reserves at the (lower) penalty rate and lend them at the (higher) market interest rate. If the market interest rate were exactly equal to the penalty rate, however, banks would be willing to hold some reserve balances toward meeting their requirements. In fact, each bank would be indifferent between holding reserves directly and borrowing at the penalty rate as long as it is sure that late-day payment inflows will not leave it holding excess balances at the end of the day. As a result, the demand curve is flat—reflecting this indifference—at the level of the penalty rate for sufficiently small levels of reserve balances.

For interest rates below the penalty rate, each bank will choose to hold a quantity of reserves that is close to the level of its requirement; hence, aggregate reserve demand will be close to the total level of required reserves. However, as described above, banks face uncertainty about their final account balance that prevents them from being able to meet their requirement exactly. Instead, each bank must balance the possibility of falling short of its requirement—and being forced to pay the penalty rate—against the possibility that it will end up holding more reserves than are required. As no interest is paid on reserves, holding excess balances is also costly. The resulting demand for reserve balances will vary inversely with the market interest rate, since this rate represents the opportunity cost of holding reserves. The less expensive it is to hold precautionary reserve balances, the greater the quantity demanded by the banking system will be. This reasoning generates the downward-sloping part of the demand curve in the exhibit.

If the market interest rate were very low—close to zero—the opportunity cost of holding reserves would be very small. In this case, each bank would hold enough precautionary reserves to be virtually certain that unforeseen payment flows will not decrease its reserve balance below the required level. In other words, each bank would choose to be "fully insured" against the possibility of falling short of its requirements. The point in Exhibit 1 where the demand curve intersects the horizontal axis represents the total of this fully insured quantity of reserve balances for all banks. The banking system will not demand more than this quantity of reserve balances as long as there is some opportunity cost, no matter how small, of holding these reserves.

If the market interest rate were exactly zero, however, there would be no opportunity cost of holding reserves. In this limiting case, there is no cost at all to a bank of holding additional reserves above the fully insured amount. The demand curve is therefore flat along the horizontal axis after this point; banks are indifferent between any quantities of reserves above the fully insured amount when the market interest rate is exactly zero.

Needless to say, our model of reserve demand abstracts from important features of reality. Holding more reserves, for example, might require a bank to raise more deposits and subject it to higher capital requirements. Nevertheless, the model is useful because it lays out, in perhaps the simplest way possible, the basic relationship between the market interest rate and the demand for reserves that results from the optimal portfolio decisions of banks. Moreover, small changes in the shape of the demand curve would have no material effect on the analysis that follows.

The equilibrium interest rate in our model is determined by the height of the demand curve at the level of reserve balances supplied by the Federal Reserve. If the supply is smaller than the total amount of required reserves, for example, the equilibrium interest rate would be near the penalty rate. If, however, the supply of reserves were very large, the equilibrium interest rate would be zero. Between these two extremes, on the downward-sloping portion of the demand curve, there is a *liquidity effect* of reserve balances on the market interest

[5] The interest rate charged on discount window loans has been set above the FOMC's target rate since the facility was redesigned in 2003. The gap between the two rates was initially set at 100 basis points, but has since been lowered to 50 basis points (in August 2007) and to 25 basis points (in March 2008). In addition, there is evidence that banks attach a substantial nonpecuniary cost to borrowing from the discount window, as they sometimes borrow in the interbank market at interest rates significantly higher than the discount window rate. These stigma costs may reflect a fear that other market participants will find out about the loan and interpret it as a sign of financial weakness on the part of the borrowing bank.

rate: a higher supply of reserves will lower the equilibrium interest rate.[6]

As shown in the exhibit, there is a unique level of reserve supply that will lead the market to clear at the FOMC's announced target rate; we call this level the *target supply*. Monetary policy is implemented through open market operations that aim to set the supply of reserves to this target level. This process requires the Fed's Open Market Desk to accurately forecast both reserve demand and changes in the existing supply of reserves attributable to autonomous factors such as payments into and out of the Treasury's account. Forecasting errors will lead the actual supply to deviate from the target and, hence, will cause the market interest rate to differ from the target rate. In our simple model, the downward-sloping portion of the demand curve may be quite steep, indicating that relatively small forecasting errors could lead to substantial interest rate volatility. In reality, a variety of institutional arrangements, including reserve maintenance periods, are designed to flatten this curve and thus limit the volatility associated with forecasting errors.[7]

Monetary policy is implemented in the United States by changing the supply of reserves in such a way that the fed funds market will clear at the desired rate. In other words, the stock of "money" is set in order to achieve a monetary policy objective.

The key point of this discussion is that monetary policy is implemented in the United States by changing the supply of reserves in such a way that the fed funds market will clear at the desired rate. In other words, the stock of "money" is set in order to achieve a monetary policy objective. This direct relationship between money and monetary policy generates the tensions that we discuss in Section 3.

[6] See Hamilton (1997), Carpenter and Demiralp (2006a), and Thornton (2006) for empirical evidence of this liquidity effect.

[7] See Ennis and Keister (2008) for a detailed discussion of interest rate volatility in this basic framework. See Whitesell (2006a) for a formal model of the "flattening" effect of reserve maintenance periods.

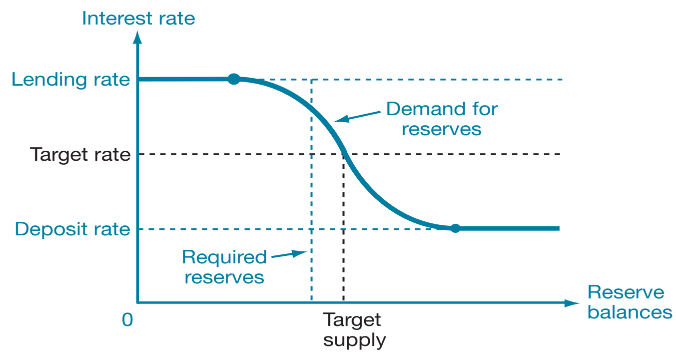
2.2 Symmetric Channel Systems

Many central banks use what is known as a symmetric *channel* (or *corridor*) system for monetary policy implementation. Such systems are used, for example, by the European Central Bank (ECB) and by the central banks of Australia, Canada, England, and (until spring 2006) New Zealand. The key features of a symmetric channel system are standing central bank facilities that lend to and accept deposits from commercial banks. The lending facility resembles the discount window in the United States; banks are permitted to borrow freely (with acceptable collateral) at an interest rate that is a fixed number of basis points above the target rate. The deposit facility allows banks to earn overnight interest on their excess reserve holdings at a rate that is the same number of basis points below the target. In this way, the interest rates at the two standing facilities form a "channel" around the target rate.

Exhibit 2 depicts the demand for reserve balances in a symmetric channel system. The curve looks very similar to that in Exhibit 1. There is no demand for reserves in the interbank market if the interest rate is higher than the rate at the lending facility.[8] For lower values of the market rate, the demand is decreasing in the interest rate—and hence the liquidity effect is present—for exactly the same reasons as before. Banks choose their reserve holdings to balance the potential costs of falling short of their requirement against the potential costs of ending with excess reserves. When the opportunity cost of holding reserves is lower, banks' precautionary demand for reserves will be larger.

Exhibit 2

A Symmetric Channel System of Monetary Policy Implementation



[8] The lending facility in a channel system is typically designed in a way that aims to minimize stigma effects. For this reason, we begin the demand curve in Exhibit 2 at the lending rate instead of at a penalty rate that includes stigma effects, as was the case in Exhibit 1.

The new feature in Exhibit 2 is that the demand curve does not decrease all the way to the horizontal axis, but instead becomes flat at the deposit rate. In other words, the deposit rate forms a *floor* below which the demand curve will not fall. If the market rate were below the deposit rate, an arbitrage opportunity would exist—a bank could borrow at the (low) market rate and earn the (higher) deposit rate on these funds, making a pure profit. The demand for reserves would be unbounded in this case; such arbitrage activity would quickly drive up the market rate until it at least equals the deposit rate.

The demand curve is flat at the deposit rate for the same reason it was flat on the horizontal axis in Exhibit 1. If the market rate were exactly equal to the deposit rate, banks would face no opportunity cost of holding excess reserves. Holding additional funds on deposit and lending them would yield exactly the same return. Banks would therefore be indifferent between any quantities of reserves above the fully insured amount. In other words, paying interest on excess reserves raises the floor where the demand curve is flat from an interest rate of zero (as in Exhibit 1) to the deposit rate (as in Exhibit 2).

[In a symmetric channel system,] the target interest rate determines, through the demand curve, a target supply of reserves, and the central bank aims to change total reserve supply to bring it as close as possible to this target.

The equilibrium interest rate is determined exactly as before, by the height of the demand curve at the level of reserve balances supplied by the central bank. Monetary policy is thus implemented in much the same way as it is in the United States. The target interest rate determines, through the demand curve, a target supply of reserves, and the central bank aims to change total reserve supply to bring it as close as possible to this target. Importantly, the link between money and monetary policy remains: the quantity of reserves is set in order to achieve the desired interest rate.

The symmetric channel systems used by various central banks differ in a variety of important details. The Bank of England and the ECB operate relatively wide channels, with the standing facility rates 100 basis points on either side of the target. Australia and Canada, in contrast, operate narrow channels, where this figure is only 25 basis points. Australia and Canada have no required reserves; in this case, the demand curve in Exhibit 2 shifts to the left so that the "required reserves" line lies on the vertical axis. The important point here, however, is that regardless of these operational details, a symmetric channel system links the quantity of reserves to the central bank's interest rate target, exactly as in the U.S. system.

3. Payments, Liquidity Services, and Reserves

The link between money and monetary policy described above can generate tension with central banks' other objectives, particularly those regarding the payments system and the provision of liquidity. Reserve balances are useful to banks, and to the financial system more generally, for purposes other than simply meeting reserve requirements. Banks use reserve balances to provide valuable payment services to depositors. In addition, these balances assist the financial sector in allocating other, less liquid assets. Since reserves are a universally accepted asset, they can be exchanged more easily for other assets than any substitute. Finally, reserve balances serve as a perfectly liquid, risk-free store of value, which is particularly useful during times of market turmoil. Because reserves play these other important roles, the quantity of reserve balances consistent with the central bank's monetary policy objective may at times come into conflict with the quantity that is desirable for other purposes. In this section, we describe some of the tensions that can arise.

3.1 Payments Policy

The value of the payments made during the day in a central bank's large-value payments system is typically far greater than the level of reserve balances held by banks overnight. (In the United States, for example, during the first quarter of 2008 the average daily value of transactions over the Fedwire Funds Service was approximately *185 times* the value of banks' total balances on deposit at the Federal Reserve.) The discrepancy has widened in recent decades as most central banks have adopted a real-time gross settlement (RTGS) design for their large-value payments system, which requires substantially larger payment flows than earlier designs based on netting of payment values.[9]

As a result, banks' overnight reserve holdings are too small to allow for the smooth functioning of the payments system

[9] See Bech and Hobijn (2007) for an analysis of the adoption of RTGS systems by various central banks.

during the day. When reserves are scarce or costly during the day, banks must expend resources in carefully coordinating the timing of their payments. If banks delay sending payments to economize on scarce reserves, the risk of an operational failure or gridlock in the payments system tends to increase. The combination of limited overnight reserve balances and the much larger daylight demand for reserves thus creates tension between a central bank's monetary policy and its payments policy. The central bank would like to increase the total supply of reserve balances for payment purposes, but doing so would interfere with its monetary policy objectives.

> *The value of the payments made during the day in a central bank's large-value payments system is typically far greater than the level of reserve balances held by banks overnight . . . . As a result, banks' overnight reserve holdings are too small to allow for the smooth functioning of the payments system during the day.*

This tension has led to a common practice among central banks of supplying additional reserves to the banking system for a limited time during the day. These daylight reserves (also called *daylight credit*) are typically lent directly to banks. Many central banks provide daylight reserves against collateral at no cost to banks. The Federal Reserve currently supplies daylight credit to banks on an uncollateralized basis for a small fee.[10] In providing daylight reserves, a central bank aims to allow banks to make their payments during the day smoothly and efficiently while limiting its own exposure to credit risk.

Under normal circumstances, this process of expanding the supply of reserves during the day and shrinking it back overnight works well; banks make payments smoothly and the central bank implements its target interest rate. However, this balancing act is not without costs. Lending large quantities of reserves to banks each day exposes the central bank to credit risk. While requiring collateral for these loans mitigates credit risk, it is an imperfect solution. If collateral is costly for banks to hold or create, the requirement imposes real costs. Moreover, collateralizing daylight loans simply moves the central bank's claims ahead of the deposit insurance fund in the event of a bank failure, without necessarily reducing the overall risk of the consolidated public sector.

Routine daylight lending by the central bank may also create moral hazard problems, leading banks to hold too little liquidity and, perhaps, take on too much risk. In addition, such lending might make regulators more reluctant to close a financially troubled bank promptly, exacerbating the well-known too-big-to-fail problem. Even if each of these costs is relatively small in normal times, their sum should be considered part of the tension generated by the link between money and monetary policy.

3.2 Liquidity Policy

In times of stress or crisis in financial markets, the tension between monetary policy and central banks' other objectives can become acute. After the destructive events of September 11, 2001, the Federal Reserve recognized that the quantity of overnight reserves consistent with the target fed funds rate was too small to adequately address banks' reluctance to make payments in a timely manner. The FOMC released a statement on September 17, 2001, that, in addition to lowering the target fed funds rate, stated:

> The Federal Reserve will continue to supply unusually large volumes of liquidity to the financial markets, as needed, until more normal market functioning is restored. As a consequence, the FOMC recognizes that the actual federal funds rate may be below its target on occasion in these unusual circumstances.[11]

In this statement, the FOMC explicitly recognized the tension between maintaining the market interest rate at its target level and supplying more reserves to meet the demand for financial market settlements. On September 18 and 19, the effective fed funds rate was close to 1¼ percent while the target rate was 3 percent.

Exhibit 1 is again useful to help illustrate what happened. To meet the demand for reserves for financial settlements in various markets, the Fed increased the supply of reserve balances. A shift in the supply curve to the right implies that intersection with the demand curve will occur at a lower interest rate.[12] In this case, it was not possible to achieve simultaneously the interest rate target and the increase in overnight reserves necessary to ensure the efficient functioning

[10] See Board of Governors of the Federal Reserve System (2008b) for a proposal to change the Federal Reserve's method of supplying daylight reserves. Under this proposal, banks would be able to obtain daylight reserves either on a collateralized basis at no cost or on an uncollateralized basis for a higher fee. For a general discussion of the Federal Reserve's policies on daylight credit, see Board of Governors of the Federal Reserve System (2007).

[11] See <http://www.federalreserve.gov/boarddocs/press/general/2001/20010917>.

of financial markets in conditions of stress. The exact same tension would arise under a symmetric channel system. Note, however, that the channel places a limit on how far the market interest rate can deviate from the target—it cannot fall below the deposit rate.

During the events of September 2001, the fed funds rate was below its target for only a few days and thus likely had no impact on monetary policy objectives, as expectations were that the target rate would quickly be reestablished. It is an instructive episode, however, in that it demonstrates how increasing the supply of reserve balances available to the banking system can support market liquidity, and how this objective can interfere with the maintenance of the target interest rate.

The Federal Reserve faced a different type of liquidity issue during the financial market turmoil that began in August 2007. In this case, there was a sharp decline in . . . broad liquidity*: the ease with which assets in general can be sold or used as collateral at a price that appropriately reflects the expected value of the asset's future dividends.*

The Federal Reserve faced a different type of liquidity issue during the financial market turmoil that began in August 2007. In this case, there was a sharp decline in what Goodfriend (2002) calls *broad liquidity*: the ease with which assets in general can be sold or used as collateral at a price that appropriately reflects the expected value of the asset's future dividends. Goodfriend argues that increasing the supply of bank reserves can also support the level of broad liquidity in financial markets. This is especially true if the central bank uses the newly created reserves to purchase (or lend against) relatively illiquid assets, thereby increasing the total quantity of liquid assets held by the private sector. However, once again the link between money and monetary policy generates a tension; the central bank cannot pursue an independent "liquidity policy" using bank reserves. Any attempt to increase reserve balances for the purpose of providing additional liquidity would lead to a lower short-term interest rate and, hence, would change the stance of monetary policy.

Goodfriend (2002, p. 4) points out that central banks can use other, less direct methods of managing broad liquidity:

> To some degree, the Fed can already manage broad liquidity under current operating procedures by changing the composition of its assets, for example, by selling liquid short-term Treasury securities and acquiring less liquid longer term securities. However, the government debt injected into the economy in this way would not be as liquid as newly created base money. More importantly, the Fed's ability to affect broad liquidity in this way is strictly limited by the size of its balance sheet.

Interestingly, one of the new facilities introduced by the Fed in response to the market turmoil closely resembled the policy described by Goodfriend. The Term Securities Lending Facility, introduced in March 2008, provides loans of Treasury securities using less liquid assets as collateral.[13] These loans increase broad liquidity by raising the total supply of highly liquid assets (reserves plus Treasury securities) in the hands of the private sector and decreasing the supply of less liquid assets. However, as Goodfriend observes, the amount of broad liquidity that can be provided through such a facility is strictly limited by the quantity of Treasury securities owned by the central bank. Thus, while a central bank can pursue a policy based on changes in the composition of its assets, such a policy has inherent limitations. As we discuss in Section 4, alternative methods of monetary policy implementation allow the central bank to overcome this limitation by pursuing a liquidity policy based directly on bank reserves.

3.3 Efficient Allocation of Resources

Another tension generated by the typical methods of monetary policy implementation described earlier relates to efficiency concerns. These methods rely on banks facing an opportunity cost of holding reserves; their balances earn no interest in the U.S. system and earn less than the prevailing market rate in a symmetric corridor. This opportunity cost helps generate the downward-sloping part of the demand curve that the central bank uses to implement its target interest rate. The fact that

[12] Needless to say, the disruption in financial markets would also tend to increase the demand for reserves, shifting the curve in Exhibit 1 to the right. The FOMC's statement indicates a desire to more than compensate for this shift, that is, to increase reserve supply beyond the point that would maintain the target interest rate given the increased reserve demand.

[13] The Fed also introduced other facilities, including the Term Auction Facility and the Primary Dealer Credit Facility. Those facilities make loans of reserve balances. In order to maintain the target interest rate, however, the Fed uses open market operations to "sterilize" these loans, leaving the total supply of reserve balances unaffected.

holding reserves is costly, however, conflicts with another central bank objective: the desire to promote the efficient functioning of financial markets and the efficient allocation of resources more generally.

Remunerating reserve balances at a below-market interest rate is effectively a tax on holding these balances. (Box 1 discusses how this tax is distortionary when applied to required reserves.) Similar logic shows that a distortion arises when banks face an opportunity cost of holding *excess* reserves. In this case, the tax leads banks to invest real resources in economizing on their holdings of excess reserves, but these efforts produce no social benefit.

Reserve balances are costless for a central bank to create through open market operations, for example, that exchange newly created reserves for Treasury securities. If banks perceive an opportunity cost of holding reserves (relative to Treasury securities, say), then they will engage in socially inefficient efforts to reduce their use of reserves. In other words, the tax places a wedge between a private marginal rate of substitution and the corresponding social marginal rate of transformation. This type of distortion was emphasized by Friedman (1959, pp. 71-5), who argues that the central bank should pay interest on all reserve balances at the prevailing market interest rate.[14]

Another tension generated by the typical methods of monetary policy implementation . . . relates to efficiency concerns.

One might be tempted to suppose that the distortions created by this tax must be small because the quantity of excess reserves held by banks is currently fairly small in the United States, around $1.5 billion. Such a conclusion is not warranted, however: the fact that the tax base is small does not imply that the deadweight loss associated with the tax is insignificant. The deadweight loss includes all efforts banks expend to avoid holding excess reserves, including closely monitoring end-of-day and end-of-maintenance-period balances so that any

[14] This logic is central to the well-known *Friedman rule*, which calls for the central bank to eliminate the opportunity cost of holding all types of money (see especially Friedman [1969]). One way to implement this rule is by engineering a deflation that makes the real return on holding currency equal to the risk-free return. In this case, no interest needs to be paid on any form of money; the deflation generates the required positive return. In practice, there are a variety of concerns about deflation that keep central banks from following this approach. When applied to the narrower question of reserve balances held at the central bank, however, Friedman's logic simply calls for remunerating all reserve balances at the risk-free rate.

Box 1

Required Reserves

Although this article emphasizes the similarities in monetary policy implementation procedures across countries, there are a number of differences. One notable difference is in the use of reserve requirements. Banks in the United States and the Euro zone are required to hold reserves in proportion to certain liabilities. In other countries, including Australia and Canada, banks are not required to hold any reserves; the only requirement is that a bank's reserve account not be in overdraft at the end of the day.

In the simple framework we describe, it is immaterial whether banks face a positive reserve requirement or the requirement is effectively zero. In reality, however, there are important differences between these approaches. One such difference is that reserve requirements allow the central bank to implement reserve averaging, whereby banks are allowed to meet their requirement on average over a reserve maintenance period rather than every day. As shown in Whitesell (2006a), reserve averaging tends to flatten the demand curve for reserves around the central bank's target supply on all days of a maintenance period except the last one; this flattening tends to reduce volatility in the market interest rate.[a] Another important difference is the extent of the distortions associated with bank reserve holdings. When required reserve balances do not earn interest, as is currently the case in the United States, the requirement acts as a tax on banks. This *reserve tax* raises banks' operating costs and drives a wedge between the price of banking services and the social cost of producing those services, creating a deadweight loss. The reserve tax also gives banks a strong incentive to find ways to decrease their requirements, such as by sweeping customers' checking account balances on a daily basis into other accounts not subject to reserve requirements. The efforts invested in these reserve-avoidance activities are clearly wasted from a social point of view.

Paying interest on required reserves at the prevailing market rate of interest, as the European Central Bank does, eliminates most of these distortions. The Bank of England goes a step further by having banks set *voluntary* balance targets. Once set, these targets can be used to implement monetary policy exactly the same way that reserve requirements are. However, because the targets are chosen by the individual banks, rather than being determined administratively, their creation generates none of the distortions associated with traditional reserve requirements.

[a] See Ennis and Keister (2008) for a detailed discussion of reserve averaging in the type of framework used here.

excess funds can be lent out, as well as actually lending the funds out. A substantial fraction of activity in the fed funds market is precisely of this type, and it is not clear whether these indirect costs associated with the tax are small.

The issues created by the reserve tax are sometimes described as a "hot potato" problem. Participants all try to get rid of excess reserves because holding them is costly. However, the supply of excess reserve balances is fixed by the central bank and, at any point in time, someone must be holding them. Extending this analogy a bit, the fact that the potato itself (that is, the quantity of excess reserve balances) is small does not imply that that the efforts spent passing it along are also small. This is especially true if the potato is very hot, that is, if excess reserve balances earn much less than the market rate of interest.

The issues created by the reserve tax are sometimes described as a "hot potato" problem. Participants all try to get rid of excess reserves because holding them is costly. However, the supply of excess reserve balances is fixed by the central bank and, at any point in time, someone must be holding them.

Lucas (2000, p. 247) describes the deadweight loss associated with the inflation tax in a similar way:

> In a monetary economy, it is in everyone's private interest to try to get someone else to hold non-interest-bearing cash and reserves. But someone has to hold it all, so all of these efforts must simply cancel out. All of us spend several hours per year in this effort, and we employ thousands of talented and highly trained people to help us. These person-hours are simply thrown away, wasted on a task that should not have to be performed at all.

Any system of monetary policy implementation that relies on banks facing an opportunity cost of holding reserves necessarily creates deadweight losses. The approaches described in the previous section thus conflict with a central bank's desire to promote an efficient allocation of resources in the economy.

We summarize by noting that a central bank's payments policy, liquidity policy, and desire to promote efficient allocation may all come into conflict with its monetary policy objectives. The tension created by these conflicts tends to be particularly strong during periods of stress in financial markets. These tensions would be reduced or would disappear altogether if banks did not face an opportunity cost of holding overnight reserves that leads them to economize on their holdings. In the next section, we describe an approach to implementing monetary policy that removes this opportunity cost and discuss some of its implications.

A central bank's payments policy, liquidity policy, and desire to promote efficient allocation may all come into conflict with its monetary policy objectives. The tension created by these conflicts tends to be particularly strong during periods of stress in financial markets.

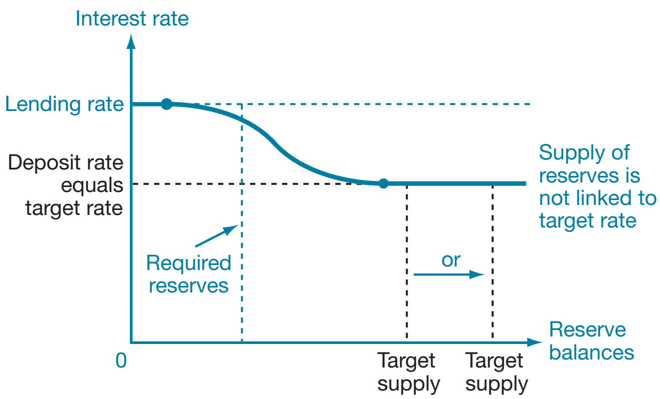
4. Divorcing Money from Monetary Policy

The tensions we described all arise from the fact that, under either current U.S. practice or a symmetric channel system, the quantity of reserve balances must be set to a particular level in order for the central bank's interest rate target to be achieved. There are, however, other approaches to monetary policy implementation in which this strict link between money and monetary policy is not present. Here we discuss one such approach, which can be described as a *floor-target channel system*, or simply a *floor system*. This approach is a modified version of the channel system described above and has been advocated in various forms by Woodford (2000), Goodfriend (2002), Lacker (2006), and Whitesell (2006b). A particular type of floor system has recently been adopted by the Reserve Bank of New Zealand.

4.1 The Floor System

Starting from the symmetric channel system presented in Exhibit 2, suppose that the central bank makes two

Exhibit 3
A Floor System of Monetary Policy Implementation



modifications. First, the deposit rate is set *equal to* the target rate, instead of below it. In other words, in this system the central bank targets the *floor* of the channel, rather than some point in the interior. Second, the reserve supply is chosen so that it intersects the flat part of the demand curve generated by the deposit rate (Exhibit 3), rather than intersecting the downward-sloping part of the curve. Supply and demand will then cross exactly at the target rate, as desired.[15]

The key feature of this system is immediately apparent in the exhibit: the equilibrium interest rate no longer depends on the exact quantity of reserve balances supplied. Any quantity that is large enough to fall on the flat portion of the demand curve will implement the target rate. In this way, a floor system "divorces" the quantity of money from the interest rate target and, hence, from monetary policy. This divorce gives the central bank two separate policy instruments: the interest rate target can be set according to the usual monetary policy concerns, while the quantity of reserves can be set independently.

If the quantity of reserves is no longer determined by monetary policy concerns, how should it be set? In general, the supply of overnight reserve balances could be used to ease any of the tensions described earlier. For example, Lacker (2006) suggests that increasing the supply of overnight reserves could reduce banks' use of daylight credit without impairing their ability to make timely payments. In fact, he argues that if overnight reserve balances are increased by the maximum amount of current daylight credit use, then "in principle, any pattern of intraday payments that is feasible under the current policy would still be feasible" even in the extreme case where access to daylight credit is eliminated altogether. Note that restricting access to daylight credit will tend to increase the demand for overnight reserves, shifting the curve in Exhibit 3 to the right. The proposal in Lacker (2006) thus calls for increasing the supply of reserves enough to ensure that it falls on the flat portion of the demand curve even after this shift is taken into account.[16]

Goodfriend (2002) takes a different view, proposing that the supply of reserve balances could be used to stabilize financial markets. The central bank could, for example, "increase bank reserves in response to a negative shock to broad liquidity in banking or securities markets or an increase in the external finance premium that elevated spreads in credit markets" (p. 4). More generally, he suggests that the supply of reserves could be set to provide the optimal quantity of broad liquidity services.[17] It should be noted that there may be complementarity between payments policy and liquidity policy with respect to reserve balances; increasing the reserve supply to support broad liquidity can simultaneously reduce the use of daylight overdrafts, which might be particularly desirable during times of market turmoil.

A floor system "divorces" the quantity of money from the interest rate target and, hence, from monetary policy. This divorce gives the central bank two separate policy instruments: the interest rate target can be set according to the usual monetary policy concerns, while the quantity of reserves can be set independently.

The floor system also promotes a more efficient allocation of resources. Not only does this approach eliminate the reserve tax, it also removes the opportunity cost of holding *excess*

[15] The fact that these supply and demand curves cross at the target rate does not imply that trades in the interbank market would occur at exactly this rate. A bank would require a small premium, reflecting transaction costs and perhaps credit risk, in order to be willing to lend funds rather than simply hold them as (interest-bearing) reserves. As a result, the measured interest rate in the interbank market would generally be slightly above the deposit rate. The target rate could instead be called the *policy rate* in order to make this distinction clear.

[16] See Ennis and Weinberg (2007) for a formal analysis of the relationship between daylight credit and monetary policy implementation, including the ability of a floor system to reduce daylight credit usage.

[17] Determining this optimal quantity is a nontrivial task, however, and would likely require more research on the notion of broad liquidity and its role in the macroeconomy. The quantitative easing policy in place in Japan from 2001 to 2006 can be viewed as an attempt to use the supply of bank reserves to influence macroeconomic outcomes.

reserve balances. This is true for any quantity of reserve balances large enough to lie on the flat portion of the demand curve in Exhibit 3. At such points, banks are indifferent at the margin between reserves and other risk-free assets. As a result, they no longer have an incentive to invest real resources in order to economize on their reserve holdings, and the deadweight loss associated with the systems described in Section 3 disappears.

Woodford (2000) points to another advantage of the floor system. Suppose that innovation in financial markets were to

The Reserve Bank of New Zealand recently became the first central bank to implement a floor system. While it is too early to evaluate the effects of this change properly, some benefits—such as improved timeliness of payments—have already been observed.

undermine the demand for reserve balances that is at the heart of our model in Section 2. In particular, suppose that a perfect substitute for central bank reserves were developed and that banks were able to avoid reserve requirements completely. In such a situation, the demand for reserves would fall to zero if there were *any* opportunity cost of holding them; banks would instead use the substitute private instrument for payment and other liquidity purposes. If the central bank supplied a positive quantity of reserves, under the current system in the United States the market interest rate would fall to zero.

Woodford argues that even in this extreme situation, the central bank can still implement its target interest rate by using a floor system. Banks would again demand zero reserves at any interest rate higher than the target rate in this situation. However, under a floor system, the demand curve would be flat at the target rate for exactly the same reasons as described above. By setting a positive supply of reserves, therefore, the central bank could still drive the market interest rate to the target value. In this way, a floor system would enable the central bank to meet its monetary policy objectives even if technological changes eliminated the special role currently played by reserves; the key once again is divorcing money from monetary policy.

The Reserve Bank of New Zealand recently became the first central bank to implement a floor system (Box 2). While it is too early to evaluate the effects of this change properly, some benefits—such as improved timeliness of payments—have already been observed. To be sure, the experience of a smaller country like New Zealand with this type of system may not be directly applicable to other central banks. Nevertheless, it will be instructive to observe this experience and, in particular, to see how it compares with the simple framework we present.

4.2 Discussion

While a floor system could potentially relieve or even eliminate the tensions between central bank objectives, there are several important concerns about how such a system would operate in practice and its potential effects on financial markets. One concern is that a floor system would likely lead to a substantial reduction in activity in the overnight interbank market, as banks would have less need to target their reserve balance precisely on a daily basis. In particular, since banks with excess funds can earn the target rate by simply depositing them with the central bank, the incentive to lend these funds is lower than it is under the other approaches to implementation discussed above. Nevertheless, an interbank market would still be necessary, as institutions will occasionally find themselves short of funds. How difficult it would be for institutions to borrow at or near the target rate is an important open question.

In addition, some observers argue that the presence of an active overnight market generates valuable information and

While a floor system could potentially relieve or even eliminate the tensions between central bank objectives, there are several important concerns about how such a system would operate in practice and its potential effects on financial markets.

that some of this information would be lost if market activity declined. For example, market participants must monitor the creditworthiness of borrowers. If the overnight market were substantially less active, such monitoring may not take place on a regular basis; this in turn could make borrowing even harder for a bank that finds itself short of funds. Such monitoring may also play a socially valuable role in exposing banks to market discipline. It is important to bear in mind, however, that the

Box 2

The Reserve Bank of New Zealand's Floor System

In July 2006, the Reserve Bank of New Zealand (RBNZ) began the transition from a symmetric channel system of monetary policy implementation to a floor system. We describe some reasons for the change and some features of the new regime, drawing heavily on Nield (2006) and Nield and Groom (2008).

From 1999 to 2006, the RBNZ operated a symmetric channel system with zero reserve requirements. It targeted a supply of NZD 20 million overnight reserve balances every day. All reserve balances were remunerated at a rate 25 basis points below the RBNZ's target interest rate, called the official cash rate (OCR). Payments system participants could borrow reserves overnight against collateral at the overnight reserve repurchase facility (ORRF), at a rate 25 basis points above the OCR. Finally, participants could obtain reserves intraday, against collateral, at an interest rate of zero using a facility called Autorepo.

The RBNZ's decision to change the framework for monetary policy implementation followed signs of stress in the money market. The Government of New Zealand had been running a fiscal surplus for a number of years and government bonds had become increasingly scarce. The scarcity of government securities available to pledge in the Autorepo facility led to delayed payments between market participants. For the same reason, there had been an increase in the levels of underbid open market operations and, consequently, in the use of the bank's standing facilities at the end of the day. Finally, the implied New Zealand dollar interest rates on overnight credit in the foreign exchange (FX) swap market—the primary market by which banks in New Zealand traded overnight—were volatile and often significantly above the target rate desired to implement monetary policy.

The Reserve Bank of New Zealand conducted a review of its liquidity management regime in 2005 and announced the new system in early 2006. Under this system, the RBNZ no longer offers daylight credit. In other words, there is no distinction between daylight and overnight reserves. The target supply of reserves has been vastly increased to allow for the smooth operation of the payments system; the new level currently fluctuates around NZD 8 billion. This represents an increase of 400 times the level under the previous regime. Reserves are now remunerated at the OCR. It is still possible to obtain overnight funds at the ORRF, but at a rate 50 basis points above the OCR.

The bulk of the transition to this new system occurred in four steps over a twelve-week period between July 3 and October 5, 2006. During that time, the target supply of reserves increased gradually to its current level. At each step, the rate earned on reserves and the rate at which funds could be borrowed at the ORRF were increased relative to the OCR in increments of 5 basis points up to their current levels. The set of securities eligible as collateral for Autorepo was reduced until the facility was discontinued on October 5.

Since the new framework was introduced, the RBNZ has implemented two changes. First, banks are now allowed to use a wider set of assets to raise cash from the central bank. In particular, a limited amount of AAA-rated paper is eligible.[a] Second, a tiered system of remuneration was introduced in response to episodes in which the market interest rate rose substantially above the OCR. The RBNZ now estimates the quantity of reserves a bank needs for its payment activity and, based on this estimate, sets a limit on the quantity that will be remunerated at the OCR. Any reserves held in excess of that limit earn a rate 100 basis points below the OCR. This policy is designed to provide an incentive for banks to recirculate excess reserve positions and to prevent them from "hoarding" reserves.

In principle, the RBNZ could have addressed this problem by increasing its supply of reserves instead of by implementing a tiered system. If the market interest rate is significantly higher than the policy rate in a floor system, increasing the supply of reserves should drive the market rate down (see Exhibit 3 in the text). However, the RBNZ uses FX swaps to increase the supply of reserves, and it found that the price in this market was moving against it; the more reserves the RBNZ created, the more costly it became to create those reserves. It is worth noting that this problem would not arise in a country with a large supply of government bonds or with a central bank that can issue its own interest-bearing liabilities. In such cases, increasing the supply of reserves need not be costly and could be an attractive alternative to a tiered system.

While it is too early to evaluate with great confidence all of the effects of the RBNZ's changes, it appears that the transition went smoothly overall. There were, of course, occasional signs of stress in money markets, mostly attributable to the learning process experienced by the Bank and its payments system participants. There are, however, definite positive signs that the liquidity of the interbank market has improved. Notably, payments have been settling significantly earlier since the transition began, suggesting a reduction in the constraints previously attributable to the scarcity of collateral available to pledge in the Autorepo facility. In addition, the implied New Zealand dollar interest rates in the FX swap market are now much less volatile and are well within the 50 basis point band between the official cash rate and the ORRF. Finally, the RBNZ conducts open market operations much less frequently, and the operations are no longer subject to the underbidding that had led to excessive use of overnight facilities.

[a] See the Reserve Bank of New Zealand's May 2008 Financial Stability Report for more details.

market for overnight loans of reserves differs from other markets in fundamental ways. As we discussed, reserves are not a commodity that is physically scarce; they can be costlessly produced by the central bank from other risk-free assets. Moreover, there is no role for socially useful price discovery in this market, because the central bank's objective is to set a particular price. Weighing the costs and benefits of a reduction in market activity is therefore a nontrivial task and an important area for future research.

If desired, the floor system could be modified in ways that encourage higher levels of activity in the overnight interbank market. For example, the central bank could limit the quantity of reserves on which each bank earns the target rate of interest and compensate balances above this limit at a lower rate. Such limits would encourage banks that accumulate unusually large balances over the course of the day to lend them out. By setting lower limits, the central bank would encourage more activity in the interbank market while marginally increasing the distortions discussed above.[18] Whitesell (2006b) presents a system in which banks are allowed to determine their own limits by paying a "capacity fee" proportional to the chosen limit. In this case, the central bank would set the fee schedule in a way that balances concerns about the level of market activity with the resulting level of distortions.

> *If desired, the floor system could be modified in ways that encourage higher levels of activity in the overnight interbank market.*

Another interesting issue is the extent to which a floor system would allow the central bank to restrict access to daylight credit, if it so desired. If access to daylight credit is substantially restricted or removed, the smooth functioning of the payments system may require banks to acquire funds in the market on a timely basis during the day. In principle, this could be accomplished by the development of either an intraday market for reserve balances or a market for precise time-of-day delivery of reserves (see McAndrews [2006] for a discussion of such possibilities). Whether such markets would actually develop and how efficiently they would operate are important open questions.

Going forward, the experience of New Zealand's floor system will provide valuable information on these issues and others that might arise. However, the differences between the financial system of New Zealand and those of economies like the United States will make it difficult to draw definite conclusions. For this reason, it is important to employ the tools of modern economic theory to develop models that are capable of addressing these issues.

5. Conclusion

This article highlights the important similarities in the monetary policy implementation systems used by many central banks. In these systems, there is a tight link between money and monetary policy because the supply of reserve balances must be set precisely in order to implement the target interest rate. This link creates tensions with the central bank's other objectives. For example, the intraday need for reserves for payment purposes is much higher than the overnight demand, which has led central banks to provide low-cost intraday loans of reserves to participants in their payments systems. This activity exposes the central bank to credit risk and may generate problems of moral hazard. The link also prevents central banks from increasing the supply of reserves to promote market liquidity in times of financial stress without compromising their monetary policy objectives. Furthermore, the link relies on banks facing an opportunity cost of holding reserves, which generates deadweight losses and hinders the efficient allocation of resources.

Our study also presents an approach to implementing monetary policy in which this link is severed, leaving the quantity of reserves and the interest rate target to be set independently. In this floor-system approach, interest is paid on reserve balances at the target interest rate. This policy allows the central bank to increase the supply of reserves, perhaps even significantly, without affecting the short-term interest rate. While the floor system has received a fair amount of attention in policy circles recently, there are important open questions about how well such a system will work in practice. Going forward, it will be useful to develop theoretical models of the monetary policy implementation process that can address these questions, as well as to observe New Zealand's experience with the floor system it implemented in 2006.

[18] Ennis and Keister (2008) describe a related approach based on "clearing bands," where banks face a minimum requirement and earn the target rate of interest on balances held up to a higher limit. This approach could be used to encourage activity in the interbank market on the borrowing side (by banks that find themselves below the minimum requirement) as well as on the lending side (by banks that find themselves above the higher limit).

References

Bartolini, L., G. Bertola, and A. Prati. 2002. "Day-to-Day Monetary Policy and the Volatility of the Federal Funds Interest Rate." Journal of Money, Credit, and Banking 34, no. 1 (February): 137-59.

Bech, M. L., and B. Hobijn. 2007. "Technology Diffusion within Central Banking: The Case of Real-Time Gross Settlement." International Journal of Central Banking 3, no. 3 (September): 147-81.

Board of Governors of the Federal Reserve System. 2007. "Overview of the Federal Reserve's Payments System Risk Policy on Daylight Credit." Sixth Edition. July. Available at <http://www.federalreserve.gov/paymentsystems/psr/overview.pdf>.

———. 2008a. "Monetary Policy Report to the Congress." February 27. Available at <http://www.federalreserve.gov/boarddocs/hh/2008/february/fullreport.htm>.

———. 2008b. "Request for Comment on the Payments System Risk Policy." Press release dated February 28. Available at <http://www.federalreserve.gov/newsevents/press/other/20080228a.htm>.

Carpenter, S., and S. Demiralp. 2006a. "The Liquidity Effect in the Federal Funds Market: Evidence from Daily Open Market Operations." Journal of Money, Credit, and Banking 38, no. 4 (June): 900-20.

———. 2006b. "Anticipation of Monetary Policy and Open Market Operations." International Journal of Central Banking 2, no. 2 (June): 25-63.

Clouse, J. A., and J. P. Dow, Jr. 2002. "A Computational Model of Banks' Optimal Reserve Management Policy." Journal of Economic Dynamics and Control 26, no. 11 (September): 1787-1814.

Ennis, H. M., and T. Keister. 2008. "Understanding Monetary Policy Implementation." Federal Reserve Bank of Richmond Economic Quarterly 94, no. 3 (summer).

Ennis, H. M., and J. A. Weinberg. 2007. "Interest on Reserves and Daylight Credit." Federal Reserve Bank of Richmond Economic Quarterly 93, no. 2 (spring): 111-42.

Friedman, M. 1959. A Program for Monetary Stability. New York: Fordham University Press.

———. 1969. The Optimum Quantity of Money and Other Essays. Chicago: Aldine Publishing Company.

Furfine, C. H. 2000. "Interbank Payments and the Daily Federal Funds Rate." Journal of Monetary Economics 46, no. 2 (October): 535-53.

Goodfriend, M. 2002. "Interest on Reserves and Monetary Policy." Federal Reserve Bank of New York Economic Policy Review 8, no. 1 (May): 13-29. Proceedings of a Conference Sponsored by the Federal Reserve Bank of New York, April 5-6, 2001.

Guthrie, G., and J. Wright. 2000. "Open Mouth Operations." Journal of Monetary Economics 46, no. 2 (October): 489-516.

Hamilton, J. D. 1997. "Measuring the Liquidity Effect." American Economic Review 87, no. 1 (March): 80-97.

Lacker, J. M. 2006. "Central Bank Credit in the Theory of Money and Payments." Remarks presented at The Economics of Payments II Conference, Federal Reserve Bank of New York, March 29. Available at <http://www.richmondfed.org/news_and_speeches/presidents_speeches/ index.cfm/2006/id=88>.

Lucas, R. E., Jr. 2000. "Inflation and Welfare." Econometrica 68, no. 2 (March): 247-74.

Martin, A. Forthcoming. "Reconciling Bagehot with the Fed's Response to September 11." Journal of Money, Credit, and Banking.

McAndrews, J. J. 2006. "Alternative Arrangements for the Distribution of Intraday Liquidity." Federal Reserve Bank of New York Current Issues in Economics and Finance 12, no. 3, April.

McAndrews, J. J., and S. M. Potter. 2002. "Liquidity Effects of the Events of September 11, 2001." Federal Reserve Bank of New York Economic Policy Review 8, no. 2 (November): 59-79.

Nield, I. 2006. "Changes to the Liquidity Management Regime." Reserve Bank of New Zealand Bulletin 69, no. 4 (December): 26-31.

Nield, I., and J. Groom. 2008. "Implementing Monetary Policy in New Zealand." Unpublished paper, Reserve Bank of New Zealand.

Poole, W. 1968. "Commercial Bank Reserve Management in a Stochastic Model: Implications for Monetary Policy." Journal of Finance 23, no. 5 (December): 769-91.

Thornton, D. L. 2006. "The Daily Liquidity Effect." Federal Reserve Bank of St. Louis Working Paper no. 2006-020A, April.

Whitesell, W. 2006a. "Interest Rate Corridors and Reserves." Journal of Monetary Economics 53, no. 6 (September): 1177-95.

———. 2006b. "Monetary Policy Implementation without Averaging or Rate Corridors." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, no. 2006-22, May.

Woodford, M. 2000. "Monetary Policy in a World without Money." International Finance 3, no. 2 (July): 229-60.

Empirical Analyses of Trends in Large-Value Payments

Articles by:

Morten L. Bech, Christine Preisig, and Kimmo Soramäki

Olivier Armantier, Jeffrey Arnold, and James McAndrews

Christopher Becher, Marco Galbiati, and Merxe Tudela

Morten L. Bech, Christine Preisig, and Kimmo Soramäki

Global Trends in Large-Value Payments

- The evolving landscape in which large-value payments systems (LVPSs) operate is having important effects on the financial system.
- An analysis of the current interbank payment environment points to three forces that are shaping ten trends common to LVPSs around the world.
- *Technological innovation* is making LVPSs safer and more efficient while allowing for new systems that are not limited to one country or currency.
- *Structural changes in banking*—such as immense growth in the financial sector, changes in the role of firms and their products, and greater globalization of financial institutions and their services—are influencing the use of LVPSs.
- *The evolution of central bank policies* is resulting in central banks becoming more active in monitoring existing and planned systems, assessing systems according to international standards, and inducing change.

1. Introduction

Globalization and technological innovation are two of the most pervasive forces affecting the financial system and its infrastructure. Perhaps nowhere are these trends more apparent than in the internationalization and automation of payments. The evolving landscape is most obvious in retail payments. The use of paper checks is in rapid decline or has been eliminated in most of the industrialized world. Credit and debit cards can be used in the most surprising places. Internet banking with money transfer capabilities is common, and several providers are competing to service consumers' payments over the Internet and mobile devices.

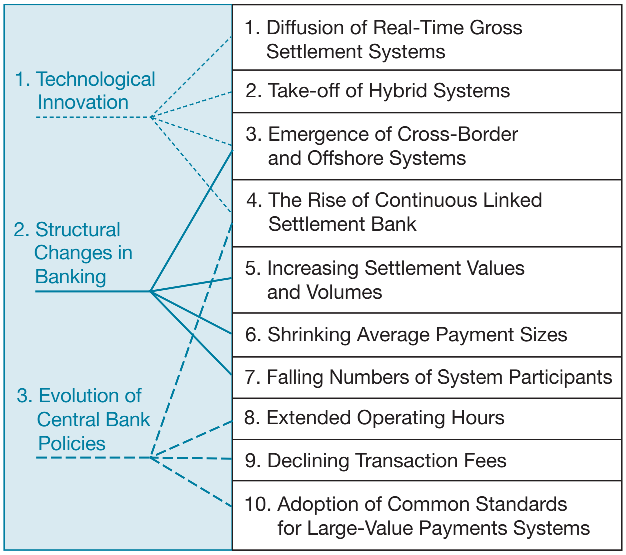
In wholesale, or interbank, payments, the effect of globalization and technological innovation is probably less obvious to the casual observer—but it has been equally impressive. Given the importance of payments and settlement systems to the smooth operation as well as resiliency of the financial system, stakeholders need to understand and assess the potential consequences of this evolution. This article offers an in-depth look at the current environment for large-value payments systems (LVPSs). We describe ten trends common to LVPSs around the world and identify the key drivers of these developments and the most important policy issues facing central banks (see box). Furthermore, we provide empirical support for each of the trends by using numerous publicly available sources, including Bank for International Settlements

Morten L. Bech is a senior economist at the Federal Reserve Bank of New York; Christine Preisig is a former senior policy expert at the Bank; Kimmo Soramäki is a Ph.D. candidate at Helsinki University of Technology on leave from the European Central Bank. This article was written while Mr. Soramäki was on leave from the European Central Bank and acting as a research consultant to the Federal Reserve Bank of New York.

Correspondence: <morten.bech@ny.frb.org>

The views expressed are those of the authors and do not necessarily reflect the position of the European Central Bank, the Federal Reserve Bank of New York, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Ten Global Trends in Large-Value Payments and Their Key Drivers



(BIS) statistics on payments and settlement systems in selected countries (the "Red Book"). We focus on large-value payments systems in countries where the central bank is a member of the Committee on Payment and Settlement Systems (CPSS), a body under the auspices of the BIS (Appendix A).

Technological innovation, structural changes in banking, and the evolution of central bank policies are the three main reasons for the recent developments in large-value payments. First, technological innovation has created opportunities to make existing large-value payments systems safer and more efficient. Such innovation has also accommodated the industry's growing need for new types of systems that are not limited to a single country or a currency. Second, the financial sector has experienced immense growth over the last few decades accompanied by changes in the role of individual firms and the products they offer. In addition, financial institutions and their services have become increasingly globalized. These structural changes have affected how participants use large-value payments systems. Third, the role of central banks in large-value payments systems has changed significantly in recent years. Central banks have become more involved in payments systems and have created formal and systematic oversight functions. The main focus lies in promoting safety and efficiency in LVPSs and in maintaining overall financial stability. Central banks therefore have taken more active roles in monitoring existing and planned systems, in assessing systems according to international standards, and, if necessary, in inducing change.

As the box illustrates, the ten trends that we describe can be assigned to three key drivers. The first four trends—the diffusion of real-time gross settlement (RTGS) systems, the take-off of hybrid systems, the emergence of cross-border and offshore systems, and the rise of Continuous Linked Settlement (CLS) Bank—are all associated with settlement technology and fall into the first category. Technological innovation has enabled new settlement methodologies to emerge that allow a better balance between settlement risks, immediacy, and liquidity requirements. RTGS systems have to a large extent replaced deferred net settlement (DNS) systems. However, the high liquidity needs associated with RTGS have led some system operators to explore liquidity-saving mechanisms and have motivated them to develop hybrid systems. Developments in payments system technology have also facilitated the emergence of systems that settle payments across national borders in one or more currencies. In addition, the clearing of payments is in some instances moving offshore and the ability of participants to connect remotely—eliminating the need for a physical "footprint" in the jurisdiction of LVPSs—is becoming more widespread. Foreign exchange (FX) settlement and counterparty risk are being managed more tightly in part because of the use of payment-versus-payment (PvP)

Technological innovation, structural changes in banking, and the evolution of central bank policies are the three main reasons for the recent developments in large-value payments.

mechanisms.[1] CLS Bank operates a multicurrency payments system for the simultaneous settlement of both sides of a foreign exchange transaction on a PvP basis. With CLS Bank, existing risks associated with FX trades are virtually eliminated.

The next three trends—increasing settlement values and volumes, shrinking average payment sizes, and falling numbers of system participants—as well as the emergence of cross-border and offshore systems (Trend 3) fall into the second category. They are determined largely by how the banking sector uses payments systems and by the structural changes taking place therein. The values and volumes originated over LVPSs grew exponentially until the turn of the century. However, in terms of value, growth has since slowed and is no longer outpacing economic growth as measured by GDP.

[1] PvP ensures that a final transfer of one currency occurs only if a final transfer of the other currency or currencies takes place (Bank for International Settlements 2003).

Because many LVPSs process a large amount of relatively low-value payments, the average payment size settled has shrunk. Hence, the dichotomy between small- and large-value payments systems is not always applicable. In addition, consolidation in the banking sector has led to fewer participants in LVPSs. Structural changes have also resulted in the emergence of global banks that require a global payment infrastructure, which in turn has led to the creation of new systems that accommodate these needs.

The last three trends and the rise of CLS Bank (Trend 4) fall into the third category. They are associated with central banks' operating policies regarding LVPSs. The service level of all systems is improving with longer operating hours. Some systems are even approaching a twenty-four-hour settlement cycle. Transaction costs in various LVPSs have been falling since the late 1990s because the savings achieved through improvements in operating efficiency have been passed on to system participants in the form of lower fees. Through the adoption of common standards, such as the CPSS' Core Principles for Systemically Important Payments Systems, risk management in LVPSs has become more standardized. Furthermore, the central bank community was the driving force behind the development of CLS Bank.

We now describe each trend in detail and conclude by commenting on the possible future of the large-value payments landscape.

2. Trend 1: Diffusion of Real-Time Gross Settlement Systems[2]

As a consequence of the rapid increase in values settled in large-value payments systems in the 1980s, central banks became concerned about settlement risks inherent in the then-prevalent deferred net settlement systems.[3] In particular, the banks were concerned about the potential for contagion (or even a systemic event) attributable to the unwinding of the net positions that would result if a participant failed to make good on its obligations when due.[4]

Over the last few decades, many countries have chosen to modify the settlement procedure employed by their interbank payments system with a view to reducing settlement risks and the potential for adverse systemwide implications. Most central banks have opted for the implementation of an RTGS system. Such a system reduces settlement risk, as payments are settled individually and irrevocably on a gross basis in real time, ensuring immediate finality. RTGS can also help reduce settlement risk by facilitating payment versus payment and delivery versus payment in the settlement of FX and securities transactions, respectively.

Fedwire is the world's oldest RTGS system. Its origins can be traced to 1918, when the Federal Reserve inaugurated a network of wire communications among the individual Reserve Banks. In the early 1970s, the Fedwire system migrated to a fully computerized platform, and settlement in "real time" was achieved.

Over the last few decades, many countries have chosen to modify the settlement procedure employed by their interbank payments system with a view to reducing settlement risks and the potential for adverse systemwide implications. Most central banks have opted for the implementation of an RTGS [real-time gross settlement] system.

A number of western European countries began implementing RTGS systems in the 1980s. By 1988, RTGS systems operated in four of the six major currencies. RTGS adoption continued at a rate of roughly one country per year during the early 1990s. In 1992, the Treaty of Maastricht created the foundation for the Economic and Monetary Union (EMU). A year later, the central banks in the European Union (EU) agreed that each member state should have an RTGS system. Furthermore, in 1995 it was decided to interlink the national RTGS systems through the Trans-European Automated Real-time Gross settlement Express Transfer (TARGET) system to facilitate the European Central Bank's (ECB) single monetary policy and to promote sound and efficient payment mechanisms in euros. This decision led to a flurry of new systems and upgrades to existing ones. TARGET went live on January 4, 1999, and even EU countries that did not join the EMU at the outset (the United Kingdom, Denmark, and Sweden) were allowed to participate in the

[2] For more on these systems, see Bech (2007) and Bech and Hobijn (2007).

[3] A DNS system effects the settlement of obligations or transfers between or among counterparties on a net basis at some later time (Bank for International Settlements 2003).

[4] Unwinding is a procedure followed in certain clearing and settlement systems in which transfers of securities and funds are settled on a net basis, at the end of the processing cycle, with all transfers provisional until all participants have discharged their settlement obligations. If a participant fails to settle, some or all of the provisional transfers involving that participant are deleted from the system and the settlement obligations from the remaining transfers are then recalculated. Such a procedure has the effect of allocating liquidity pressures and losses attributable to the failure to settle to the counterparties of the participant that fails to settle (Bank for International Settlements 2003).

system. As the ECB made RTGS a prerequisite for membership in the EMU, prospective members in the rest of Europe began to implement RTGS. Furthermore, as hostilities ended in the Balkans in the late 1990s, governments began to rebuild their respective economies. They considered the establishment of sound and efficient financial systems a priority. RTGS systems were implemented with support from the EU, the International Monetary Fund (IMF), and the World Bank. With ongoing projects in Russia and Cyprus, the diffusion of RTGS in Europe is nearly completed.

Outside Europe, the rate of RTGS adoption since the mid-1990s has been equally impressive. Australia and New Zealand implemented RTGS in 1998. In Asia, the rate of implementation has been fairly steady; on average, about one country per year has adopted RTGS. Six countries in the Middle East have done likewise. In Africa, the South African Reserve Bank (SARB) spearheaded RTGS adoption in 1998. Through the South African Development Community (SADC),[5] SARB has participated in developing and strengthening the financial infrastructure in the rest of southern Africa. As of 2006, eleven African central banks have implemented RTGS, many with the support of the World Bank.

In the Western Hemisphere, Canada is the only Group of Ten (G-10) country that has decided not to implement an RTGS system. Instead, Canada opted for a hybrid system.[6]

Uruguay was the first country in South America to adopt RTGS in 1995. By 2006, seven of thirteen South American countries had followed suit. Implementation in Central America and the Caribbean has started only recently, but the Inter-American Development Bank is assisting RTGS efforts in the region.

The global diffusion of RTGS systems since the mid-1980s is evident from Exhibit 1. By 1985, three central banks—the Federal Reserve, Danmarks Nationalbank, and the Netherlandsche Bank—had implemented RTGS systems. A decade later, that number had increased to sixteen, but RTGS was still utilized predominantly by industrialized countries. In recent years, however, transitional as well as developing countries have begun investing heavily in improving their financial systems, and now RTGS is a common choice for interbank payments. At the end of 2006, 93 of the world's 174 central banks were using RTGS systems. The RTGS adoption rate was about one central bank per year in the latter part of the 1980s and the beginning of the 1990s. In the mid-1990s, the rate accelerated with the addition of three central banks in 1995 and five in 1996. Since then, the annual adoption rate has not dipped below three new central banks. It peaked in 2002, when a total of fifteen central banks implemented new RTGS systems.

3. Trend 2: Take-off of Hybrid Systems

As payments are settled individually in an RTGS system, sufficient liquidity needs to be available to fund each payment. Real-time gross settlement thus reduces settlement risks but results in an increased need for intraday liquidity to smooth nonsynchronized payment flows.[7]

Initially, central banks provided intraday credit free to commercial banks. This policy is no longer considered a viable option by the banks, as it exposes them (and ultimately taxpayers), as guarantor of the finality of payments, to credit risk (see, for example, Humphrey [1986] and Bech and Soramäki [2005]). Thus, intraday liquidity is costly for participants either in the form of explicit fees or implicitly as the opportunity cost of collateral that participants need to pledge for an intraday credit line from the central bank. To reduce the need for intraday liquidity, several systems have developed different types of queue management and liquidity-saving features. The queue management features include different priority categories for payments and the possibility of reordering payments once in the queue. The liquidity-saving

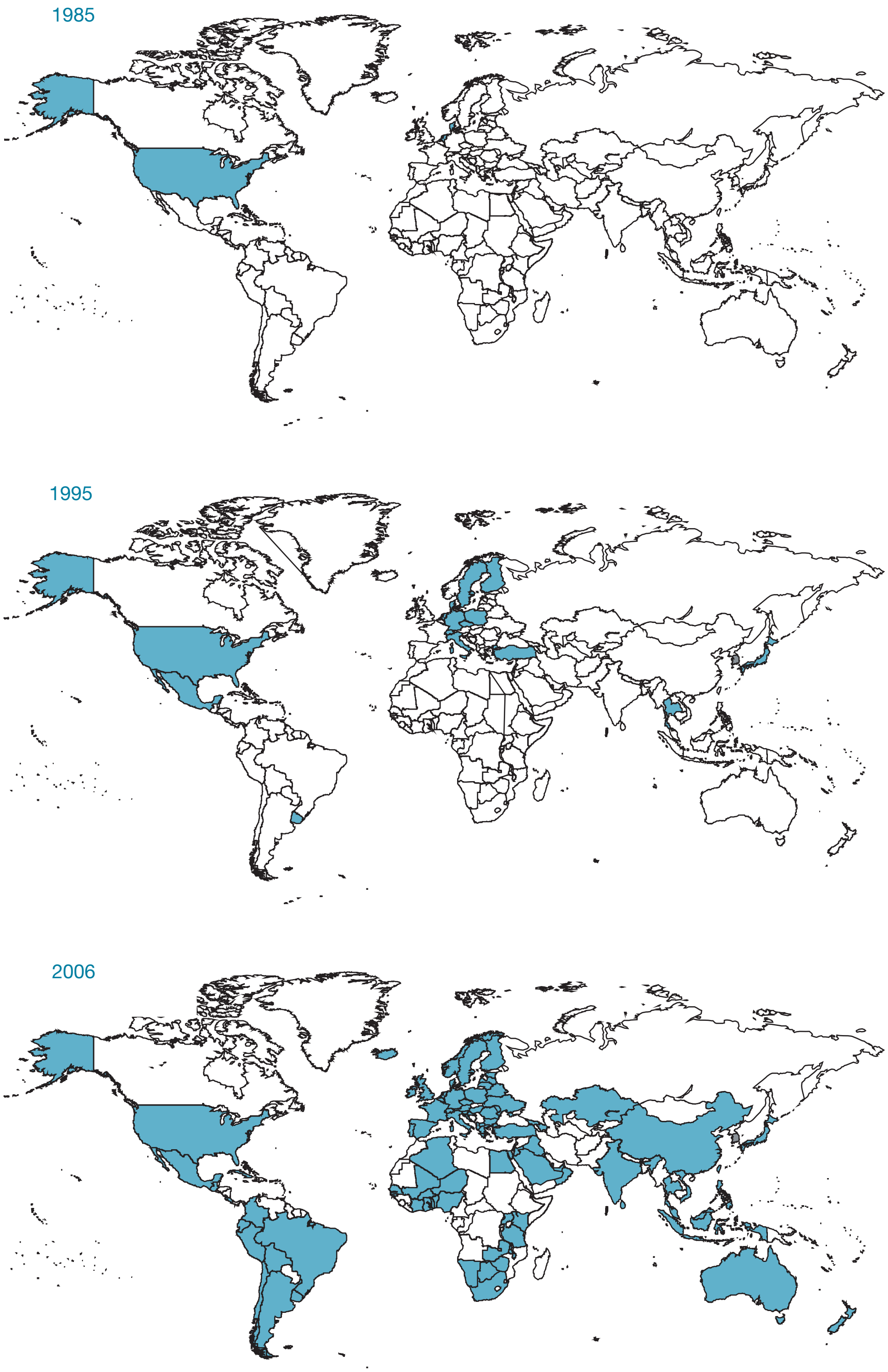
To reduce the need for intraday liquidity, several systems have developed different types of queue management and liquidity-saving features.

[5] The member states of the SADC are Angola, Botswana, Democratic Republic of Congo, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia. and Zimbabwe.

[6] The Canadian Large-Value Transfer System (LVTS) processes payments with finality in real time, while settlement occurs on a multilateral net basis at the end of the day. Immediate intraday finality is achieved because settlement is guaranteed under all circumstances. This is facilitated by the use of collateral to secure participants' intraday net debit (negative) positions and by a residual guarantee provided by the Bank of Canada (see Arjani and McVanel [2006]). LVTS is considered equivalent to RTGS in terms of finality, as the Bank of Canada provides an explicit guarantee of settlement in case of participant failure.

[7] Payments systems can operate at different levels of liquidity requirements, delays, and risks. Depending on the system design, these can be traded off against each other. Liquidity requirements are highest when payments are settled continuously against full cover (as in RTGS). Liquidity requirements in an RTGS system can be reduced by delaying payments until incoming payments allow settlement. In deferred net settlement systems, payments are settled only periodically by transferring only net amounts, thus payments are delayed from time of receipt until time of settlement. In such systems, banks can reduce delays by crediting customer accounts before final settlement. This will, however, come at the expense of credit risks, as final settlement may not take place as expected. For an in-depth discussion of these trade-offs, see Leinonen and Soramäki (1999).

Exhibit 1
Diffusion of Real-Time Gross Settlement Systems Worldwide

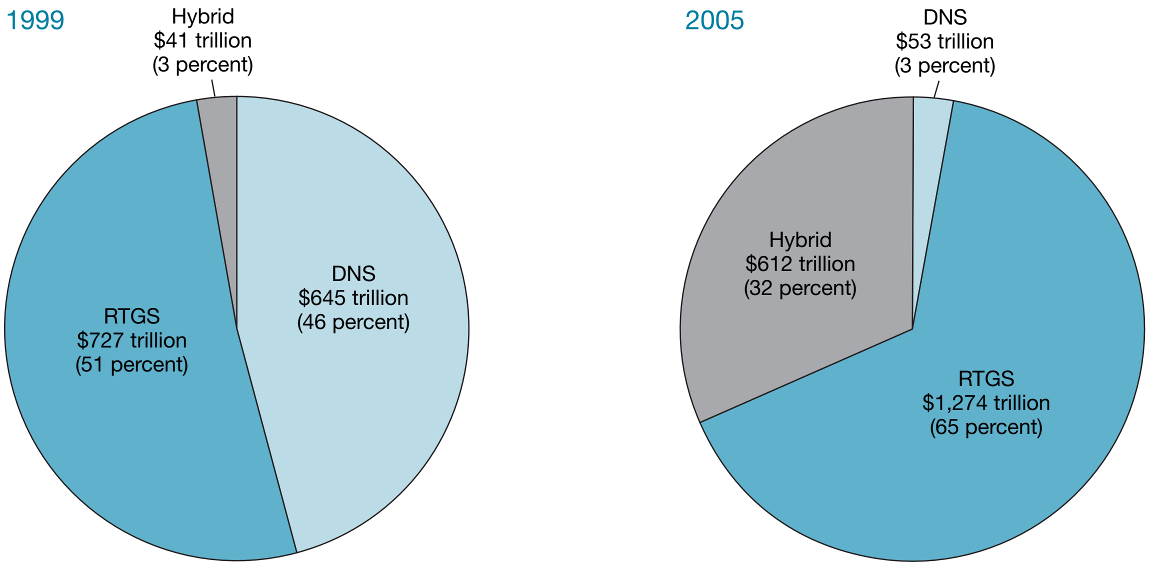


Source: Bech and Hobijn (2007).

Chart 1

Payments Settled in RTGS, DNS, and Hybrid Modes, 1999 and 2005

U.S. Dollar Equivalent



Sources: Bank for International Settlements; authors' calculations.

Notes: RTGS is real-time gross settlement; DNS is deferred net settlement. Figures represent Fedwire, CHIPS, CHAPS, SIC, TARGET, MEPS, K-RIX, LVTS, PNS, Euro1, and HKD CHATS. DNS in 1999: Euro1 and BoJ-NET. Hybrids in 2005: PNS, LVTS, CHIPS, and RTGSplus. Systems are described in Appendix A. The value for RTGSplus was subtracted from the value for TARGET in computing the value settled via RTGS.

features typically involve the netting of payments in the queue, a feature commonly referred to as *gridlock resolution* (see Bech and Soramäki [2001]).[8]

Another approach to balancing risks, payment delays, and liquidity needs more efficiently has been the development of hybrid systems (see McAndrews and Trundle [2001] and Leinonen and Soramäki [1999]). Hybrid systems employ advanced settlement algorithms that combine components of both net and real-time gross settlement. Some payments may be settled individually, as in RTGS, while others, usually less urgent payments, may be pooled together and netted.[9] Other features may include bilateral limits to manage credit exposures and reciprocity. The distinction between RTGS systems and hybrid systems can be fluid. In this article, hybrid systems are defined as systems with either separate payment streams for urgent or nonurgent payments and/or systems that employ advanced bilateral or multilateral offsetting algorithms on a continuous basis. Prominent examples of hybrid systems include RTGSplus in Germany, LVTS in Canada, CHIPS in the United States, and Paris Net Settlement (PNS) in France.

With regard to settlement method, the major development in payments systems in CPSS countries has been the sharp increase in value settled by hybrid systems (Chart 1). In 1999, 3 percent of payment value was settled by these systems. RTGS settlement accounted for approximately 50 percent of payments and DNS settlement for roughly 45 percent.[10] By 2005, hybrid systems accounted for close to one-third of value settled, whereas RTGS increased to almost two-thirds. The only remaining DNS system, EURO1, accounted for less than 3 percent of total value settled in CPSS large-value payments systems.

[8] Gridlock is a situation that can arise in a funds or securities transfer system when the failure of certain transfer instructions to be executed (because the needed funds or securities balances are unavailable) prevents the execution of a substantial number of instructions from other participants (Bank for International Settlements 2003).

[9] For example, the RTGS share of the German hybrid system RTGSplus is about 10 percent, whereas the netted payments share is 90 percent.

[10] Prior to 2001, BoJ-NET provided both DNS and RTGS settlement modes. However, the RTGS settlement mode was seldom used by banks because of its higher liquidity costs. In 2001, the Bank of Japan introduced a reconfigured BoJ-NET and abolished DNS. Just before the change, only 3 percent of Japan's wholesale payments were settled via RTGS (Selgin 2004).

4. Trend 3: Emergence of Cross-Border and Offshore Payments Systems

Large-value payments systems have traditionally settled payments in the local currency among participants located within the same national borders as the system. However, since the late 1990s, systems have emerged that allow payments to settle across national borders, facilitate settlement in multiple currencies, and permit participants to be located in a foreign jurisdiction. The two key drivers of the development of cross-border and offshore systems have been the introduction of the euro and the demand for payment settlement in foreign currencies as part of payment-versus-payment FX transactions, particularly in the Asian time zones.

To analyze these new types of systems, we classify them according to the location of their participants and the currencies in which they settle payments. A simple taxonomy is presented in Table 1. The type of participants and currencies that a system services are each divided into three groups. Participants are categorized as *domestic, remote*, or *cross-border*, whereas the currencies settled are classified as *local*, *foreign*, or *multiple*.

We refer to participants located in the same country as the system as *domestic* participants. A *remote* participant has neither its head office nor any of its branches located in the country where the transfer system is based (Bank for International Settlements 2003). *Cross-border* participants are payees and payers located in different countries. The groups are not exclusive, as systems with remote or cross-border participants also service domestic participants. Further on, we distinguish between systems that settle in the *local currency* of the country where the system is located, systems that settle in a single *foreign currency*, and systems that settle in a set of *multiple currencies* that typically includes the local currency.

Table 1
Taxonomy of Payments Systems

	Settlement Currency		
Participant Type	Local	Foreign	Multiple (PvP in Foreign Exchange Settlement)
Domestic	Fedwire, CHIPS, CHAPS Sterling, LVTS, RIX, PNS, BoJ-NET	USD CHATS, EUR CHATS	HKD, USD, and EUR CHATS (PvP arrangement in CHATS)
Remote	SIC	Euro-SIC	—
Cross-border	TARGET, EURO1	CHAPS Euro	CLS Bank, RENTAS-CHATS link

Notes: Systems are described in Appendix A. PvP is payment-versus-payment.

Fedwire and CHIPS in the United States, CHAPS in the United Kingdom, and BoJ-NET in Japan, to name a few, are traditional large-value payments systems that settle payments in the local currency for domestic participants.

Originally, participation in the Swiss Interbank Clearing (SIC) system was limited to banks domiciled in Switzerland and the Principality of Liechtenstein. However, since 1998 remote access to SIC has been granted to banks domiciled outside Switzerland. Any bank worldwide can participate in SIC as long as it meets the admission criteria (see Heller, Nellen, and Sturm [2000]). Among other things, remote access has allowed foreign banks that participate in the futures and options exchange EUREX to process Swiss franc transactions directly via SIC without having a physical "footprint" in Switzerland. At the end of 2006, of the 331 SIC participants, 72 were so-called *remote* members.

The two key drivers of the development of cross-border and offshore systems have been the introduction of the euro and the demand for payment settlement in foreign currencies as part of payment-versus-payment FX transactions, particularly in the Asian time zones.

In 1999, with the introduction of the euro, two pan-European interbank payments systems were introduced. TARGET and EURO1 both settle cross-border payments in euros. Currently, TARGET consists of 17 RTGS systems with 1,058 direct participants[11] (see Appendix B for more on TARGET and its individual components). In 2006, TARGET processed on average 326,000 daily payments worth about 2.1 trillion euros (see Appendix C for more on cross-border payments within TARGET).

In conjunction with EU central banks' efforts, the European Banking Association (EBA) established, much like CHIPS, a private sector complement to TARGET: the EURO1 system. Today, the system has 70 participating banks and processes on average 185,000 payments a day with a total value of around 195 billion euros. TARGET and EURO1 are examples of LVPSs that settle cross-border payments in a local currency.

[11] The systems are Belgium ELLIPS (Belgium), KRONOS (Denmark), RTGSplus (Germany), HERMES (Greece), SLBE (Spain), EP RTGS (Estonia), TBF (France), IRIS (Ireland), New BIREL (Italy), LIPS-Gross (Luxembourg), TOP (the Netherlands), Artis (Austria), SORBNET-EURO (Poland), SPGT (Portugal), Payments System (Slovenia), BoF-RTGS (Finland), and CHAPS (United Kingdom).

The United Kingdom, Sweden, and Denmark did not join the EMU at the outset. However, in expectation that these countries would eventually join, separate euro RTGS systems were built and connected to TARGET: CHAPS Euro in the United Kingdom, E-RIX in Sweden, and DEBES in Denmark. From the perspective of the Table 1 taxonomy, these systems can be thought of as facilitating cross-border payments in a foreign currency. However, with the advent of the next generation of the TARGET system, the euro functionality of all three systems will be phased out. E-RIX already closed at the end of 2006.

Another implication of the common currency was the introduction of euroSIC in 1999. The system allows Swiss banks to conduct euro transactions. It operates on the same platform as the Swiss franc system, but settlement takes place on the books of the Swiss Euro Clearing Bank (SECB) in Frankfurt, Germany. SECB provides a link to the euro area by being a direct participant in RTGSplus, through which access to TARGET is established.[12] In addition, remote access is possible in euroSIC. Within the taxonomy, euroSIC is a system that settles a foreign currency for domestic and remote participants.

CLS Bank operates a system that settles multiple currencies for participants located in different countries (we discuss it in more detail as part of Trend 4). A similar system is the link between the ringgit real-time gross settlement system in Malaysia (the RENTAS system) and the U.S. dollar real-time gross settlement system in Hong Kong (USD CHATS). The link, established by the Hong Kong Monetary Authority and Bank Negara Malaysia in 2006, allows for PvP foreign exchange settlement of USD-MYR trades. In the Table 1 taxonomy, the link can be considered a cross-border PvP system.

Since 2000, several new systems that settle a foreign currency for participants located within the same national borders as the system itself have emerged:

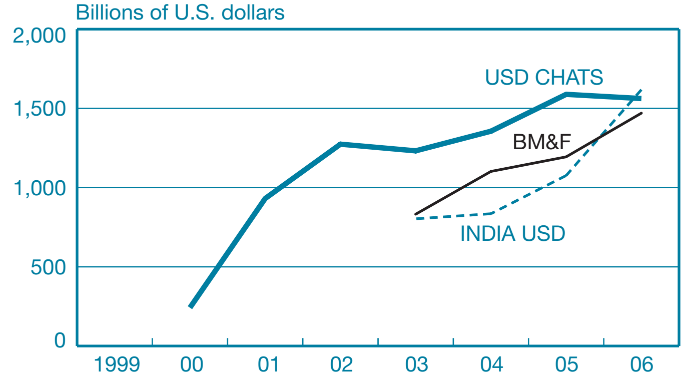
- In 2000, Hong Kong Interbank Clearing Ltd. (HKICL) introduced the U.S. Dollar Clearing House Automated Transfer System (USD CHATS) to clear USD payments in the Asian time zone. The system is technically the same as the Hong Kong dollar CHATS system, except that settlement currently takes place on the books of a private bank (HSBC) in New York City.
- In April 2002, the Brazilian Bolsa de Mercadorias & Futuros (BM&F) FX Clearinghouse initiated operations for settling USD-BRL trades. The Clearinghouse maintains a settlement account in the local currency with the Central Bank of Brazil and settlement accounts in the foreign currency with correspondent banks abroad. The transfer of funds takes place within the same settlement window.
- In April 2003, HKICL introduced EUR CHATS for the settlement of euro payments. The system is similar to USD CHATS, with the settlement bank being Standard Chartered in London. A novel feature of the combined CHATS systems is that it allows for PvP settlement of USD-HKD and EUR-HKD as well as USD-EUR foreign exchange trades. In the Table 1 taxonomy, this system enables the settlement of multiple currencies for domestic system participants.
- In August 2003, Clearing Corporation of India, Ltd. (CCIL), introduced a system that clears and settles interbank FX trades, including Indian rupee (INR) and USD. CCIL is a third-party member of CLS Bank and currently uses ABN AMRO in New York as its settlement bank.[13]

Many of these systems can be understood as substitutes for traditional correspondent banking services. The systems offer a customized local service in the native language, operating hours that accommodate their customers' needs, and potentially better risk management through the use of more formal rules and procedures.

We refer to systems that settle a foreign currency for domestic, remote, or cross-border participants as *offshore systems*. Offshore systems settling USD have experienced strong growth rates, albeit from a low initial level (Chart 2). All three USD offshore systems settled about USD 1,500 billion

Chart 2

Annual Value Settled by U.S. Dollar Offshore Systems, 1999-2006



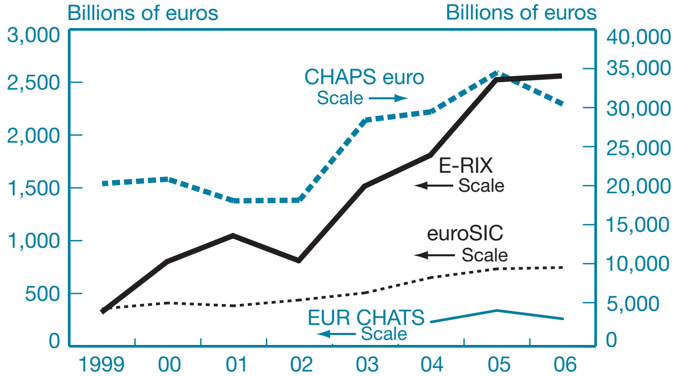
Sources: Hong Kong Monetary Authority; Clearing Corporation of India, Limited; Bolsa de Mercadorias & Futuros.

[12] Interestingly, the largest Swiss bank, UBS AG, processes its cross-border euro payments via proprietary access to RTGSplus and not via euroSIC (see the interview with Stephan Zimmerman, Head of Operations at UBS AG, at <http://www.sic.ch/dl_tkicch_clearit14interview.pdf>).

[13] Other offshore USD systems in operation are the Moscow Interbank Currency Exchange in Russia and the Philippine Domestic Dollar Transfer System.

Chart 3

Annual Value Settled by Euro Offshore Systems, 1999-2006



Sources: Bank for International Settlements; authors' calculations.

in 2006. In euro offshore systems, growth has been more modest—especially considering the fact that the fastest-growing systems, CHAPS Euro and E-RIX, have been discontinued (Chart 3). All of the offshore systems (excluding CHAPS Euro), however, are still of smaller orders of magnitude than the smallest domestic LVPS in the CPSS countries.

5. Trend 4: The Rise of Continuous Linked Settlement Bank

Traditionally, foreign exchange settlement was carried out bilaterally between trade parties through the use of correspondent banking arrangements.[14] Such arrangements lead to exposures because there is no direct link between the payment of the two currency legs; thus, there is a risk (called *Herstatt risk*) of paying the currency sold but not receiving the currency bought. This risk, combined with the vast size of daily FX trading and the global interdependence of FX markets and payments systems, raised concerns among central banks.

In March 1996, the BIS issued a report titled "Settlement Risk in Foreign Exchange Transactions." The report analyzed in particular the risks associated with FX settlement operations and outlined a strategy for reducing them. Based on this report, the G-10 central banks endorsed a three-track strategy to reduce these risks. First, individual banks were requested to take measures to control their FX settlement exposures by improving their practices for measuring and managing exposures. Second, industry groups were encouraged to develop well-constructed multicurrency services that would contribute to the risk-reduction efforts of individual banks. Third, central banks committed themselves to encouraging and fostering private sector development in this field. They also agreed to improve national payments systems to facilitate private sector risk-reduction efforts.

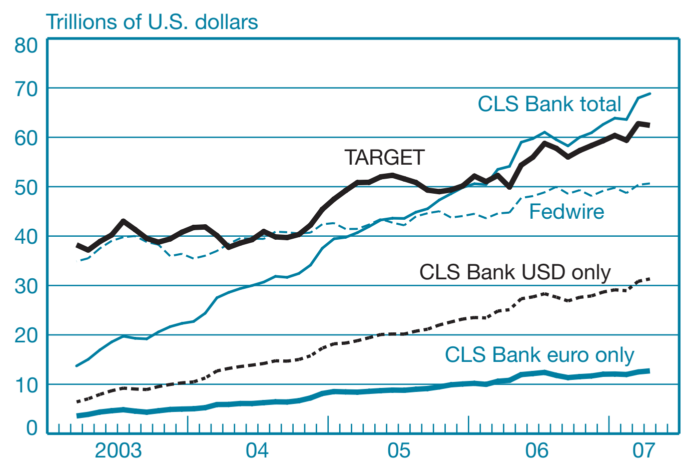
Since 1996, there has been considerable progress on all three tracks of the strategy. In particular, the launch of CLS Bank in September 2002 was a significant move forward by the industry to reduce foreign exchange settlement risk. CLS Bank operates a multicurrency system that settles payments for participants located on every continent. As transactions are settled on a PvP basis, Herstatt risk associated with these trades is virtually eliminated. CLS Bank is a special-purpose U.S. bank supervised by the Federal Reserve and under the cooperative oversight of the central banks of the fifteen currencies included in the system. It has grown extremely fast and is on par with Fedwire and TARGET in terms of value settled (Chart 4).

On March 19, 2008, CLS Bank settled a record 1,113,464 payment instructions with a gross value of USD 10.3 trillion. CLS Bank has grown steadily since its inception as a result of increasing FX volumes, new currencies, and greater market penetration. It now settles more than USD 200 trillion per quarter and had surpassed both Fedwire and TARGET in terms of value settled in the second half of 2005. Chart 4 also displays the euro and USD values settled in CLS Bank. Whereas TARGET

Chart 4

Value Settled in CLS Bank Compared with Fedwire and TARGET, 2003-06

Three-Month Moving Average



Sources: Bank for International Settlements; authors' calculations.

[14] Correspondent banking is an arrangement under which one bank (correspondent) holds deposits owned by other banks (respondents) and provides payment and other services to those respondent banks (Bank for International Settlements 2003). Correspondent relationships are sometimes formalized; the benefit is that parties to a trade do not have to know each other because the system's rules replace individual agreements between them.

Table 2

Total Foreign Exchange Obligations Settled, by Method, 1997 and 2006

Settlement Method	Value (Trillions of U.S. Dollars)	Percentage of Total Value 1997	Percentage of Total Value 2006
CLS Bank (PvP)	2,091	0	55
Traditional correspondent banking	1,224	85	32
Bilateral netting	304	15	8
Other	203		5
Total	3,821	100	100

Source: Bank for International Settlements (2007b).

Note: PvP is payment-versus-payment.

settled five times the euro value settled in CLS Bank, Fedwire settled about 1.7 times the USD value settled in CLS Bank.

Ten years after the strategy was launched, the Committee on Payment and Settlement Systems conducted a survey to assess the extent to which systemic risk had been reduced (Bank for International Settlements 2007b). The survey included 109 institutions that represent 80 percent of the FX market in the fifteen currency areas.[15]

The institutions reported average daily gross values of FX settlement obligations totaling USD 3.8 trillion (Table 2). Of these obligations, 32 percent (USD 1.2 trillion) were settled by traditional correspondent banking arrangements and are still subject to settlement risk. Nonetheless, this is a significant improvement from the time of the 1997 survey, when an estimated 85 percent of the obligations were settled by this method. The major reason for the decline in FX obligations subject to settlement risk is the increasing use of CLS Bank. In 2006, CLS Bank settled 55 percent (USD 2.1 trillion) of the total FX settlement obligations of the surveyed institutions. Furthermore, 8 percent of FX obligations were settled by bilateral netting.[16] Other settlement methods, such as the PvP arrangement available in Hong Kong's USD CHATS and EUR CHATS systems or on-us settlement,[17] accounted for the remaining 5 percent of obligations.

[15] The survey updated and extended previous CPSS surveys conducted in 1996 (Bank for International Settlements 1996) and 1997 (Bank for International Settlements 1998).

[16] Provided it is conducted under legally robust arrangements, bilateral netting can also be a safe and efficient method for reducing settlement exposures (Bank for International Settlements 2007b).

[17] In on-us settlement, both legs of an FX trade are settled on the books of a single institution (Bank for International Settlements 2007b).

Even though major progress has been made in reducing aggregate foreign exchange settlement exposures, the size and duration of the exposures still settled by traditional correspondent banking are significant. In its assessment, the CPSS recommends actions to address the remaining exposures that may continue to present systemic risk.

6. Trend 5: Increasing Settlement Values and Volumes

Available data suggest that the values transferred over large-value payments systems in Committee on Payment and Settlement Systems countries grew substantially during the 1980s and 1990s. Table 3 summarizes the annual growth in nominal payment values in local currencies since 1985 for four systems for which such long-ranging data are available. In the United States, the number of transfers originated over both Fedwire and CHIPS grew by 6 percent per year on average from 1985 to 2000. In terms of value, turnover increased by an average of 9 percent per year. Since 2000, both systems have experienced a slight slowdown in growth, with volumes increasing by 4 percent and 5 percent and values by 7 percent and 9 percent, respectively.

Table 3

Annual Turnover Growth for Selected Large-Value Payments Systems, Nominal Values
Percent

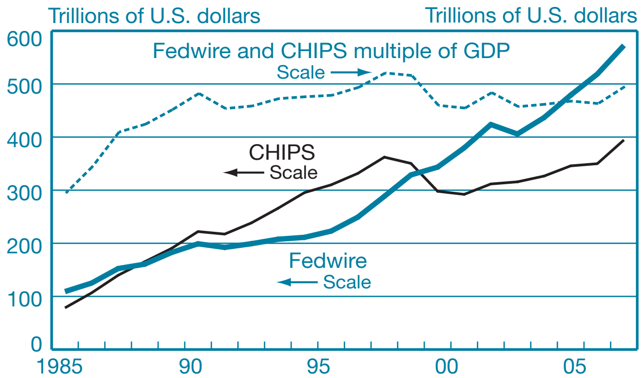
	Average Annual Growth			
	CHAPS	Fedwire	CHIPS	SIC
	Volume			
1985-2006	14	6	6	11
1985-2000	16	6	6	9
2000-06	7	4	5	13
	Value			
1985-2006	17	8	8	2
1985-2000	22	9	9	4
2000-06	3	7	5	0

Sources: CHIPS; Federal Reserve Bank of New York; Bank of England; Swiss National Bank; authors' calculations.

Notes: Systems are described in Appendix A. For SIC, the growth rates are calculated using data from 1989 to 2006.

Chart 5

Time Series of Value of Transfers in CHIPS and Fedwire since 1985



Sources: CHIPS; Federal Reserve Bank of New York; authors' calculations.

In the United Kingdom, CHAPS saw double-digit turnover growth in terms of volumes and values from 1985 to 2000. However, since the turn of the century, growth has decreased. In Switzerland, the picture is a bit different, as volumes have been growing faster since 2000. If growth is measured in Swiss francs, though, turnover since 2000 has been almost flat, at around CHF 45 trillion.

Chart 5 shows the annual turnover for Fedwire and CHIPS since 1985. Fedwire turnover increased from around USD 100 trillion in 1985 to more than USD 570 trillion in 2006. Values settled in CHIPS were smaller than those settled in Fedwire from 1985 to 1988. However, between 1988 and 1998, settlement values in CHIPS surpassed those in Fedwire. After settlement values in CHIPS dipped in the late 1990s,[18] they have steadily risen again and amounted to USD 395 trillion in 2006.

A major determinant of the value of interbank payments is general economic activity. Since 1990, the combined settlement value on Fedwire and CHIPS has kept pace with economic activity, at around seventy times GDP (Chart 5, right axis). However, from 1985 to 1990, combined turnover on the two systems rose from forty-five times GDP to almost seventy-five times GDP. Likely explanations are technological development, deregulation of financial markets, and innovation in financial instruments.[19]

[18] This point is addressed in more detail in our Trend 9 discussion.

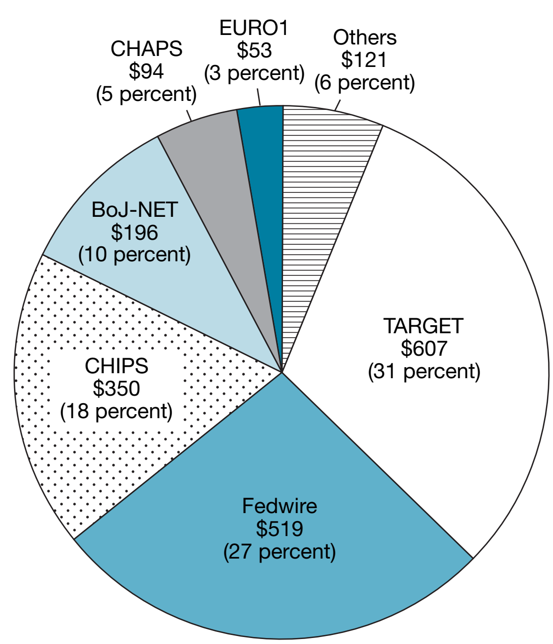
[19] 6Similar developments also took place outside the United States. Annual turnover in the United Kingdom's CHAPS system rose from seven times GDP in its first full year of operation, 1985, to more than fifty times GDP at the turn of the century. In Switzerland, annual turnover reached 120 times GDP in 1997; it has since fallen to around 90 times GDP in 2006, the same level as in 1989.

Chart 6 shows the relative importance of the CPSS countries' interbank payments systems in terms of value settled. Currently, the three largest systems represent more than 75 percent of value transferred. The six largest systems represent almost 95 percent. The largest individual LVPS is TARGET, with an annual settlement value of EUR 489 trillion (USD 607 trillion). The two U.S. systems, Fedwire and CHIPS, respectively rank second and third. However, their combined share of turnover, 45 percent, exceeds the share of combined turnover of the euro payments systems TARGET and EURO1, 34 percent.

The number of transfers settled increased in all of the CPSS countries' payments systems (Chart 7). The Swiss SIC system stands out, with more than 250 million payments settled in 2005. SIC is notable because there is no separate system for settling retail payments in the Swiss payment infrastructure. Thus, retail as well as wholesale payments are settled in SIC. (The settlement of low-value payments in LVPSs

Chart 6

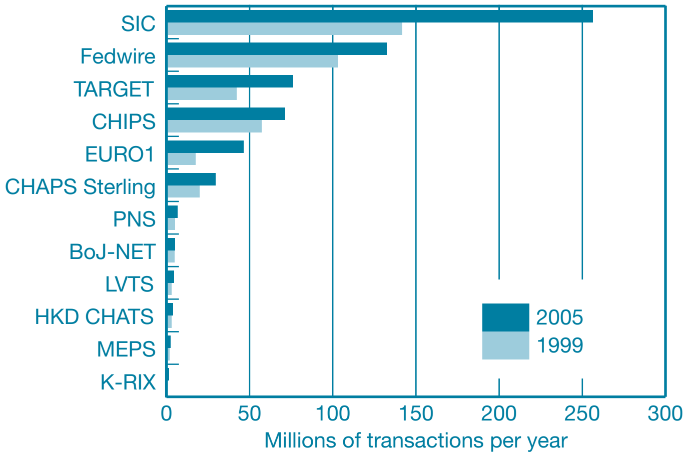
Percentage of Transfer Value in CPSS Countries' Large-Value Payments Systems, 2005



Sources: Bank for International Settlements; authors' calculations.

Notes: Dollar amounts are in trillions of U.S. dollars. "Others" includes SIC, LVTS, PNS, HKD CHATS, K-RIX, MEPS, and USD CHATS (in size order). Systems are described in Appendix A. CPSS is the Committee on Payment and Settlement Systems.

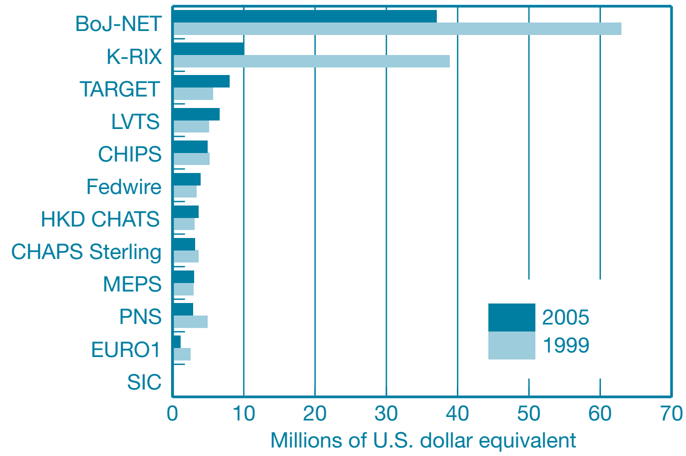
Chart 7
Transactions Settled by CPSS Countries' Large-Value Payments Systems, 1999 and 2005



Sources: Bank for International Settlements; authors' calculations.

Notes: Systems are described in Appendix A. CPSS is the Committee on Payment and Settlement Systems.

Chart 8
Value of Average Payment in CPSS Countries' Large-Value Payments Systems, 1999 and 2005



Sources: Bank for International Settlements; authors' calculations.

Notes: Systems are described in Appendix A. CPSS is the Committee on Payment and Settlement Systems.

is discussed in more detail in the next section.) Fedwire settled the second highest number of transactions, 132 million. In 2005, TARGET settled 72 million transactions and CHIPS settled 76 million.

7. Trend 6: Shrinking Average Payment Sizes

Payments processed through large-value payments systems come in many sizes. The maximum value of a payment allowed over Fedwire is one cent less than USD 10 billion; on occasion, payments of less than USD 1 are processed. In fact, most LVPSs process a significant amount of relatively low-value payments. As a result, the dichotomy between small-value (retail) and large-value (wholesale) payments systems is often blurred. In Fedwire and CHIPS, the median payment size is less than USD 35,000, and almost two-thirds of transfers are for amounts less than USD 100,000. Hence, both systems are important for making low-value payments. The appeal of making low-value transfers in Fedwire derives from the speed, certainty, and finality of settlement and, in some cases, from the ease of reconciliation.[20] In addition, the value distribution has a fat right-hand tail. In other words, a small number of payments account for a large share of value. In Fedwire, 5 percent of the largest payments account for 95 percent of the total value.

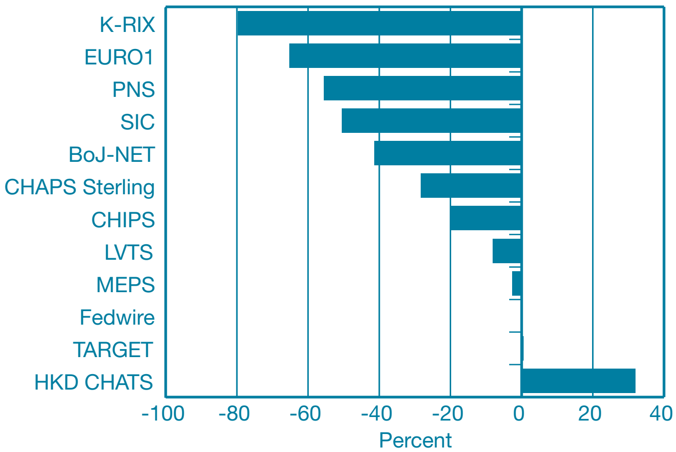
The considerable use of low-value wire transfers is not just a U.S. phenomenon. The distribution of payments handled in the Canadian LVTS is similar to that of the U.S system. The mean value is CDN 8 million (USD 6.6 million), while the median value is about CDN 50,000 (USD 41,300). In the United Kingdom, the mean value of wire transfers processed by CHAPS Sterling is GBP 1.9 million (USD 3.45 million), and the median value is estimated to be approximately GBP 25,000 (USD 45,500). The bulk of payments in Switzerland's SIC system is less than CHF 5,000 (USD 4,000).

Looking across large-value payments systems in CPSS countries, we observe a remarkable dispersion in average payment value (Chart 8). In BoJ-NET, the average payment was in excess of 3.8 billion JPY (USD 35 million) in 2005—the highest among the LVPSs surveyed and three times larger than the second-ranking K-RIX system of the Swedish Riksbank. Nonetheless, this is a significant drop from 1999, when the

[20] See Federal Reserve Board, "A Summary of the Roundtable Discussion on the Role of Wire Transfers in Making Low-Value Payments," May 16, 2006 (available at <http://www.federalreserve.gov/paymentsystems/lowvaluepay/default.htm>).

Chart 9

Inflation-Adjusted Change in Average Payment Size for CPSS Countries' Large-Value Payments Systems, 1999-2005



Sources: Bank for International Settlements; authors' calculations.

Notes: Systems are described in Appendix A. CPSS is the Committee on Payment and Settlement Systems.

average payment value was more than JPY 6,941 million (USD 63 million). The drop is attributable to the change from a deferred net settlement to a real-time gross settlement system in 2001. An upper limit of JPY 5 billion per payment was set to allow for smoother settlement in the new RTGS system. TARGET is ranked third and followed by the North American systems. At the other end of the spectrum is the SIC system, with SFR 160,000 (USD 130,000), along with the euro systems PNS (USD 2.9 million) and EURO1 (USD 1.2 million).

Adjusting for inflation, Chart 9 shows that the average payment size has fallen for most systems from 1999 to 2005. For five of the twelve systems, it fell more than 40 percent. For Fedwire and TARGET, the average size has remained unchanged. The only increase in payment size adjusted for inflation occurred in the HKD CHATS system; it is to a large extent attributable to the period of deflation that Hong Kong experienced between 1999 and 2004.[21]

[21] The consumer price index in Hong Kong fell 12 percent between 1999 and 2004.

8. Trend 7: Falling Numbers of System Participants

One of the most pervasive trends in international banking is consolidation. All else equal, larger banks imply fewer participants in LVPSs. In addition, an increasing focus on costs has made banks more selective in terms of the systems in which they participate. It is no longer considered "a must" for many foreign banks to clear their USD payments themselves as it was in the 1980s and 1990s. However, technological advances have made it possible to participate in more systems in new ways, as we describe in the emergence of cross-border and offshore systems (Trend 3).

According to figures reported to the Bank for International Settlements as part of its "Red Book" statistics, Fedwire had 10,000 participants in the late 1990s and 6,819 participants in 2005. Despite this decrease, Fedwire is still by far the largest LVPS in the CPSS group. The second largest system is TARGET, with 2,628 participants. With only twelve direct participants, the Canadian LVTS is the smallest system in the CPSS group (Table 4).

Most LVPSs in the CPSS group saw participation decline from 1999 to 2005 (Table 4). However, there was a substantial

Table 4

Ranking of CPSS Countries' Large-Value Payments Systems by Number of Participants, 1999 and 2005

System	Ranking	Total Participants, 1999	Ranking	Total Participants, 2005	Percentage Change
Fedwire	1	9,994	1	6,819	-32
TARGET	2	5,144	2	2,628	-49
BoJ-NET	3	409	3	357	-13
SIC	5	291	4	325	12
CHAPS	4	404	5	241	-40
HKD CHATS	6	151	6	129	-15
MEPS	7	136	7	111	-18
EURO1	9	72	8	75	4
CHIPS	8	77	9	48	-38
PNS	10	25	10	45	80
K-RIX	11	23	11	21	-9
LVTS	12	14	12	15	7

Sources: Bank for International Settlements; authors' calculations.

Notes: Systems are described in Appendix A. CPSS is the Committee on Payment and Settlement Systems.

increase in the number of participants in SIC and, even more so, in PNS.[22] SIC's growth in participation is attributable to its opening to "remote membership" in 1998 (see our Trend 3 discussion). The number of remote participants in SIC has since increased, whereas the number of Swiss participants has remained largely unchanged.

9. Trend 8: Extended Operating Hours

A direct effect of the globalization of financial markets is the extension of operating hours in large-value payments systems. Between 1997 and 2005, there were two waves of operating hour extensions in CPSS countries' large-value payments systems. The first occurred in 1998 and 1999, when several European LVPSs changed or extended their operating hours to coincide with TARGET's business hours. The second wave coincided with the launch of CLS Bank in 2002. BoJ-NET, MEPS, and LVTS extended their operating hours to synchronize with CLS Bank settlement hours during the European morning. The CLS Bank settlement process takes place during a five-hour window from 7:00 a.m. to noon (CET). Its operating hours partially overlap the operating hours of all the participating RTGS systems (Chart 10).

A direct effect of the globalization of financial markets is the extension of operating hours in large-value payments systems.

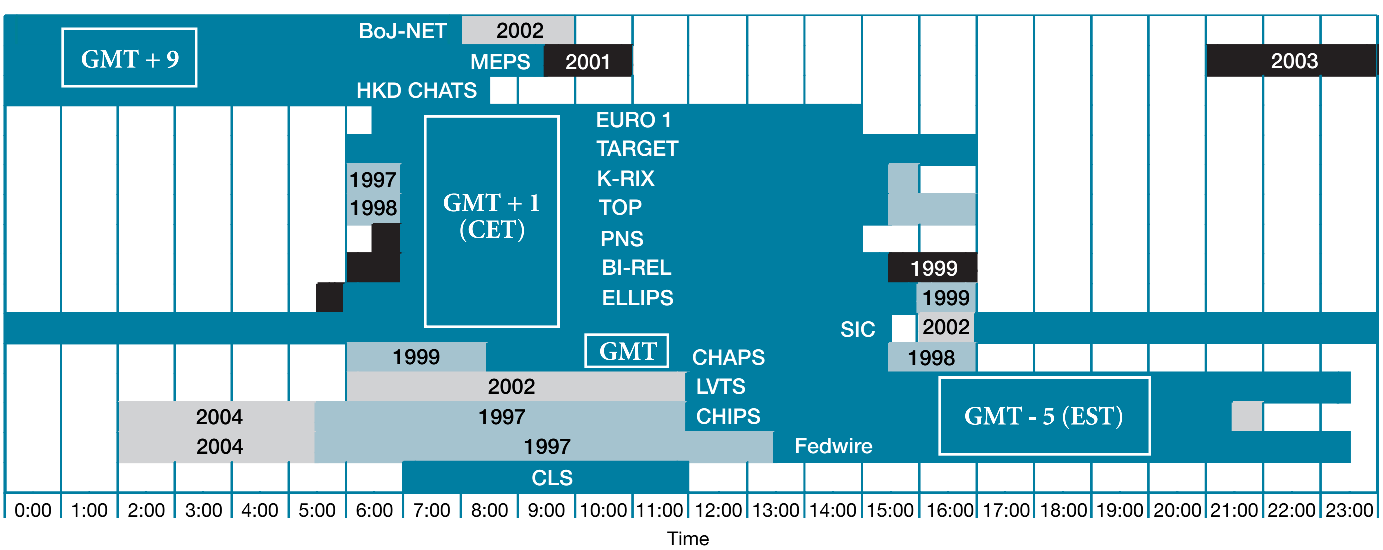
To meet industry requests to achieve greater overlap of U.S. wholesale payments system operating hours with those of the Asia-Pacific markets, Fedwire and CHIPS expanded their operating hours in 1997 and 2004. In both instances, the opening was moved to earlier while the closing remained unchanged.

In 2007, SIC had the longest operating hours—23 hours and 15 minutes—approaching a twenty-four-hour settlement cycle. The three North American systems—Fedwire, CHIPS,

[22] For PNS, the increase in participants is attributable to an increase in the number of indirect participants; by and large, it can be explained by a reclassification of their status that took place over the period.

Chart 10

Opening Hours of CLS Bank and Selected CPSS Countries' Large-Value Payments Systems

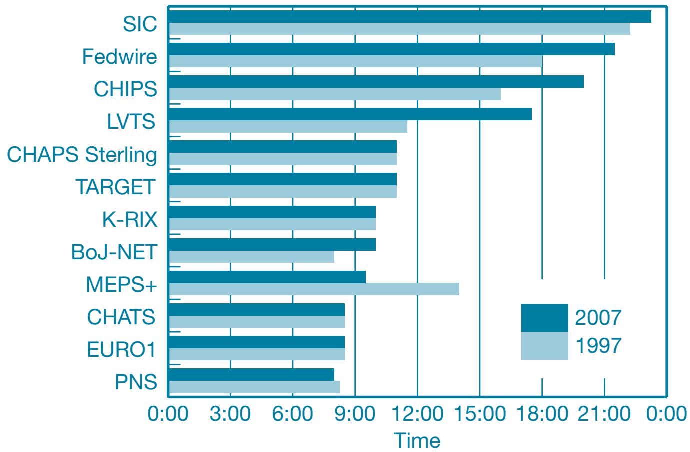


Sources: Bank for International Settlements; authors' calculations.

Notes: The blue and black shading represents the operating hours in 1996 or when the system was implemented. The full extent of all shaded areas (excluding the black areas, which represent opening hour shortenings in 1999 and before) is operating hours in 2005. Green shading indicates extensions in 1999 and before; grey shading shows extensions after 1999. Systems are described in Appendix A.

Chart 11

Length of Operating Hours of Selected CPSS Countries' Large-Value Payments Systems, 1997 and 2007



Sources: Bank for International Settlements; authors' calculations.

Notes: Systems are described in Appendix A. CPSS is the Committee on Payment and Settlement Systems.

and LVTS—followed with 21.5, 20, and 17.5 hours, respectively. At the other end of the spectrum were the European systems PNS and EURO1 and the Asian systems MEPS and HKD CHATS, with operating hours of between 8 and 9.5 hours. A reduction in operating hours occurred only in the Singaporean system MEPS and, to a lesser extent, in the French PNS system (Chart 11).

10. Trend 9: Declining Transaction Fees

Large-value payments systems are characterized by large economies of scale, as there are considerable fixed costs in terms of setting up and maintaining the systems. In contrast, individual payments generate diminutive costs. This creates a potential problem for efficient pricing. Standard economic theory suggests that each transaction should be priced at its marginal cost. However, marginal cost pricing implies that the fixed costs are not recovered (see Holthausen and Rochet [2006]). Central banks and system operators around the world have found different solutions to this challenge, depending on their mandate for the provision of payment services.

Most systems charge a fixed membership or admission fee and the majority charge a per-transaction fee. However, differences exist in terms of whether both the originator and the receiver are charged for a transaction. TARGET only charges the originator whereas both Fedwire and SIC levy fees on both the originator and the receiver. A simple, flat transaction fee schedule is often used, but several systems base the fee on a combination of the volume submitted by the participant, the value of the particular payment, the submission time of the payment, and the mode of delivery, such as online and offline. In addition, participants may have to pay separate communications charges, for example, to SWIFT.

In 1999, the Federal Reserve implemented a volume-based fee schedule to reflect more accurately the cost structure of Fedwire services and its demand elasticity. This type of structure remains in place today, in which offline participants are also assessed a surcharge to initiate or receive a funds transfer. In Switzerland, the originator of the payment is charged differently depending both on the time when the payment is submitted and its value. The SIC pricing schedule is illustrated in Table 5. The receiver is charged a flat fee regardless of the settlement time and value. The fee structure provides an incentive for early input and settlement of payments, which in turn prevents the demand for settlement from peaking at the end of the day (see Heller, Nellen, and Sturm [2000]). For cross-border payments, TARGET has a transparent, volume-based pricing structure. Domestic payments are currently priced by each TARGET component independently.

Large-value payments systems are characterized by large economies of scale, as there are considerable fixed costs in terms of setting up and maintaining the systems. In contrast, individual payments generate diminutive costs. This creates a potential problem for efficient pricing.

In the second half of the 1990s, the Federal Reserve undertook a five-year project—akin to the current TARGET2 initiative—to consolidate its processing facilities. The project resulted in significant savings that were passed on to users in the form of lower fees. The average transaction fee (nominal)

Table 5
Pricing Principles in Fedwire, TARGET, and SIC

	Volume-Based	Value-Based	Time of Day	Mode of Delivery	Transaction Fee (U.S. Dollar Equivalent) Lowest Sender	Lowest Total	Highest Sender	Highest Total
Fedwire	Y	N	N	Y	0.1	0.2	0.3	0.6
TARGET	Y	N	N	N	1.1	1.1	2.4	2.4
SIC	N	Y	Y	N	0.01	0.03	3.7	3.7

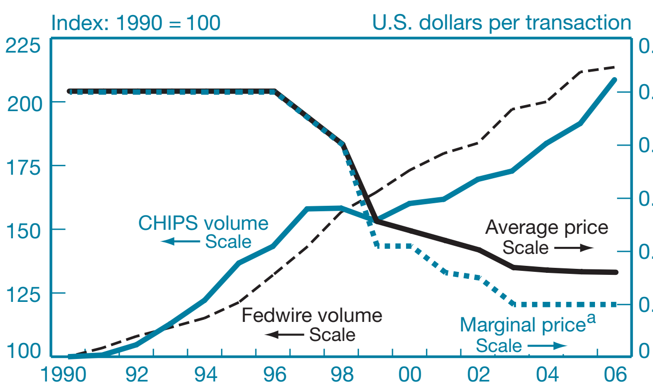
Sources: European Central Bank; Federal Reserve Bank of New York; Swiss National Bank.

Note: Total = sender + receiver, exchange rate of May 25, 2007.

in Fedwire dropped from USD 0.50 in 1996 to less than USD 0.16 in 2006—a reduction of 68 percent (Chart 12). SIC participants have seen a similar drop, as the average transaction fee fell from 17 centimes in 1999 to 7 centimes in 2005—a reduction of 60 percent. TARGET transaction fees have been fixed at the same nominal rates since 1999.

Pricing in LVPSs is important not only from a revenue and cost perspective but also from a competitive perspective when private and public systems co-exist as they do in Europe and the United States. As an example, the growth in the respective volumes of payments submitted to Fedwire and CHIPS since 1990 is illustrated in Chart 12. Both systems have more than doubled the amount of payments they process. However, the trajectory of volumes processed over CHIPS declined significantly in the late 1990s and did not recover until 2001. At the same time, Fedwire was experiencing steady growth. This change coincides with the reduction in Fedwire fees and the move to volume-based pricing. Conversely, in late 2005, CHIPS announced new incentive pricing for existing and new participants. Based on one year of data, incentive pricing appears to have had some effect, as Fedwire volume grew by only 1 percent—the lowest rate in twenty years—while CHIPS volume grew by 9 percent.

Chart 12
The Effect of Fedwire Price Reductions on CHIPS Volume



Sources: CHIPS; Federal Reserve Bank of New York; authors' calculations.

[a] 80,000 or more transactions.

11. Trend 10: Adoption of Common Standards for Large-Value Payments Systems

In 1980, the governors of the central banks of the G-10 countries established the Group of Experts on Payment Systems. One of the Group's first projects was to conduct a detailed review of payments system developments in the G-10 countries. It was published by the BIS in 1985 in the first of a series that has become known as the "Red Book." The Group also analyzed interbank netting schemes in the Angell Report ("Report on Netting Schemes") and in the Lamfalussy Report ("Report of the Committee on Interbank Netting Schemes of the Central Banks of the Group of Ten Countries"), published by the BIS in 1989 and 1990, respectively. In 1990, the G-10 governors established the Committee on Payment and Settlement Systems to assume and extend the activities of the Group of Experts on Payment Systems. The CPSS was set up as one of the BIS' permanent central bank committees reporting to the G-10 governors.

The CPSS has focused on disseminating information on payments system design and has been defining payment norms and best practices for the central bank community. In 1997, the CPSS published a report on real-time gross settlement systems and in 2005 a report on new developments in large-value payments systems that focused on changes since the RTGS report. Furthermore, in January 2001, the CPSS published the "Core Principles for Systemically Important Payment Systems." The document was developed to serve as guidelines for promoting safety and efficiency in the design and operation of systemically important payments systems (SIPSs).[23]

> *The standards published by the CPSS [Committee on Payment and Settlement Systems] provide the main principles for the design and operation of payments and settlement systems.*

The standards published by the CPSS provide the main principles for the design and operation of payments and settlement systems. They are currently being used as a reference by central banks and international organizations in their efforts to improve the safety and efficiency of payments systems worldwide. They are part of a set of key standards that the international community considers essential to strengthening and preserving financial stability. As such, these standards are used by the joint IMF and World Bank Financial Sector Assessment Program (FSAP) and the Reports on the Observance of Standards and Codes.

A growing number of LVPSs around the world have been assessed according to the Core Principles. This has occurred either as a self-assessment by the system operator, the central bank, or as part of an FSAP review. In some cases, systems have even been self-assessed and assessed by the IMF and the World Bank. Some central banks, such as the Bank of England and the Swedish Riksbank, use the Core Principles for annual assessments of their SIPSs as part of their payments system oversight.

The first country to assess its payments system according to the Core Principles was Canada. In late 1999, Canada participated in an FSAP pilot that included an assessment of the Canadian payments system, LVTS, using a draft version of the Core Principles. The payments systems in Cameroon and Estonia were assessed in 2000. In January 2001, the Governing Council of the European Central Bank adopted the Core Principles as the minimum standards for the Eurosystem's common oversight policy on systemically important payments systems. As a consequence, the Governing Council decided that all nineteen SIPSs in the euro area would be assessed against the Core Principles in mid-2003.

Exhibit 2 on the next page displays the countries that have applied the Core Principles to assess their LVPSs. At the end of 2006, fifty-nine countries had done so—up from twenty-two at the end of 2002. Of the fifty-nine countries, twelve had been both self-assessed and under the scrutiny of the IMF and World Bank. In addition, payments systems in another forty-two countries had been assessed as part of an FSAP review, while five had conducted self-assessments. The United States belongs to the latter group.

According to the assessments, practically all SIPSs in developed countries meet the standards and codes in the Core Principles. This is also true for payments systems in many emerging economies. Private sector payments systems in major currencies also comply with the Core Principles; notable examples are CHIPS and EURO1.

12. Future Developments

This article describes ten long-range trends in the settlement of large-value payments. The questions worth considering are how these trends will evolve and what new developments can be foreseen. We offer some thoughts on these questions.

Currently, the diffusion of RTGS is well under way. RTGS and net settlement systems each have characteristics that make them desirable, thus the hybridization of RTGS is likely to continue as long as liquidity is costly. Many central banks require collateral for intraday credit. With the ongoing development of financial markets, collateral is likely to find new, more profitable uses than payment settlement. This will likely drive the cost of liquidity up and, as a consequence, increase the demand for liquidity saving that netting and offsetting in conjunction with RTGS can offer. The trend toward greater hybridization of systems is therefore likely to continue.

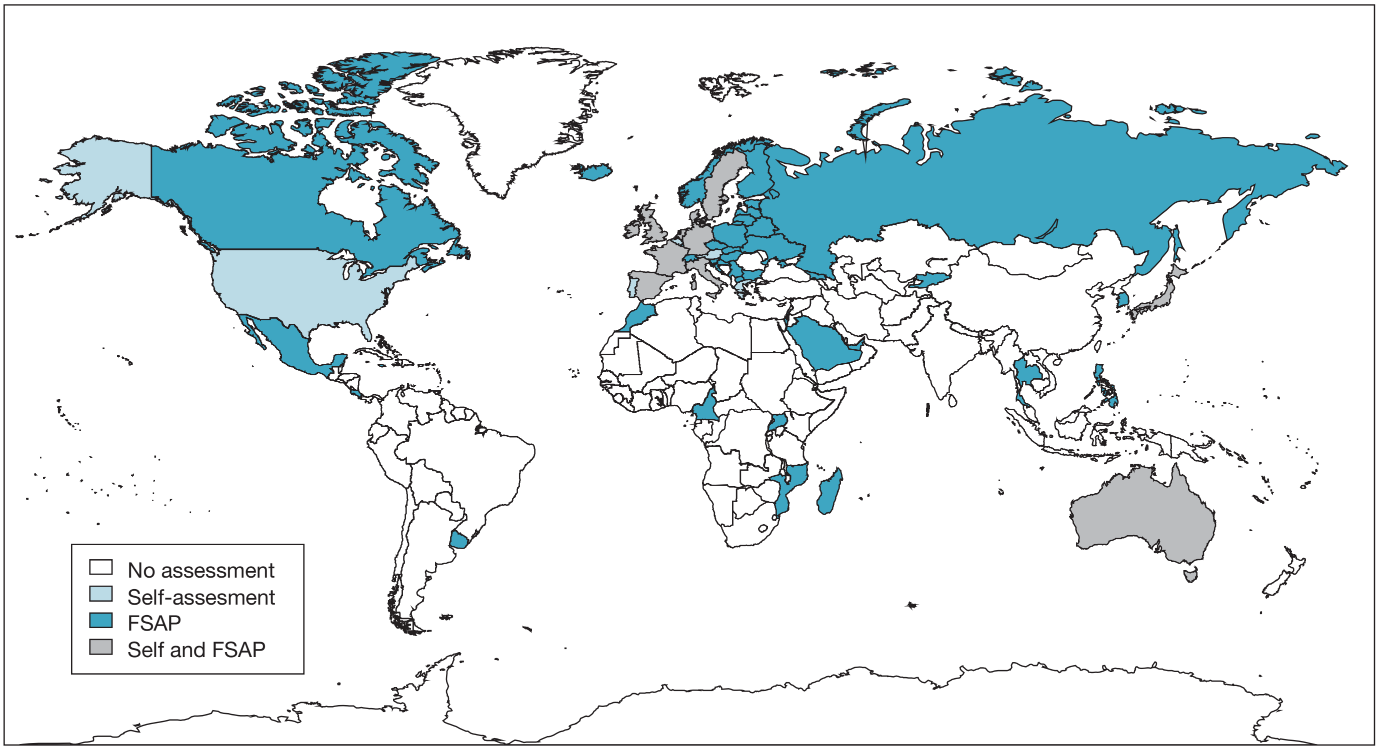
The introduction of cross-border systems has been associated with unique events linked to the introduction of the euro and the establishment of CLS Bank. Cross-border systems are likely to remain rare in the future. However, remote participation may become more prevalent.

Offshore systems that settle a foreign currency are presently small and serve niche markets—mainly a local FX market or

[23] A payments system is systemically important when, if the system were insufficiently protected against risk, disruption within it could trigger or transmit further disruptions among participants or systemic disruptions in the financial area more widely (Bank for International Settlements 2003).

Exhibit 2

Assessment of Compliance with the "Core Principles for Systemically Important Payments Systems," 2006



Sources: International Monetary Fund (IMF); websites of countries' central banks.

Note: FSAP is the joint IMF-World Bank Financial Sector Assessment Program.

the needs of banks in the area and time zone to settle payments in a foreign currency among each other. Such demands may arise in the context of the establishment of new financial centers, for instance, in the Middle East or China, where the People's Bank of China is developing a USD clearing system. Most existing or planned offshore systems are limited to a single country. With improvements in information and communications technologies, the fixed cost of setting up such systems is being reduced. As a consequence, we may see more offshore systems emerge, but they are likely to remain niche players, much like the existing ones are.

CLS Bank settlements have grown rapidly and are likely to continue to do so. Currently, CLS Bank captures around 50 percent of all FX trades and is pursuing the settlement of other types of transactions. As most of its costs stem from fixed investments, CLS Bank has incentives to continue fostering growth in settlement volumes.

Settlement values are likely to continue growing at the pace of GDP in the long run, and be cyclical to financial market activity in the short run—as they have done over the past ten years. The rapid growth in values attributable to financial deregulation and innovation in the 1980s and early 1990s has largely been absorbed.

The average real value of payments processed in LVPSs has declined. As transaction prices seem to be declining too, it can be expected that the benefits of real-time settlement will outweigh the costs for a wider variety of smaller financial transactions. Thus, we expect that the average value of large-value payments will continue to fall.

Consolidation in financial services is continuing. Especially in Europe, the process of cross-border mergers has not yet taken off. In addition, the introduction of TARGET2 and the consolidation of all the EU RTGS systems into a single entity will substantially reduce the number of LVPS participants, as banks operating in several EU countries will be better positioned to manage their payments centrally.

One reason for the emergence of offshore systems is that the operating hours of domestic LVPSs do not coincide with the business hours of LVPSs in other countries. Often, the operating hours of euro or U.S. dollar LVPSs do not sufficiently overlap those of Asian financial centers. Fedwire and CHIPS have reacted to this disparity by extending their operating hours. The European systems have had less of a need to do so because their operating hours already overlap those of Asian financial centers by a few hours. The overlap with current operating hours, however, is now wider for the U.S. dollar than for the euro—in spite of the more advantageous time differences in Europe.

Evidence from systems for which price data are available suggests that the cost of payments in LVPSs has declined rapidly. The underlying reasons are associated with regulatory changes, lower costs of information and communications technology, and perhaps competition between the public and private systems that operate side by side in some countries. These reasons are not likely to change, and the cost of making payments is likely to continue to fall.

The final trend we discussed was the standardization of large-value payments systems through the use of common standards. The "Core Principles for Systemically Important Payment Systems" is already widely accepted and will continue to be applied around the world.

Appendix A: Large-Value Payments Systems in CPSS Member Countries

Country	System Name	Abbreviation	Year of Implementation	Annual Number of Transactions (2005, in Thousands)	Annual Value of Transactions (2005, in Billions of U.S. Dollars)
Belgium	Electronic Large-Value Interbank Payments System	ELLIPS[a]	1996	1,800	21,448
Canada	Large-Value Transfer System	LVTS	1999	4,600	30,321
France	Transferts Banque de France	TBF[a]	1997	4,300	151,425
France	Paris Net Settlement	PNS	1999	6,800	19,432
Germany	RTGSplus	RTGSplus[a]	2001	35,800	172,023
Hong Kong	HK Dollar Clearing House Automated Transfer System	HKD CHATS	1996	4,100	14,936
Hong Kong	U.S. Dollar Clearing House Automated Transfer System	USD CHATS	2000	1,500	1,588
Hong Kong	Euro Clearing House Automated Transfer System	Euro CHATS	2003	1,000	422
Italy	BI-REL	BI-REL1	1997	10,400	40,840
Japan	BoJ-NET Funds Transfer System	BoJ-NET	1988	5,300	196,452
Netherlands	TOP	TOP[a]	1997	4,700	38,126
Singapore	Monetary Authority of Singapore (MAS) Electronic Payments System	MEPS	1998	2,500	7,564
Sweden	K-RIX	K-RIX	1990	1,500	14,867
Sweden	E-RIX	E-RIX[a]	1999	1,000	279
Switzerland	Swiss Interbank Clearing	SIC	1987	256,400	32,956
Switzerland	EuroSIC	EuroSIC	1999	9,130	630
United Kingdom	CHAPS Sterling	CHAPS Sterling	1984	29,600	94,299
United Kingdom	CHAPS Euro	CHAPS EURO1	1999	5,100	66,859
United States	Fedwire Funds Service	Fedwire	1918	123,400	518,547
United States	Clearing House Interbank Payments System	CHIPS	1970	71,500	349,871
European Union	Trans-European Automated Real-time Gross settlement Express Transfer system	TARGET	1999	76,200	607,254
European Union	EURO1	EURO1	1999	46,400	53,334
European Union	ECB Payment Mechanism	EPM[a]	1999	41	5
International	Continuous Linked Settlement Bank	CLS Bank[b]	2002	47,900	785,300

Notes: In this article, we analyze the European TARGET system but do not consider its components (ELLIPS, TBF, RTGSplus, BI-REL, and TOP). CPSS is Committee on Payment and Settlement Systems.

[a] The system is a component of TARGET, which consists of fifteen national real-time gross settlement systems and the EPM system of the European Central Bank.

[b] CLS Bank data are based on the aggregation of both sides of a foreign exchange transaction.

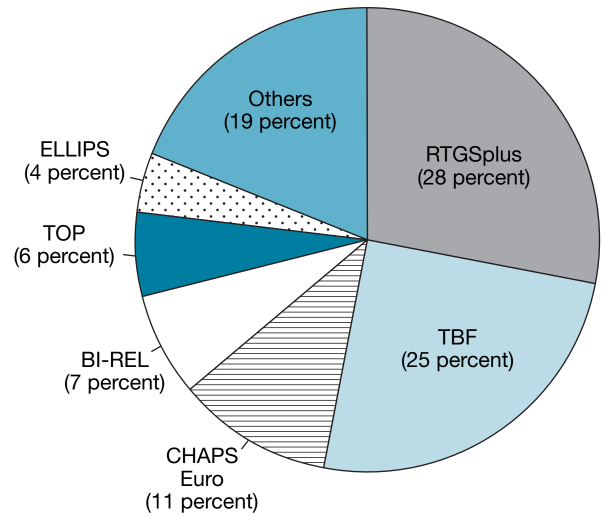
Appendix B: TARGET and TARGET2

TARGET (Trans-European Automated Real-time Gross settlement Express Transfer system) is the RTGS system for the euro, owned and operated by the Eurosystem. It went live in January 1999. The system is used for the settlement of central bank operations, large-value euro interbank transfers, and other euro payments. TARGET was created by interconnecting national euro RTGS systems and the European Central Bank payment mechanism.

Within TARGET, the two largest components—the German RTGSplus and the French TBF—process more than half of the values transacted. The next three largest systems—the United Kingdom's CHAPS Euro, the Italian BI-REL, and the Dutch TOP—process another quarter of transfer values, while the remaining ten smaller components are responsible for the remaining quarter or less.

The decentralized infrastructure of TARGET is being replaced by TARGET2, which is based on a single technical platform. TARGET2 went live in November 2007. All Eurosystem central banks and Kronos Euro will participate in the new system. E-RIX discontinued operations on January 1, 2007, and CHAPS Euro will not connect to TARGET2.

Share of Values Transferred by TARGET Components



Source: European Central Bank, TARGET 2006 Annual Report.

Appendix C: Cross-Border Payments in TARGET

In 2006, TARGET processed on average 326,000 payments a day worth about €2.1 trillion. Inter-member state or cross-border payments accounted for 23 percent of the volume and 35 percent of the value of payments transferred (see table).

Not surprisingly, the cross-border flow of payments within TARGET correlates with economic output and the size of the financial sectors across countries. The geographical network of payment flows in Europe is shown in the exhibit. Data are from 1999-2002 and hence exclude some of the newer accession countries. The size of each node is proportional to the value of intra-border or domestic payment flows, and the width of links between countries is proportional to the value of cross-border flows. The largest flows are between the three largest economies: Germany, the United Kingdom, and France. Although the United Kingdom has not yet adopted the euro, it has a prominent role in cross-border euro payment flows.

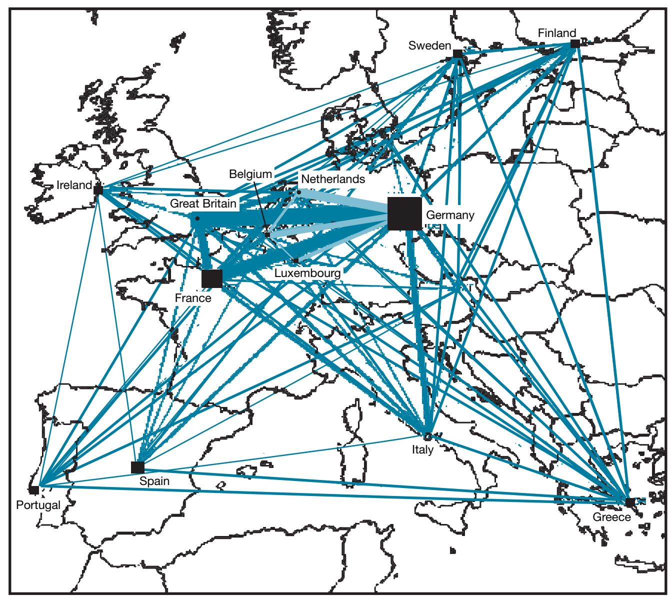
Inter- and Intra-Member State Payments in TARGET, 2006

	Value		Volume	
	Daily Average (Billions of Euros)	Share (Percent)	Daily Average (Thousands of Payments)	Share (Percent)
Total	2,092		326	
Inter-member state	725	35	75	23
Intra-member state	1,368	65	252	77

Source: <www.ecb.int>.

This simply reflects the importance of London as a financial center. Interestingly, euro payment flows within the United Kingdom are minuscule in comparison. In contrast, the value of cross-border payments to and from Italy, Spain, and Portugal is less than the countries' respective shares of EU gross domestic product. The opposite is true for the Benelux countries, all of which host important financial centers (light blue lines).

Geography of TARGET Cross-Border Payments



Sources: <http://www.ecb.int>; Rosati and Secola (2006).

References

Arjani, N., and D. McVanel. 2006. "A Primer on Canada's Large-Value Transfer System." Bank of Canada.

Bank for International Settlements. 1996. "Settlement Risk in Foreign Exchange Transactions." Committee on Payment and Settlement Systems, Publication no 17. Basel, Switzerland.

———. 1998. "Reducing Foreign Exchange Settlement Risk: A Progress Report." Committee on Payment and Settlement Systems, Publication no 26. Basel, Switzerland.

———. 2001. "Core Principles for Systemically Important Payment Systems." Committee on Payment and Settlement Systems, Publication no 43. Basel, Switzerland.

———. 2003. "A Glossary of Terms Used in Payment and Settlement Systems." Committee on Payment and Settlement Systems. March. Basel, Switzerland.

———. 2005. "New Developments in Large-Value Payment Systems." Committee on Payment and Settlement Systems, Publication no 67. May. Basel, Switzerland.

———. 2007a. "Statistics on Payment and Settlement Systems in Selected Countries—Figures for 2005." Committee on Payment and Settlement Systems, Publication no 78. Basel, Switzerland.

———. 2007b. "Progress in Reducing Foreign Exchange Settlement Risk." Committee on Payment and Settlement Systems, Publication no 81. Basel, Switzerland.

Bech, M. 2007. "The Diffusion of Real-Time Gross Settlement." In S. Millard, A. Haldane, and V. Saporta, eds., The Future of Payments. Routledge International Studies in Money and Banking. London: Routledge.

Bech M., and B. Hobijn. 2007. "Technology Diffusion within Central Banking: The Case of Real-Time Gross Settlement." International Journal of Central Banking 3, no. 3 (September): 147-81.

Bech, M., and K. Soramäki. 2001. "Liquidity, Gridlocks, and Bank Failures in Large-Value Payments Systems." Central Banking Publications, E-Money and Payments System Review: 111-26.

———. 2005. "Systemic Risk in a Netting System Revisited." In H. Leinonen, ed., Liquidity, Risks, and Speed in Payment and Settlement Systems—A Simulation Approach, 275-96. Bank of Finland Studies in Economics and Finance.

Heller, D., T. Nellen, and A. Sturm. 2000. "The Swiss Interbank Clearing System." Swiss National Bank, Payment System Subsection, June. Available at <http://www.sic.ch/dl_tkicch_sicartikelsnb.pdf>.

Holthausen, C., and J.-C. Rochet. 2006. "Efficient Pricing of Large-Value Interbank Payment Systems." Journal of Money, Credit, and Banking 38, no. 7 (October): 1797-818.

Humphrey, D. 1986. "Payments Finality and Risk of Settlement Failure." In A. Saunders and L. J. White, eds., Technology and the Regulation of Financial Markets: Securities, Futures, and Banking, 97-120. Lexington, Mass.: Heath.

Leinonen, H., and K. Soramäki. 1999. "Optimizing Liquidity Usage and Settlement Speed in Payment Systems." Bank of Finland Discussion Paper no. 16, December.

McAndrews, J., and J. Trundle. 2001. "New Payment System Designs: Causes and Consequences." Bank of England Financial Stability Review 11, December: 127-36.

Rosati, S., and S. Secola. 2006. "Explaining Cross-Border Large-Value Payment Flows— Evidence from TARGET and EURO 1 Data." Journal of Banking and Finance 30, no. 6 (June): 1753-82.

Selgin, G. 2004. "Wholesale Payments: Questioning the Market-Failure Hypothesis." International Review of Law and Economics 24, no. 3 (September): 333-50.

Olivier Armantier, Jeffrey Arnold, and James McAndrews

Changes in the Timing Distribution of Fedwire Funds Transfers

- The Federal Reserve's Fedwire funds transfer service has long displayed a concentrated peak of activity in the late afternoon.
- Sending large payments late in the day can heighten operational risk by increasing the potential magnitude of liquidity dislocation and risk if operational disruptions occur.
- A study of the distribution of Fedwire payments finds that the peak of the timing distribution has become more concentrated, has shifted to later in the day, and has been divided into two peaks.
- These trends are likely explained by a higher value of payments transferred, the settlement patterns of private settlement institutions, and increased industry concentration.
- The study uncovers no specific evidence of heightened operational risk associated with late activity, but it points to a high level of interaction between Fedwire and private settlement institutions.

1. Introduction

The Federal Reserve's Fedwire funds transfer service is the biggest large-value payments system in the United States in terms of participants, value, volume, and use by other settlement systems. Although Fedwire funds activity has long been concentrated in the late afternoon, recently there has been a noticeable shift to later in the day. The value of funds activity after 17:00 has increased from 20 percent in 1998 to more than 30 percent in 2005 (Board of Governors of the Federal Reserve System 2006). In 2006, the Federal Reserve commented on the risk posed by this change:

> "From an operational risk perspective, delaying the sending of large payments until late in the day increases the potential magnitude of liquidity dislocation and risk in the financial industry if late-in-the-day operational disruptions should occur. An increase in such risk is particularly troublesome in an era of heightened concern about operational disruptions from a range of sources" (Board of Governors of the Federal Reserve System 2006).

There is a complex set of trade-offs between risks and costs in large-value payments systems (Bank for International Settlements 2005). Theory suggests that the concentration of late-afternoon Fedwire activity is the result of coordination among banks to reduce liquidity costs, delay costs, and credit risks. As these costs and risks change over time, we would

Olivier Armantier is a senior economist at the Federal Reserve Bank of New York; Jeffrey Arnold was a research associate at the Bank at the time this article was written; James McAndrews is a senior vice president at the Bank.
<olivier.armantier@ny.frb.org>
<james.mcandrews@ny.frb.org>

The authors thank the managements of CHIPS and DTC for permission to report data associated with their operations. They also thank Craig Furfine and other seminar participants at the Federal Reserve Bank of Chicago and the Bank of England for helpful discussions as well as an anonymous referee for valuable comments. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

expect the timing of payment activity to be affected. In this article, we seek to quantify how the changing environment in which Fedwire operates has affected the timing of payment value transferred within the system.

We observe several trends in payment timing from 1998 to 2006. After 2000, the peak in payment activity shifts to later in the day. Indeed, post-2000, a greater concentration of payments occurs after 17:00. At the same time, however, several factors have been associated with increased payment activity early in the day, such as the creation of the Continuous Linked Settlement (CLS) Bank, an institution that settles U.S. dollar payments early in the morning; changes to the Clearing House Interbank Payments System's (CHIPS) settlement practices; and expanded Fedwire operating hours. Despite these developments, we find that the distribution of payment activity across the day still peaks more in the late afternoon.

Payments made through Fedwire are distributed throughout the operating day, so no single statistic fully captures the changes in timing that we observe. To analyze the timing

We seek to quantify how the changing environment in which Fedwire operates has affected the timing of payment value transferred within the system.

of payments on Fedwire, we measure the times at which each percentile of value transferred on a particular day was completed. In addition, we use regression analysis to examine factors associated with the intraday timing of each percentile of value transferred by Fedwire over time. Explanatory variables here include changes in the Federal Reserve's Payments System Risk Policy and the activity of settlement systems. We measure the effects of multiple explanatory variables on the whole time distribution of payments to understand more fully why parts of the Fedwire timing distribution have changed.

Our study focuses on two notable changes that have affected the higher percentiles of value transferred on Fedwire: the 60th-90th percentiles. First, the most concentrated period of value transfer on Fedwire has moved to later in the day, from 16:48 to 17:11; second, the concentration of Fedwire value transferred has increased. Together, these changes have resulted in greater percentages of value transferred after 17:00.[1] Our analysis suggests that these changes are explained largely by the change in the timing of CHIPS end-of-day settlement activity, growth in the volume and value of Fedwire payments, and growth in the pattern of industry concentration (a measure of the amount of Fedwire value submitted by different system participants).

Our study proceeds as follows. The next section reviews the literature on models of banks' payment timing decisions.[2] Section 3 considers how Fedwire payments are currently distributed across the time of day and how the distribution has changed over time. In Section 4, we use regression analysis to examine which factors have been most relevant in explaining these trends. Section 5 focuses on the influence that settlement institutions have on the timing of Fedwire payments. We conclude with a brief summary of our observations and offer suggestions for further research.

2. Literature Review

Banks seek to minimize the costs associated with sending payments. The existing models of banks' payment-sending behavior generally focus on four factors: the cost of liquidity, the cost of delay, uncertainty (both strategic and structural) and settlement risk, and the instruction arrival process.

The cost to banks of settling a payment is a function of the cost of liquidity used to send the payment and the cost associated with delaying it. The cost of liquidity is the cost a bank faces when using account balances (usually modeled as being held at the central bank) in a payments system. This cost is often zero as long as the bank has a positive balance in its account and nonzero if the bank needs to borrow money to settle the payment. In Fedwire, the Federal Reserve provides intraday (daylight) credit in the form of overdrafts for which banks are charged a marginal fee, if the bank is eligible for daylight credit and has exceeded the amount of credit that falls within a deductible amount (Board of Governors of the Federal Reserve System 2004). Many other payments systems provide collateralized intraday liquidity at zero marginal cost after the bank posts collateral. An additional source of liquidity in payments systems is the overnight money market for federal funds and repos. The cost of delay is the compensation a bank must pay a customer for a delayed payment or the reputational damage and loss of future business a bank suffers from delaying customer payment requests (Angelini 1998).

Angelini (1998, 2000) considers the behavior of banks vis-à-vis liquidity and delay costs in a real-time gross settlement (RTGS) payments system. He shows that the equilibrium in an RTGS system involves excessive delay of payments, as banks do not fully internalize the benefits to other banks from the receipt

[1] We measure this peak as the median daily time of the top ten minutes of Fedwire value.

[2] For simplicity of exposition, we use the term *banks* to refer to direct participants in the payments system, although these participants may not necessarily be banks.

of funds. Bech and Garratt (2003) find that RTGS systems can be characterized by multiple equilibria, with the relative costs of liquidity and delay determining whether the system has an early or late equilibrium. Additionally, whether intraday liquidity is provided as priced or collateralized credit influences the type of "game" and the associated equilibria that would result. Other papers that consider how changes in liquidity prices affect payment timing are Bech (2008) and Mills and Nesmith (2008). Green (2005) discusses the welfare implications of these models and questions whether there are social costs to delay.

An observed late-day distribution of payments could occur either because banks delay payments or because banks receive payment instructions late in the day.

Kahn, McAndrews, and Roberds (2003), Mills and Nesmith (2008), and Bech (2008) focus on settlement risk in payments systems. Settlement risk is the uncertainty to which banks are subject when they face the choice of submitting or delaying a payment. Participant A may expect an offsetting payment from Participant B to occur later in the day. However, uncertainty about whether Participant B might either default or delay sending payment until the next day can result in Participant A's decision to delay delivery of its payment to the RTGS system.

An important but often overlooked assumption of these models is the time at which customers submit payment instructions to their bank. It is only after a bank receives an instruction from its customer (including the bank itself) that the bank decides whether to send settlement instructions to the payments system or delay settlement. An observed late-day distribution of payments could occur either because banks delay payments or because banks receive payment instructions late in the day. For example, after CLS Bank began operations, banks had a new stream of payments to submit at a particular time of day, which can cause significant changes in the overall value time distribution of payments.

A factor that has not been discussed in the literature but may influence payment timing decisions is the industrial structure of banks that participate in the payments system. Industrial structure can differ for a number of reasons, many of which may be exogenous to banks' activities in the payments system. Such differences can result in varying costs of liquidity. In addition, the number of banks in a payments system—whether 10 or 1,000—can influence the likelihood of successful coordination of payments.

A number of studies provide empirical evidence on liquidity costs. McAndrews and Rajan (2000) document the payment and value timing distributions on the Fedwire funds service. McAndrews and Potter (2002) show the effects of the terrorist attacks of September 11, 2001, on the timing of payments over Fedwire. Mills and Nesmith (2008) find that the charging of overdraft fees on Fedwire in 1994 sped up the settlement of payments on the securities service but did not change the timing of payments on the funds service. Heller, Nellen, and Strum (2000) show that the introduction of intraday credit to banks in the Swiss Interbank Clearing system dramatically shifted the value time distribution of payments to earlier in the day. McAndrews (2006) finds that high-payment-value days lead to later value-weighted average payment value settlement times on Fedwire, consistent with a model of higher shadow prices of liquidity on high-payment-value days. Becher, Galbiati, and Tudela (2008), examining the timing of sterling payments on the Clearing House Automated Payment System (CHAPS), find that CHAPS has a less pronounced late-in-the-day peak than the Fedwire funds service does. The divergent patterns in Fedwire and CHAPS are broadly consistent with the models of Bech and Garratt (2003) and Bech (2008), but they also likely reflect the imposition and maintenance of throughput guidelines by CHAPS participants; the guidelines govern the percentage of value to be submitted at different times during the day. Finally, several papers use simulations to examine the trade-off between liquidity costs and delays in theoretical payments systems, including Koponen and Soramäki (2005) and Leinonen and Soramäki (2005).

3. Descriptive Analysis

Here we analyze payment and value time distributions on the Fedwire funds service. We focus on the number and value of payments transferred during a minute and contrast the time series of these variables in 1998 and 2006. Our work is purely descriptive; in the next section, we conduct regression analysis exploring the reasons behind the changes observed.

For our examination, we remove all payments to or from the settlement institutions: CHIPS, CLS Bank, and the Depository Trust Company (DTC). This allows us to focus on the non-settlement institutions' funds transfers on Fedwire, as these are subject to the strategic decisions of the sending party. Notably, this approach excludes the early-morning activity of CHIPS and CLS Bank.

3.1 General Pattern in the Timing of Payments

Charts 1 and 2, respectively, present the probability distribution functions for the percentage of the daily number of transfers and the percentage of daily value settled. Each point on Chart 1 (Chart 2) represents the average number (value) of payments transferred during that minute expressed as a percentage of the total number (value) of payments transferred that day. The charts show that the timing of Fedwire payments exhibits a general pattern that remained essentially stable between 1998 and 2006, whereby both the number and value of payments peaked in the late afternoon. We start by describing this general pattern.

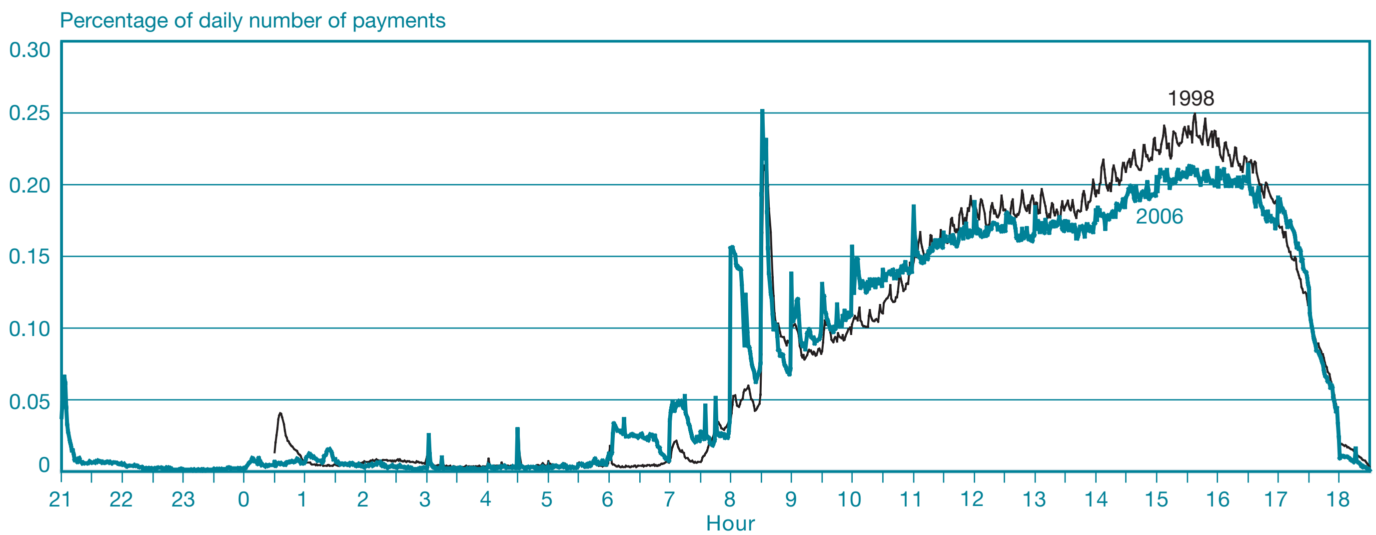
Chart 1 shows that relatively few payments are sent before 08:00, with the notable exception of the period following the opening of Fedwire (00:30 in 1998 and 21:00 in 2006), when many banks submit their CHIPS prefunding payments (recall that funds transfers to or from CHIPS and to or from CLS Bank are excluded from our measure of funds transfers in these charts). In the ten-minute period beginning at 08:30, there is a large spike in payment activity in 1998 and in 2006. The spike partially results from increases in bank balances attributable to the Federal Reserve Bank of New York's making principal and interest payments on behalf of the U.S. Treasury and government-sponsored enterprises (GSEs) as well as customer activity associated with the opening of the Fedwire securities service. After 08:30, payment volume initially declines, then steadily rises to a plateau between 15:00 and 16:00, when it reaches a level of activity similar to that of the 08:30 period. After 16:00, payment volume drops sharply until the close of Fedwire.

The value of payments settled also peaks in the late afternoon, but the peak is both sharper and later than that of the volume of payments (Chart 2). Like volume, value settled is low during the early-morning hours and it rises to its peak around the time DTC and CHIPS close. Note that while payment volume is falling after 16:00, payment value is rising. This is the result of the concentration of large-value payments late in the day. Finally, although discernible, the peaks around the opening of Fedwire and at 08:30 are proportionally much lower for the value of payments than for the volume of payments.

The timing of Fedwire payments exhibits a general pattern that remained essentially stable between 1998 and 2006, whereby both the number and value of payments peaked in the late afternoon.

Chart 1

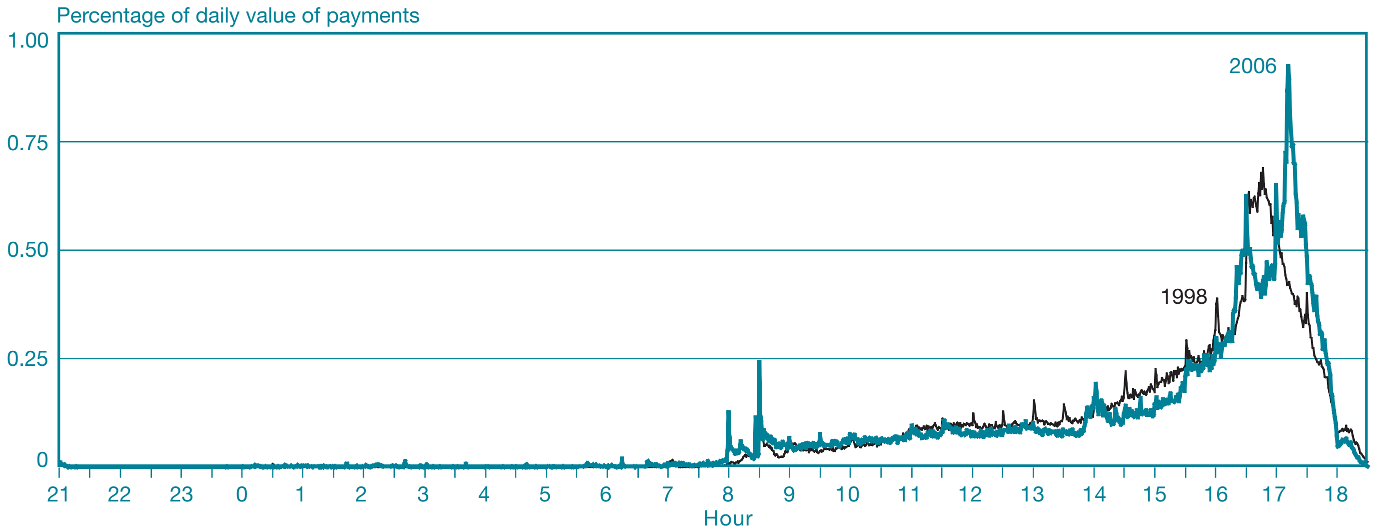
Fedwire Funds Payment Time Distribution, 1998 and 2006



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The chart shows the mean daily percentage of total payments settled in each minute. Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding.

Chart 2
Fedwire Funds Value Time Distribution, 1998 and 2006



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The chart shows the mean daily percentage of total payments settled in each minute. Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding.

3.2 Changes in the Distribution of Value

A comparison of Charts 1 and 2 indicates that the value time distribution has changed much more than the volume time distribution. Therefore, we now focus on the evolution of the value time distribution. Table 1 presents the times at which the deciles of payment value settled on average in 1998 and 2006. The earliest and latest deciles—10, 20, and 100—settled earlier

Table 1
Percentiles of Value Time Distribution

	1998		2006		
Percentile	Mean	Standard Deviation	Mean	Standard Deviation	Difference (Minutes)
10	11:15:21	00:09:34	10:37:37	00:15:22	-37.7
20	12:58:33	00:14:09	12:44:10	00:18:02	-14.4
30	14:24:21	00:15:41	14:33:51	00:17:60	9.5
40	15:19:59	00:14:58	15:38:09	00:13:25	18.2
50	15:59:29	00:14:33	16:15:04	00:10:44	15.6
60	16:28:18	00:12:09	16:37:29	00:08:22	9.2
70	16:46:05	00:10:46	16:58:58	00:07:09	12.9
80	17:02:47	00:10:43	17:14:06	00:04:20	11.3
90	17:26:25	00:11:27	17:30:12	00:03:53	3.8
100	18:37:28	00:23:05	18:31:14	00:07:15	-6.2

Sources: Federal Reserve Bank of New York; authors' calculations.

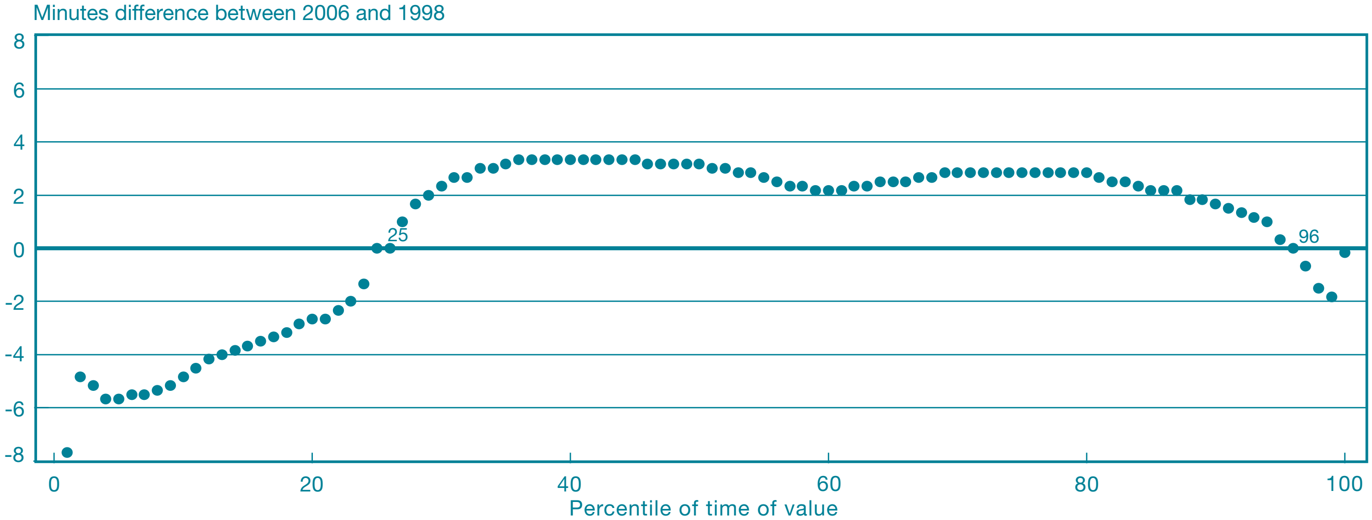
in 2006 than in 1998. The remaining deciles—30-90—moved to later-day settlement. Additionally, there is decreased variability in the settlement time of deciles 40-100, implying much more regularity in the later part of the value time distribution.

To understand the significance of these value shifts, we calculate the time at which each percentile of daily value settles for each day in 1998 and 2006. Then, for each percentile, we compare the distributions of the samples collected in 1998 and 2006 nonparametrically using the Mann-Whitney two-sample statistic. Our results appear in Chart 3. As the x-axis indicates, each of the 100 points corresponds to a percentile, moving left to right from percentile 1 to percentile 100. The y-axis corresponds to the number of minutes that must be added to or subtracted from the 2006 sample until the Mann-Whitney test is insignificant at the 5 percent level. For instance, the first point at the bottom of Chart 3 indicates that the first percentile in 2006 settled significantly earlier than the first percentile in 1998 by at least 7 minutes and 30 seconds. However, a point on the origin line indicates that the corresponding percentile cannot be distinguished statistically between the two samples.

Percentiles 1-24 are located below the origin axis, indicating that the lower percentiles of value were transferred earlier in the day in 2006 relative to 1998. Percentiles 27-95 lay above the origin line, indicating that most of the value distribution shifted to later in the day between 1998 and 2006. However, the points

Chart 3

Mann-Whitney U Test on Percentiles of Value Time, 1998 to 2006



Sources: Federal Reserve Bank of New York; authors' calculations.

Note: Minutes are subtracted from/added to each percentile until a Mann-Whitney rank-sum test is insignificant at the 5 percent level.

corresponding to the last percentiles, 96-99, are located below the origin line. It appears that the last percentiles of value were transferred later in the day in 1998. This result may be explained by a decline in the number and length of extensions since 1998.

The probability density function of the value time distributions in 1998 and 2006 is presented in Chart 2. Observe first that the supports of the 1998 and 2006 distributions in the chart have different lower bounds. Indeed, Fedwire operating hours were expanded when the opening hour was moved from 00:30 to 21:00 on May 17, 2004. Chart 2 shows that the change in the opening hour did not dramatically affect the distribution of value settled prior to 07:00. After 07:00, the two distributions intersect several times, suggesting no clear pattern to how the timing of payments changed between 1998 and 2006. We identify five distinctive features from our analysis of Chart 2:

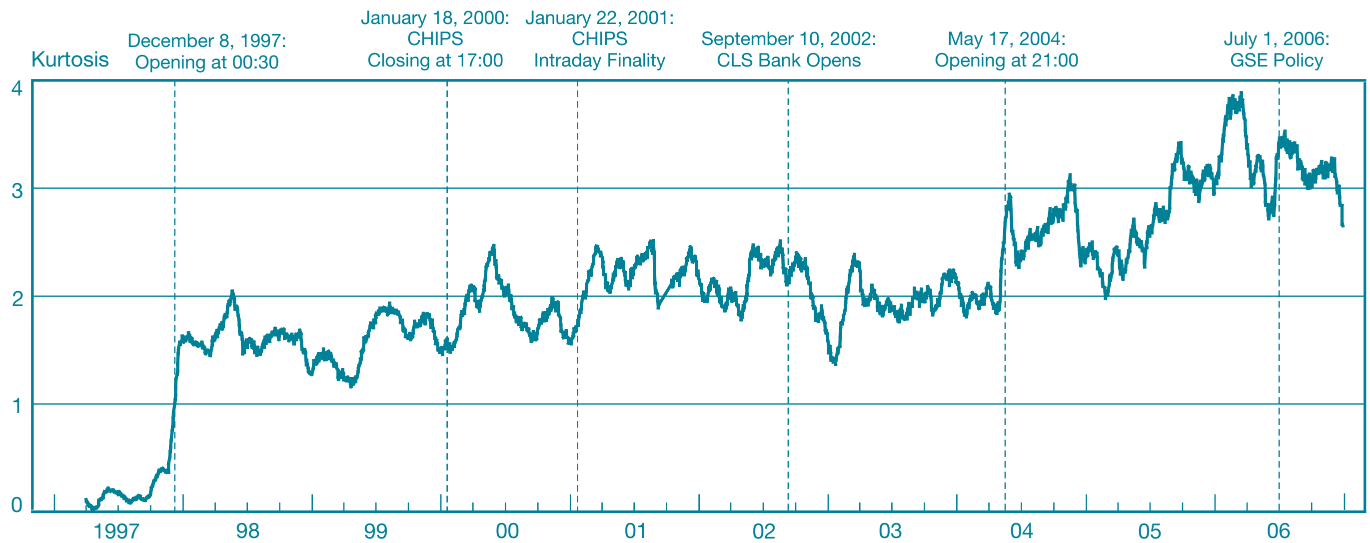
The highest peak of the distribution shifts to a later time, from around 16:48 in 1998 to around 17:11 in 2006.

1. The distribution of value settled becomes more concentrated. In particular, observe that the magnitude of the highest peak is greater in 2006 than in 1998. To confirm this observation, we plot in Chart 4 the evolution of the kurtosis of the payment value distribution. The chart shows a clear positive time trend between 1998 and 2006. If we interpret the kurtosis as a measure of peakedness, Chart 4 confirms that the distribution of value settled has become more concentrated over time. This result is important from a policy perspective, as it reflects a greater coordination in the timing of payments among Fedwire participants. A by-product of the greater coordination of payment activity, the amount of value transferred on Fedwire that is offsetting within a ten-minute period rose significantly, from 56 percent in 1998 to 58 percent in 2006. As the amount of offsetting payments rises, banks enjoy greater economy in the use of liquidity.
2. The highest peak of the distribution shifts to a later time, from around 16:48 in 1998 to around 17:11 in 2006. In other words, the minute during which most of the daily value is transferred is now twenty-three minutes closer to closing time. Chart 5 shows the distribution of value settled by time for each year between 1998 and 2006. It is clear from the chart that the highest peak moved to a later time between 1999 and 2000, and it was not a gradual move. The mean daily time of the top ten contiguous minutes of Fedwire funds value moved twenty minutes later, from 16:48 to 17:08, and the median daily time moved twenty-three minutes later, from 16:48 to 17:11, supporting the presence of a significant shift toward a later time in the peak of the distribution.[3]

[3] This difference is statistically significant at the 1 percent level for the Mann-Whitney two-sample statistic.

Chart 4

Kurtosis of Fedwire Funds Value Time Distribution

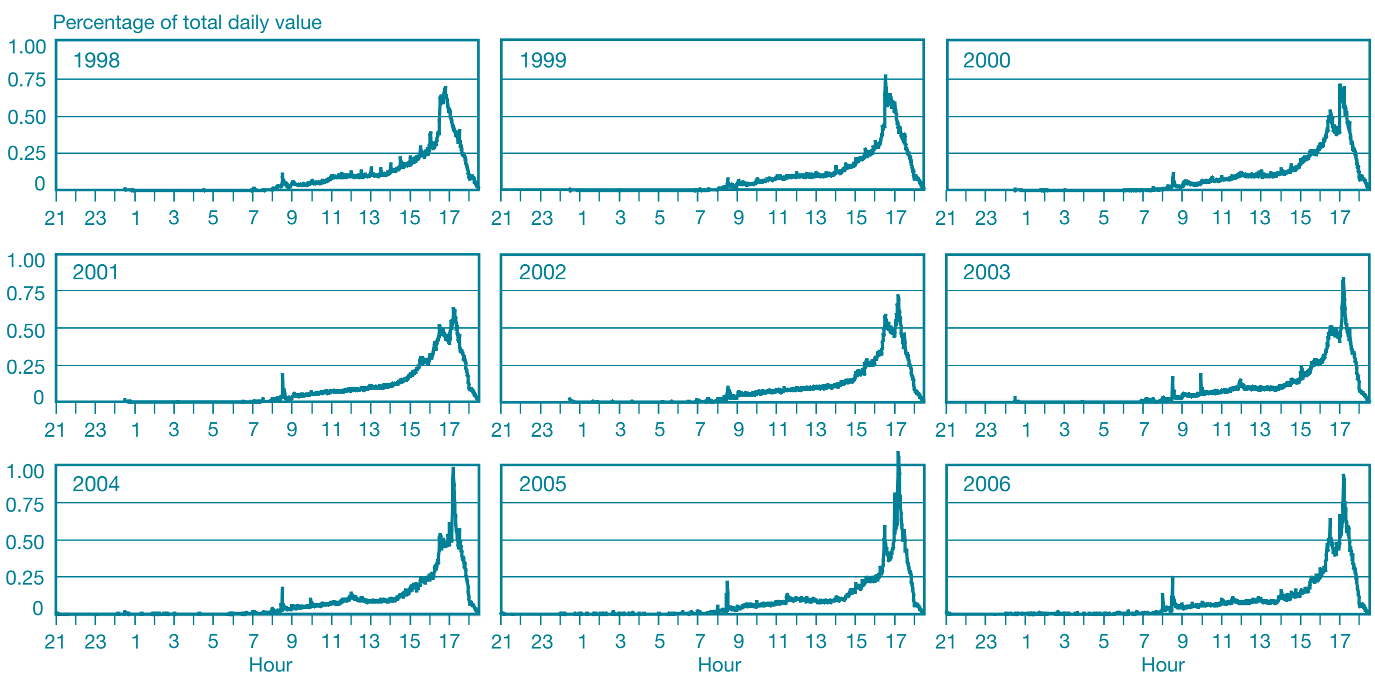


Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: Kurtosis is the excess kurtosis. A twenty-one-day centered moving average is used. Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding. GSE is government-sponsored enterprise.

Chart 5

Fedwire Funds Value Time Distribution by Year



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The panels show the mean daily percentage of total payment value settled in each minute. Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding.

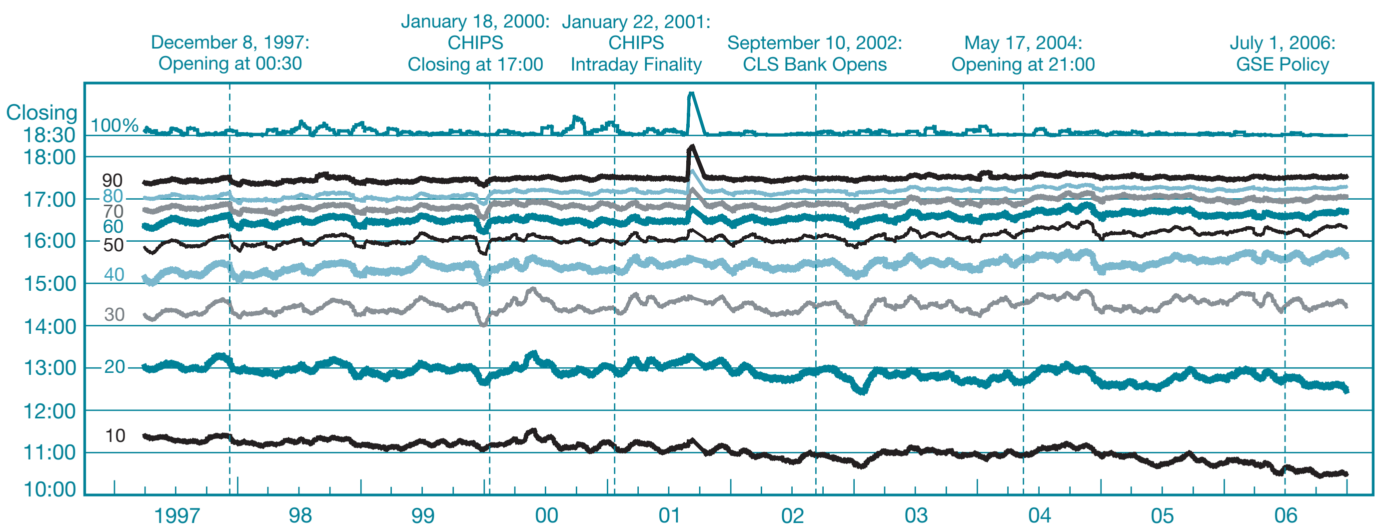
3. At the time of highest activity (between 16:00 and 17:30), the 2006 distribution exhibits three distinct peaks: two of comparable magnitude at precisely 16:30 and 17:00 and the third (the sharpest) at around 17:11. By comparison, the 1998 distribution possesses only two peaks: the sharpest at around 16:45 and a slightly smaller peak at 16:30. An analysis of the distribution of value settled for each year between 1998 and 2006 indicates that: 1) the 16:30 peak is present every year throughout the sample period; 2) as documented above, the highest peak moved from 16:48 to 17:11 between 1999 and 2000; and 3) the emergence of the 17:00 peak can be traced to 2004.
4. The 1998 distribution of value settled exhibits regular clock effects. Indeed, as indicated by the equidistant spikes in the 1998 probability distribution in Chart 2, there seems to be a flurry of activity every half-hour on the half-hour between 11:30 and 16:30. In contrast, these clock effects are not as discernible in the 2006 distribution. An analysis of the distribution of value settled for each year between 1998 and 2006 indicates that the clock effects gradually dissipate until they virtually disappear in 2002. We have not been able to identify what causes these clock effects and why they have faded away over time. In particular, it is unclear whether the effects are attributable to technological factors, the behavior of Fedwire participants, or institutional constraints.
5. The 2006 distribution exhibits a higher amount of activity at precisely 08:00 and 08:30. An analysis of the distribution of value settled for each year between 1998 and 2006 indicates that the 08:30 peak increased gradually over time, while the 08:00 peak is present only in 2006. The 08:00 peak is likely associated with the Federal Reserve's July 2006 change to its Payments System Risk Policy regarding GSEs, while the 08:30 peak is likely associated with the increased importance of the securities markets and the opening of the Fedwire securities service at 08:30.

We conclude this section with an analysis of Chart 6, in which the deciles of daily value settled are presented as a time series spanning the period from September 1997 to February 2007. The chart identifies several discrete events that may have affected the timing of Fedwire payments: changes in operating hours, changes in CHIPS operations, changes in CLS Bank operations, and changes in the Federal Reserve's Payments System Risk Policy regarding GSEs. Chart 6 provides a slightly different perspective on how the timing of payments evolved over time. Five points in particular are worth noting:

1. The deciles exhibit different trends. The first two deciles show a negative trend while the deciles between 40 percent and 90 percent indicate a slightly positive trend. These results are consistent with the dual shift we identified earlier when comparing the value of payments in 2006 and 1998. Indeed, we found that in 2006 the value settled moved toward earlier payments at the beginning of the day and toward later payments later in the day.

Chart 6

Deciles of Fedwire Funds Value Time Distribution



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: A twenty-one-day centered moving average is used. Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding. GSE is government-sponsored enterprise.

2. The first four deciles are not grouped together as much as the five deciles between 50 percent and 90 percent are. This observation is consistent with the shape of the payment value time distribution in Chart 2, which is highly concentrated toward the end of the day.
3. After 2004, the 100th percentile rarely exceeds the 18:30 closing time, indicating that extensions of Fedwire operating hours occur less frequently.
4. The deciles exhibit various peaks and valleys. The reasons for some of these peaks and valleys are clear, such as September 11, 2001, which led to later payments. Others are not as obvious.
5. The events documented in Chart 6 do not have a clear effect on the evolution of the times at which the various deciles of payment value settle. In particular, it is difficult to conclude unambiguously from the chart whether trends in the percentiles of value can be imputed directly to any one of these events. In the next section, we conduct a regression analysis to disentangle the effects of various factors on the timing of payments.

4. Regression Analysis

4.1 Model and Data

Our regression analysis identifies the factors that affected the distribution, or at least part of the distribution, of payment values. After experimenting with different specifications, we settled upon an easily interpretable yet robust model consisting of 100 linear regressions, each estimated separately for a given percentile of value.[4] To address possible serial correlation and heteroskedasticity problems, we relied on the approach developed by Newey and West (1987) to correct the estimated standard errors.[5] Finally, we conducted various diagnostic tests and compared the results of several alternative specifications to ensure that our results are robust.

[4] We recognize that there may be better specifications as well as more efficient inference techniques for analyzing the Fedwire value time distribution. In particular, since we do not estimate the joint distribution of all percentiles, we are not able to compare statistically point estimates across neighboring percentiles. Instead, we contrast only how a variable of interest affects different parts of the distribution, such as the low percentiles (corresponding to the morning) and the high percentiles (corresponding to the late afternoon). Observe, however, that: 1) our specification does not imply that the times at which the percentiles of value settle are independent and 2) if we can assume that our system of regression equations has the structure of a Seemingly Unrelated Regressions (SURE) model, then there is no loss of efficiency in estimating the regressions separately rather than jointly by GLS, since the explanatory variables are the same in each percentile regression.

[5] Note that we also estimated the model with lagged (up to ten lags) dependent variables. The results remain virtually unchanged and the differences are strongly insignificant.

Our sample consists of daily observations for virtually every business day between March 1998 and November 2006.[6] In a given regression, the dependent variable is defined as the time at which the corresponding percentile of value settled on a specific day, which we measure in the number of seconds since the day's Fedwire opening. The same set of explanatory variables is used in each of the 100 regressions. A formal definition of these variables as well as their sources can be found in Appendix A. Drawing on the literature we reviewed earlier, we include a number of potentially relevant variables in our analysis, which we organize into five categories: value and volume, Federal Reserve policies and operations, settlement system activities, other control variables, and calendar effects. Summary statistics for the independent variables are presented in Table 2.

Our regression analysis identifies the factors that affected the distribution, or at least part of the distribution, of payment values.

Value and Volume

The value as well as the number, or volume, of payments transferred over Fedwire may play a role in determining when Fedwire participants submit payments. To account for these effects, we disaggregate the nonsettlement daily value of Fedwire funds into four mutually exclusive groups. More specifically, we differentiate: 1) the total value transferred by banks on behalf of their customers, 2) the total value of deliveries of federal funds purchases and sales, 3) the total value of federal funds returns (of the prior day's deliveries), and 4) the value of all other interbank transfers, thereby consisting of payments not included in the groups made on behalf of customers or as part of a federal funds purchase, sale, or return. All else equal, and controlling in particular for the number of Fedwire payments, we observe that a higher value of transfers should result in a higher demand for daylight credit, which would lead to a higher shadow cost of liquidity. As a result, one may anticipate that higher values of payments lead to later settlement of payments.

[6] We are missing data for the following dates: April 1, 1997; December 22-24, 1997; March 1-3, 1999; April 20-22, 1999; October 14-15, 1999; October 18, 1999; and November 9, 1999.

Table 2
Summary Statistics of Independent Variables

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
Target federal funds rate (percent)	3.61	4.00	1.88	1.00	6.50
Operating hour extension (minutes)	00:04:05	00:00:00	00:19:06	00:00:00	05:16:00
Interbank payment value (billions of dollars)	487.08	476.19	75.40	148.79	865.82
Customer payment value (billions of dollars)	610.25	585.75	144.39	152.27	1334.25
Federal funds deliveries (billions of dollars)	250.83	257.65	65.70	0.92	472.64
Federal funds returns (billions of dollars)	250.09	257.84	64.21	0.92	432.61
Payments greater than or equal to $10 million	0.908	0.909	0.011	0.853	0.934
Number of payments (thousands)	465.237	453.817	79.101	186.895	904.726
Federal funds deviation	-0.01	0.00	0.13	-1.56	1.81
HHI of Fedwire value	529.8	516.5	107.0	220.2	795.3

Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: Payments greater than or equal to $10 million is defined as the fraction of daily value from payments greater than or equal to $10 million, excluding CHIPS, CLS Bank, DTC, and principal and interest funding payments. HHI is the Herfindahl-Hirschmann Index, a measure of the concentration of payment activity among banks.

As the size of payments varies greatly, we also include in the regressions the share of daily value consisting of individual payments in excess of $10 million in value. We chose $10 million as a threshold because it is quite close to the top 1 percent of individual payments when ranked by value. We hypothesize that a large proportion of high-value payments would cause a greater need for daylight credit, resulting in higher shadow costs of liquidity and later payments.

Finally, we control for the number of payments.[7] After we hold the values of payments constant, an increase in volume results in smaller individual payments on average. Therefore, one may expect that higher volume would be associated with faster Fedwire settlement, as smaller individual payments are less likely to require drawing in daylight credit. An alternative hypothesis is that a higher volume of transfers places higher operational demands on banks to check the credit lines of customers and other processes associated with submission of payments to Fedwire, and it could therefore result in later settlement.

Federal Reserve Policies and Operations

The Federal Reserve Banks offer many Fedwire participants access to daylight overdrafts for a fee and subject to upper limits, as described in the Federal Reserve's Payments System Risk Policy. These upper limits on the amount of daylight overdrafts that can be extended to a bank participant are called net debit caps. We include as a dummy variable the date of the liberalization of net debit caps that occurred on February 21, 2002, which allowed foreign banking organizations to modify the net debit cap calculation for U.S. branches and agencies of foreign banks (we call that variable the Foreign Capital Equivalency Policy). That change resulted in a one-time increase of approximately 10 percent in the aggregate net debit caps of Fedwire participants. We argue that this change in policy lowered the costs of liquidity for these banks and should result in earlier settlement of payments. Because the policy change was applied primarily to foreign banks—many of which participate in CHIPS and CLS Bank, which have early-in-the-day activity—we expect this faster settlement to affect mainly the lowest percentiles of Fedwire activity.

Another change in the Payments System Risk Policy, which occurred on July 20, 2006, restricted GSEs and certain international organizations from incurring daylight overdrafts. In general, this change represents a restriction of access to daylight credit, and we would expect it to correspond to an increase in liquidity costs. Thus, we hypothesize that this change, which we capture with a dummy variable, would result in slower payment settlement. Not all GSEs are exactly comparable, however. In particular, a distinctive feature of Fannie Mae and Freddie Mac is their payment of principal and interest on the 15th and the 25th of each month (or on dates close to the 15th and 25th when they fall on weekends or banking holidays). We therefore include an additional dummy variable to measure whether the change in policy affects the

[7] Note that although the value and volume variables drift during the sample period, we find no evidence suggesting that they may be nonstationary. More specifically, after conducting a series of Dickey-Fuller tests (with ten lags) and Phillips-Perron unit-root tests, we rejected the nonstationary hypothesis at the 1 percent significance level.

timing of Fedwire payments differently on these specific dates. We do not have an unambiguous prediction about the effect of this policy change on the timing distribution. On the one hand, the change may generate more delays, as they represent a restriction on access to liquidity; on the other hand, it may accelerate Fedwire payments, because GSEs must have funds delivered to them prior to releasing their principal and interest payments.

We include a dummy variable for all dates after the Federal Reserve extended, in May 2004, the opening hours of Fedwire to 21:30. We contend that this change may increase the submission of early-morning payments to Fedwire. Occasionally, the Federal Reserve Banks decide to extend the hours of Fedwire operation because of significant operational problems of a participant or the system. As such, we include a variable that measures the duration of the extension in minutes. We hypothesize that extensions may increase settlement risk and uncertainty and are associated with later payments—especially for the final few percentiles of payment value—and therefore they slow the settlement of the later percentiles of funds transfers. Finally, we include the target fed funds rate as a variable to control for any effect that monetary policy decisions might have on the timing of payments.

Settlement System Activities

Every day, most financial institutions are active simultaneously in a number of markets and payments systems—in particular, in the settlement systems CLS Bank, CHIPS, and DTC. (See Appendix B for a description of these systems.) This activity may affect the liquidity available to these participants at a given time during the day, which in turn may influence the time at which Fedwire participants decide to submit payments. To control for the influence of settlement system activity on the timing of Fedwire payments, we include a dummy variable for the dates after which CLS Bank began operation. Because CLS Bank operates early in the day in the United States, we reason that it may quicken the settlement of Fedwire payments submitted in the morning.

We also include variables measuring the times at which CHIPS and DTC conduct their late-afternoon settlements.[8] We hypothesize that the times of these settlement systems are associated with decreased uncertainty and with rapid redistribution of balances in various banks' accounts after these settlements are complete. Because of these effects, these times can also act as focal points for the settlement of other Fedwire payments (we discuss these points in more detail below). We conjecture that the time of these settlements will positively influence the timing of Fedwire payments: as their times move, so will the timing of Fedwire payments.

In addition to time variables associated with CHIPS, DTC, and CLS Bank, we include the value of U.S. dollar settlements conducted each day through CLS Bank, the values of the initial and final prefunding values in CHIPS, and the net-net settlement values in DTC. In our view, an increase in values settled through the settlement systems would increase demand for daylight credit, increasing its shadow cost, and result in later settlement times.

[Our analysis includes] variables measuring the times at which CHIPS and DTC conduct their late-afternoon settlements . . . . These times can also act as focal points for the settlement of other Fedwire payments.

Other Control Variables

We include both a constant and a time trend in our regression. The time trend is meant to control for trends, such as technological change, other than those captured by other covariates (for example, the volume and values of payments). In addition, we include the Herfindahl-Hirschmann Index (HHI) of payment market shares, a measure of the concentration of payment activity among banks. We expect this variable to control for industry mergers and other changes in the pattern of payments between banks. As discussed earlier, industrial structure can affect payment timing in a number of ways not fully examined in the literature. A more concentrated industrial structure might be able to coordinate payments more easily, but at either an earlier or later time of day. A more concentrated structure could result in more payment value being transferred by the larger bank that is more likely to have exceeded its deductible portion of the overdraft fee schedule, and therefore is more likely to economize liquidity actively. As a result, a greater concentration in industrial structure in the payment market could lead to later settlement. Finally, we include the interest rate spread between the effective federal funds rate and the target federal funds rate. We reason that when this spread is high, the net demand for end-of-day

[8] CLS Bank settles at multiple times in the early morning; because CLS Bank operates so early and its settlement time is so diffuse and therefore difficult to characterize, we do not include a settlement time variable for it.

balances is relatively high, which we would expect to be associated with later payment timing.

Calendar Effects

There are many predictable differences in payment activity across the days of the week and over the year. For example, Mondays predictably have higher volume than Fridays on average; days at the beginning of the month are likewise expected to be high value. To control for these effects on the timing of Fedwire payments, we include a number of dummy variables for various calendar effects. Included are dummies for the days of the week, the days preceding and following holidays, the first of the month, the last day of the quarter, the last day of the year and, separately, the last five days of the year. In addition, we include dummies for days on which the New York Stock Exchange is closed and, separately, days on which it closes early. Dummies are also included for each of the final five days of the reserve maintenance period (the day-of-week dummy captures both the effect of the day of the week and the effect of the first five days of the reserve maintenance period); we include that variable in case the reserve maintenance period influences payment activity. Finally, we include a separate dummy for the two-week period including and following September 11, 2001. During that period, Fedwire operating hours were regularly extended and were expected to experience later activity than normal.

Before analyzing our regression results, we emphasize one point. The variables described above may be expected to affect some parts, but not necessarily all, of the timing distribution. For instance, CLS Bank value may be considered likely to affect the payment distribution early in the day, but it may not necessarily have a lasting effect on the final percentiles of the distribution. Conversely, CHIPS final payout value may be considered likely to affect only the upper tail of the value distribution.

Results

To streamline our analysis, we present our estimation results graphically in Charts 7-11. We start by providing information about the interpretation of each chart. The x-axes of each chart display two scales. The bottom scale represents the percentile of value, or equivalently the regression number, moving from 1 to 100. The top scale represents the average time in 2006 at which the corresponding percentile settled during the day. The afternoon peak of the Fedwire value time distribution is evident when comparing these two scales. 12:00, which is fifteen hours after the opening of Fedwire, is only the 15th percentile of value time. By comparison, the hour between 16:00 and 17:00 includes more than twenty percentiles. Each chart corresponds to an explanatory variable; for instance, Chart 7.1 corresponds to interbank payment value. Each chart plots twenty points indicating the point estimate of the coefficient for that variable in the corresponding linear regression. As indicated in the chart notes, the color of the point identifies the level of statistical significance of the point estimate. On each side of a point estimate, we add a band representing the 95 percent confidence interval for this point estimate. Finally, we plot in Chart 12 the adjusted R^2 for each regression.

In terms of interpretation, a parameter significantly greater (lower) than zero in a regression for a given percentile indicates that the marginal effect of the corresponding explanatory variable delays (accelerates) the time at which that percentile settles.

In these charts, the results of multiple percentiles are shown on the same scale. This gives a full sense of each variable's effect on the timing of payments across the entire day. However, as we mentioned, comparisons between percentiles could be misleading and may overstate the economic effect of variables in the middle percentiles. Delaying a payment by the same amount of time becomes more costly as the end of the day approaches, when there is less time left to settle all remaining payments. For example, five minutes of delay at 17:30, when there is only one hour of the business day left, can be considered a larger economic effect than five minutes of delay at 12:00.

Value and Volume

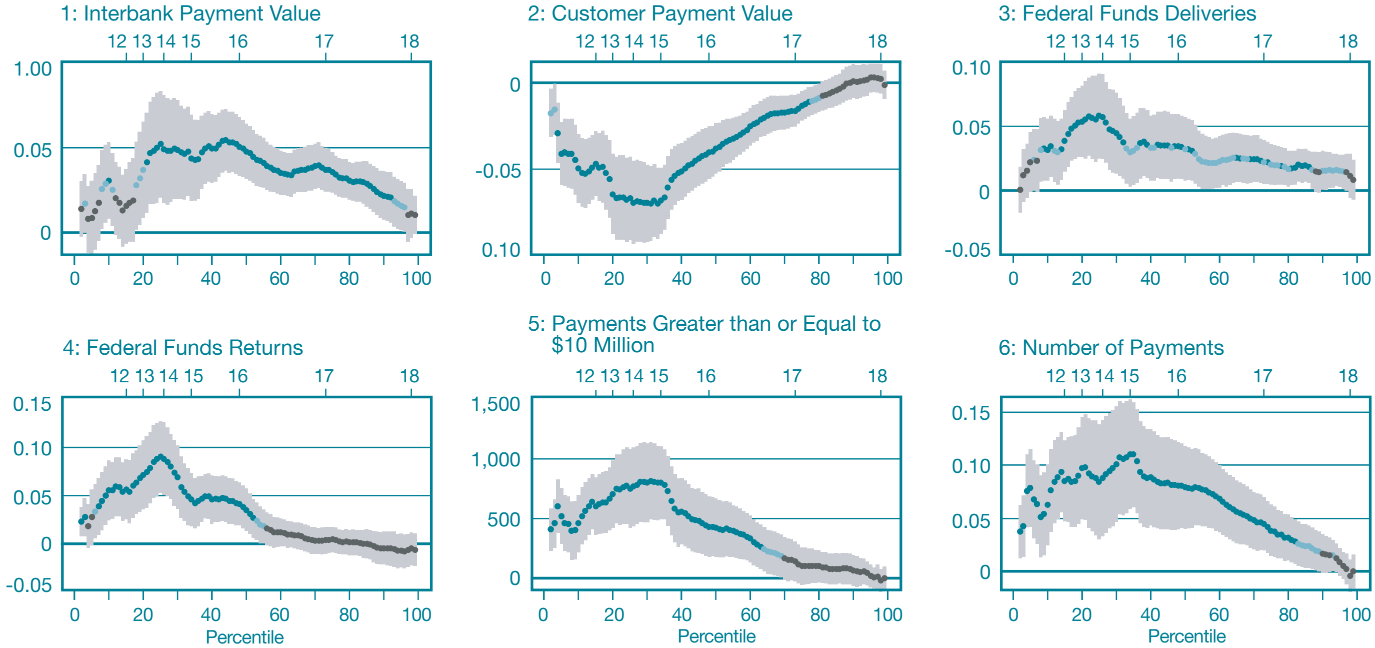
The parameters associated with the value of interbank payments settled over Fedwire are significant and greater than zero for virtually all percentiles (Chart 7.1). In other words, it appears that more interbank transfers over Fedwire tend to slow down the settlement of payments generally throughout the day. This result is consistent with the argument that banks have an incentive to delay their interbank payments, which are of high average value and may incur little delay cost because no customer may be demanding early settlement because of the cost of daylight credit.

In contrast, the parameters corresponding to the total value transferred by banks on behalf of their customers over Fedwire are negative and significant for all percentiles below 85 percent, that is, for payments submitted before 17:45 on average in 2006 (Chart 7.2). Fedwire payments therefore seem to settle earlier, when the value of transactions transferred by banks' customers is high. This result may be explained by the fact that, compared

Chart 7

Regressions of Fedwire Funds Value Time Percentiles

Value and Volume



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The upper x-axis displays the mean 2006 time for selected percentiles. The y-axis displays the value of the coefficient. The band represents the 95 percent confidence interval for the point estimates. We use independent ordinary least squares regressions with Newey-West standard errors (maximum lag = 10) for the 2nd to 99th percentiles of value. The color of the point indicates the significance of the coefficient: blue = 1 percent, light blue = 5 percent, dark gray = insignificant. There are 2,200 observations for each regression.

with interbank transfers, banks face a higher delay cost when acting on customers' requests for payments. In particular, banks may be asked by their customers to execute their transfers by a certain time.

The results suggest that both forms of federal funds activity tend to delay the timing of Fedwire payments. Indeed, the significant parameters in Charts 7.3 and 7.4 are systematically greater than zero. Observe, however, that federal funds returns appear to have a slightly larger effect earlier in the day, while the effect of federal funds deliveries persists throughout the day. These results may be considered surprising since both types of activities tend to occur later in the day. We conjecture that federal funds purchases and sales also capture the demand for overnight credit. This would therefore explain why higher federal funds deliveries are associated with delayed Fedwire payments throughout the day. Likewise, it is possible that banks expecting a return of federal funds may tend to delay their Fedwire payments in the morning until their accounts have been credited.

All else equal, an increase in the number of Fedwire transfers results in delayed payments for most of the day.

We now turn to the effect of the number of Fedwire payments transferred in a day. Virtually all parameters in Chart 7.6 are positive and significant for percentiles up to 80 percent, or equivalently for payments transferred before 17:45 on average in 2006. In other words, all else equal, an increase in the number of Fedwire transfers results in delayed payments for most of the day. This result seems to contradict our hypothesis that a greater number of transfers may expedite Fedwire payments, as it implies lower average size payments once we control for the total value transferred. Instead, the result possibly points toward greater operational costs, whereby banks must delay payments because it takes more time to process a greater number of payments.

The size of individual payments, however, is not completely neutral. Indeed, we find that the fraction of individual payments exceeding $10 million affects the timing of payments for percentiles up to 65 percent (Chart 7.5). In other words, large individual transfers delay payments submitted before 16:45 on average in 2006. This result does not unambiguously support our view that large individual payments lead to delayed settlement because they increase the likelihood of daylight overdrafts. We argue that our finding of no such delays after 16:45 may be explained by the fact that opportunities for multilateral netting are greater at the peak of Fedwire activity. As a result, banks may be less likely to delay large individual payments at the end of the day, as the risk of daylight overdraft decreases.

Federal Reserve Policies and Operations

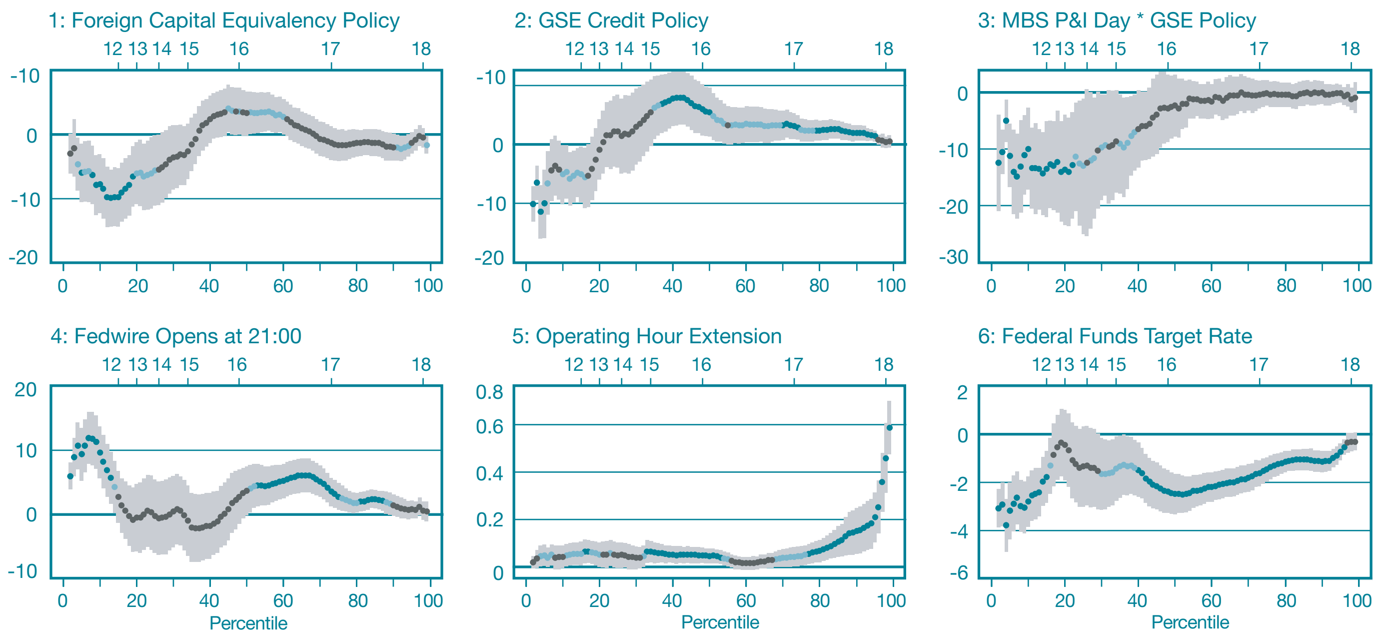
The July 1, 2006, modifications to the Federal Reserve's Payments System Risk Policy with regard to GSEs seem to have contributed to the delays in payments submitted after 15:00 on average in 2006 (Chart 8.2). Indeed, most estimated parameters for percentiles above 35 percent are significant and greater than zero. This result is therefore consistent with our hypothesis that the removal of access to intraday credit by GSEs and international organizations resulted in a shift toward later Fedwire payments. It is also consistent with the observation that the Federal Home Loan Banks decided to delay settlement of their principal and interest payments from 08:30 to approximately 14:00 after the implementation of the policy (but the delay effects persist throughout most of the remainder of the day).

As Fannie Mae and Freddie Mac are somewhat distinct from other GSEs, we also test whether the change in the Payments System Risk Policy had a specific effect on the 15th and 25th of the month, dates on which these two institutions make their principal and interest payments. Controlling for these specific dates over our entire sample period (see our discussion below) as well as for the policy change for all days following its implementation, we find that the timing of Fedwire payments shifted to earlier in the morning, but remained unchanged in the afternoon. Indeed, only the percentiles below 20 percent

Chart 8

Regressions of Fedwire Funds Value Time Percentiles

Federal Reserve Policies and Operations



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The upper x-axis displays the mean 2006 time for selected percentiles. The y-axis displays the value of the coefficient. The band represents the 95 percent confidence interval for the point estimates. We use independent ordinary least squares regressions with Newey-West standard errors (maximum lag = 10) for the 2nd to 99th percentiles of value. The color of the point indicates the significance of the coefficient: blue = 1 percent, light blue = 5 percent, dark gray = insignificant. There are 2,200 observations for each regression. GSE is government-sponsored enterprise; MBS is mortgage-backed securities; P&I is principal and interest.

are significant and negative (Chart 8.3). This result may be explained by a combination of two factors. First, compared with other GSEs, Fannie Mae and Freddie Mac did not delay markedly their payments of principal and interest after July 1, 2006. Second, after the policy change, these two GSEs found ways to have funds delivered to them earlier in the morning than they did prior to the change, in order to avoid having to draw on daylight overdrafts.

In another significant change to the Payments System Risk Policy, the February 21, 2002, Foreign Capital Equivalency Policy change increased the net debit cap of foreign banking organizations. We hypothesize that because foreign banking organizations conduct a larger percentage of their payments in the early Fedwire operating hours (from 21:00 to 08:00), which overlap with the operating hours of European and Asian markets, this change would affect the low percentiles of the Fedwire value time distribution. Chart 8.1 shows that the estimated parameters for percentiles below 20 percent are significant and negative, indicating that the change in Federal Reserve policies accelerated the submission of Fedwire payments in the morning. This result is therefore consistent with our prediction that the increase in net debit caps benefited mostly foreign banks submitting payments early in the morning, partly because of their active participation in CHIPS and CLS Bank.

We also find that the parameters associated with the duration of occasional extensions of Fedwire opening hours are significant and positive for most percentiles throughout the day. Observe also that the magnitude of the effect is significantly larger as the official closing time nears (that is, after 17:30). In other words, and in line with intuition, the duration of an extension is in general positively correlated with delays in the timing of payments, and it is particularly powerful for explaining delays at the end of the day.[9]

Finally, and somewhat surprisingly, our regression results in Chart 8.4 suggest that payments submitted in the morning were settled significantly later, not sooner, after Fedwire extended its opening hours from 00:30 to 21:00.

Settlement System Activities

We find no evidence supporting the hypothesis that the opening of CLS Bank had an effect on the timing of payments submitted through Fedwire (again, recall that payments to and from CLS Bank itself are removed from our measures). Indeed, all but a small number of estimated parameters associated with either the dummy variable capturing the opening date of CLS Bank (Chart 9.8) or the variable capturing the value of payments exchanged over CLS Bank (Chart 9.7) are insignificant. Therefore, the conjecture that the creation of CLS Bank may have helped speed up other Fedwire payments because it settles early in the day turned out to be unfounded.

In contrast, the other two settlement systems—CHIPS and DTC—appear to play major roles in shaping the value time distribution of Fedwire funds transfers, especially toward the end of the day. In particular, we identify significant delays in Fedwire payments submitted late in the afternoon (between 16:15 and 17:15 on average in 2006) on days DTC settles later.

CHIPS and DTC appear to play major roles in shaping the value time distribution of Fedwire funds transfers, especially toward the end of the day.

Indeed, the point estimates in Chart 9.6 are significantly greater than zero in the regressions conducted for the percentiles 55 percent to 85 percent. As we discuss in the next section, this result is particularly relevant, as DTC settlement typically occurs near the time of highest Fedwire activity.

Likewise, we find that the time of CHIPS settlement plays a significant role in explaining the upper tail of the value time distribution of Fedwire payments. Indeed, the estimations reported in Chart 9.1 suggest a strong positive effect, highly concentrated around the 75th percentile, which is very close to the time at which the highest peak of Fedwire activity occurs in 2006 (Chart 2). As we explore in greater detail in the next section, the emergence of the after-17:00 peak in Fedwire value transferred coincides with the change in CHIPS settlement time. In other words, the end-of-day shift in the timing of payments toward a later time may be traced in large part to the change in the timing of CHIPS settlement. Our regression results suggest that the January 2000 change in the timing of CHIPS settlement led to later settlement of the 65th-95th percentiles of Fedwire value.

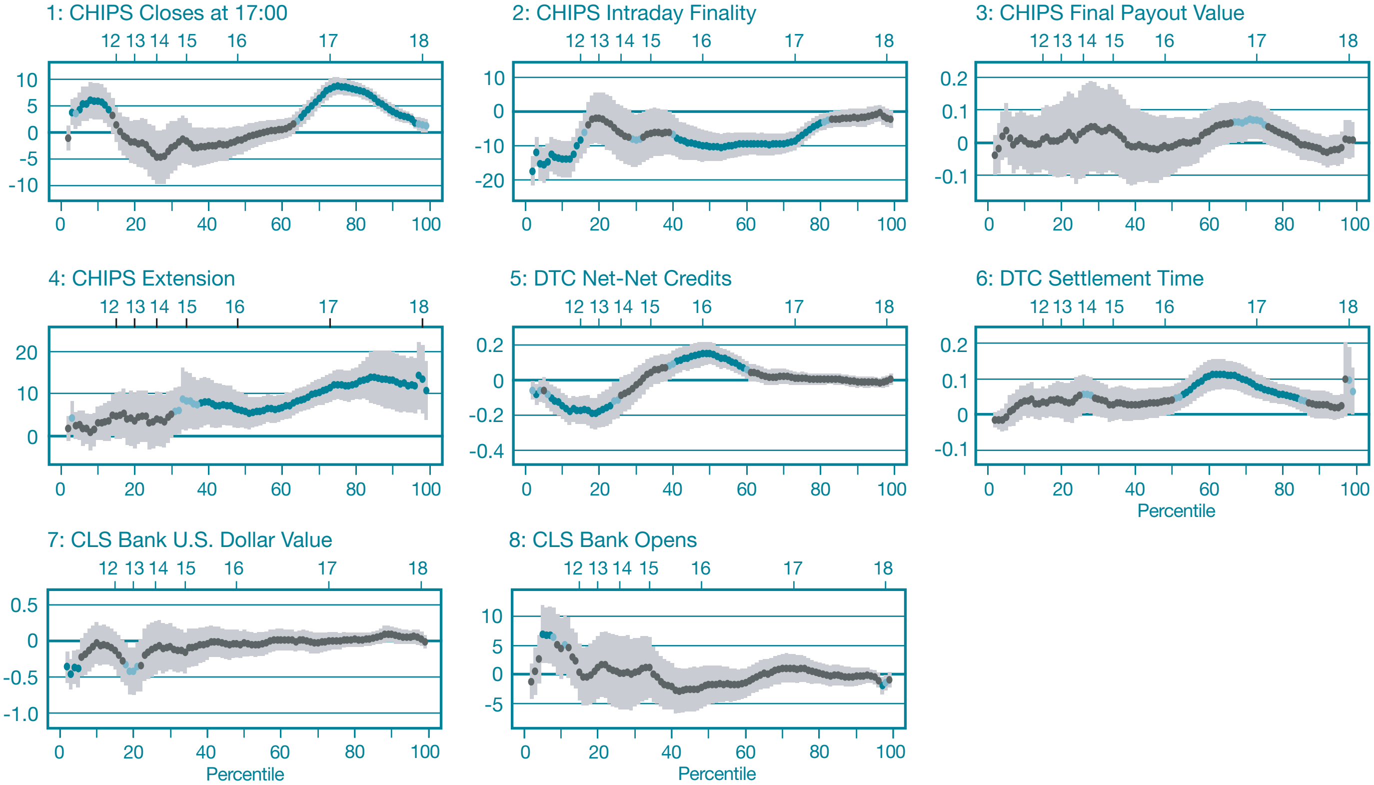
In contrast with the effects of the change in CHIPS settlement time, the change in the CHIPS settlement mechanism (to provide intraday finality of payments made via CHIPS) quickened the settlement of Fedwire payments throughout most of the day. Most of the estimated parameters in Chart 9.2 are significant and negative. This result may point to a consequence that the new CHIPS settlement mechanism has on customers: It may enable banks to credit their customers for payments made on CHIPS earlier in the day than was the practice before January 22, 2001. As a result of this earlier

[9] We find no indication that the occasional extensions of Fedwire operations may be endogenous. In other words, this variable may be considered as capturing only technical failure in Fedwire operations.

Chart 9

Regressions of Fedwire Funds Value Time Percentiles

Settlement Institutions



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The upper x-axis displays the mean 2006 time for selected percentiles. The y-axis displays the value of the coefficient. The band represents the 95 percent confidence interval for the point estimates. We use independent ordinary least squares regressions with Newey-West standard errors (maximum lag = 10) for the 2nd to 99th percentiles of value. The color of the point indicates the significance of the coefficient: blue = 1 percent, light blue = 5 percent, dark gray = insignificant. There are 2,200 observations for each regression.

redistribution of liquidity, banks' customers may now be able to submit other payments over Fedwire earlier.

Less clear is the influence of the variables capturing CHIPS and DTC values on the timing of Fedwire payments. For instance, we find that large CHIPS final payouts slow Fedwire payments at the end of the day (Chart 9.3). This result is not consistent with the hypothesis that, by releasing funds after it settles, CHIPS may accelerate payments made through Fedwire. Instead, it could suggest that CHIPS participants that have to make payments to settle their positions may experience a temporary liquidity squeeze that leads them to delay their Fedwire payments. Alternatively, it may reflect some greater uncertainties in banks' positions on days of high CHIPS settlements that cause increased delays of Fedwire payments.

The effect of DTC net-net credits is complicated. Chart 9.5 shows that larger DTC net-net credits appear to: 1) expedite Fedwire payments submitted in the morning (before 13:30), 2) slow mid-afternoon Fedwire payments (between 15:30 and 16:30), and 3) have no effect on payments submitted at the end of the day (after 16:30). We have no explanation for the first result. The second result may be explained by the fact that, as the level of activity on DTC increases throughout the day, liquidity available to other banks to make Fedwire payments is removed. Finally, the third result suggests that at the end of the day, the timing of Fedwire payments is affected only by the time at which DTC settles, not by the value of DTC net-net credits.

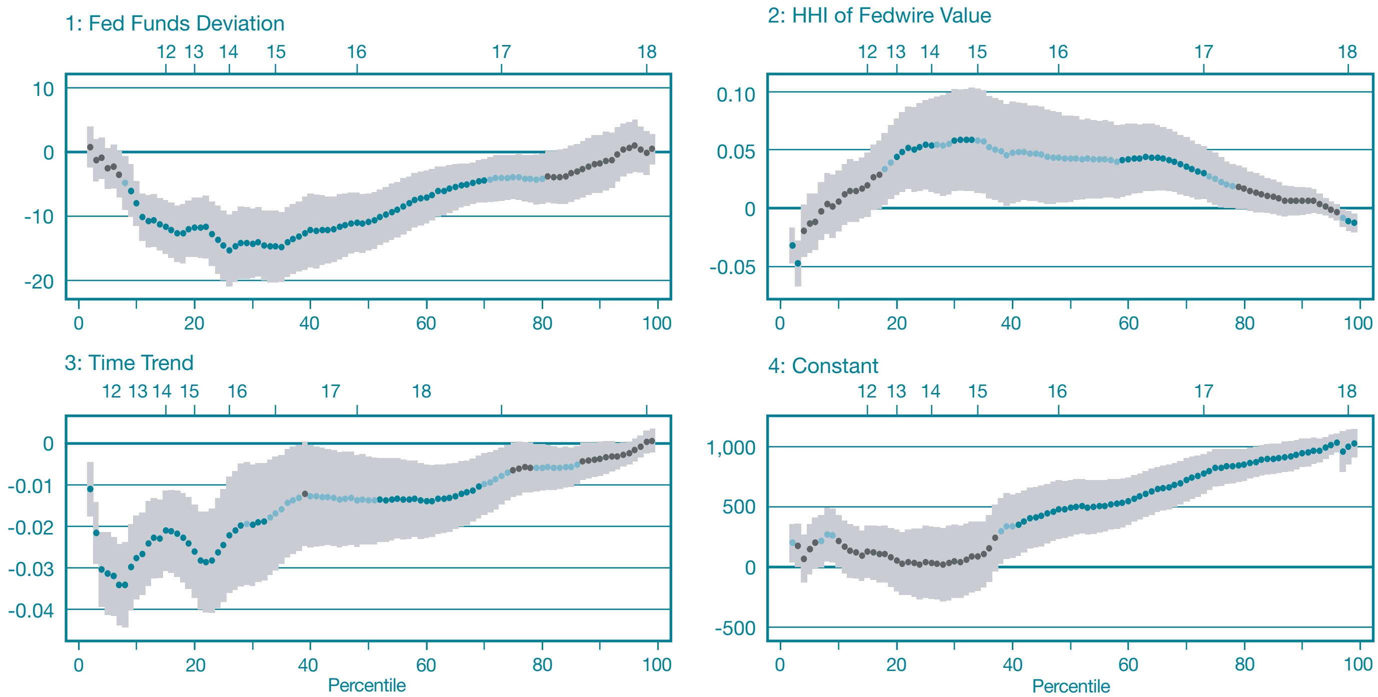
Other Control Variables

We find that a higher degree of industry concentration, as measured by the HHI, slows the transfer of Fedwire payments submitted after 12:00 up until the time of highest Fedwire activity. Indeed, most of the estimated parameters below the 75th percentile in Chart 10.2 are positive and significant. One

Chart 10

Regressions of Fedwire Funds Value Time Percentiles

Other Control Variables



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The upper x-axis displays the mean 2006 time for selected percentiles. The y-axis displays the value of the coefficient. The band represents the 95 percent confidence interval for the point estimates. We use independent ordinary least squares regressions with Newey-West standard errors (maximum lag = 10) for the 2nd to 99th percentiles of value. The color of the point indicates the significance of the coefficient: blue = 1 percent, light blue = 5 percent, dark gray = insignificant. There are 2,200 observations for each regression. HHI is the Herfindahl-Hirschmann Index, a measure of the concentration of payment activity among banks.

hypothesis is that an increase in industry concentration helps improve coordination and thereby facilitates the transfer of payments around a single point in time. Our finding that increased concentration in payment market shares tends to slow the settlement of payments might still be consistent with increased coordination, but it reflects the fact that coordination occurs around a later time of day, possibly at the peak of Fedwire activity.

An alternative hypothesis is that the increased concentration results in greater economization of liquidity by banks. The largest banks are more likely to pay positive overdraft fees and therefore face a positive marginal cost of liquidity. As payment shares move from banks with a zero marginal cost of liquidity to those with a positive marginal cost, we would expect settlement to occur later in the day. The nature of the increase in concentration is consistent with this hypothesis; the payment value market share of the top four banks has increased by 13 percentage points, while the share of banks ranked 5th through 50th has declined by 2 percentage points over the period. This shows that payment activity has moved from relatively small banks (market shares ranked below 50th), which face low marginal overdraft costs, to the largest banks (top four), which regularly face a higher positive fee for the use of their marginal daylight overdrafts.

The federal funds rate deviation shifts most of the distribution of Fedwire payments earlier (Chart 10.1). This result does not support the hypothesis that a positive deviation reflects a higher than anticipated demand for intraday liquidity by Fedwire participants and therefore would be associated with later Fedwire payments. Instead, we conjecture that the effect of federal funds deviations could be explained by Fedwire participants having an incentive to purchase federal funds early if they are trading at a higher than anticipated price.

Finally, the time trend is found to be significantly lower than zero for most percentiles during the day (Chart 10.3). Before we interpret this result, recall that in the descriptive analysis we identified a dual adjustment process between 1998 and 2006, with a trend toward earlier payments for low percentiles and a trend toward later payments for higher percentiles. The estimated time trend therefore appears to capture part of the first effect, but it is not consistent with the second. In other words, the move toward later Fedwire payments at the end of the day is explained in our regressions by explanatory variables other than the exogenous time trend. In addition, observe that the influence of the time trend on the timing of Fedwire transfers provides some support for the hypothesis that technological improvements, such as in queuing mechanisms used at various banks, may have contributed to accelerating transfers over Fedwire.

Calendar Effects

We now comment briefly on some of the major calendar effects. We find that compared with Thursdays, the timing of payments on Mondays (especially in the morning) and Tuesdays is delayed (Charts 11.5 and 11.6), but for the most part the timing is not significantly different on Wednesdays and Fridays, except for a marked end-of-week effect at the close of Fedwire on Fridays (Charts 11.7 and 11.8). Controlling for the days of the week, we find that the timing of Fedwire payments is virtually identical during the second week of the maintenance period (Charts 11.9-11.12), except for the Monday of the second week (Chart 11.11), when payments settle earlier.

Chart 11.1 indicates that Fedwire payments tend to settle earlier when Fannie Mae and Freddie Mac make principal and interest payments (on the 15th and 25th of the month). This effect was anticipated, because these payments are typically issued around 08:30, and therefore they provide Fedwire participants with an influx of liquidity early in the morning.

In addition, we find that, all else equal, Fedwire payments settle earlier: 1) on days when the New York Stock Exchange is either closed or closes early (Charts 11.2 and 11.3), 2) on the days preceding and following a holiday (Charts 11.14 and 11.15), 3) on the last days of each quarter (Chart 11.17), and 4) on the last five days of the year (Chart 11.18). In contrast, Fedwire payments tended to be submitted later on the first of the month (Chart 11.16) and during the week following September 11, 2001 (Chart 11.4).

Finally, observe that the adjusted R^2s are generally high (Chart 12), indicating that our regression models are able to capture a large part of the daily variations in the percentiles of Fedwire value. Note also that the adjusted R^2s tend to be closer to 1 for higher percentiles. This result is consistent with the fact that low percentiles, corresponding to payments settled in the morning, are in general more volatile from one day to the next.

4.2 Economic Significance

To put our regression results into a more general perspective, we conduct two exercises. First, we measure which variables can explain a later-than-normal settlement of Fedwire value on a given day in 2006. The second exercise measures the approximate economic contribution of various factors in explaining the shift in late-day payments between 1998 and 2006. To start, both exercises confine our attention to those variables that largely explain the variation in the timing of payments submitted after 17:00. This may be considered particularly relevant in light of the Board of Governors of the Federal Reserve System's 2006 consultation paper, which points out that the recent shift of the Fedwire activity peak to closer to the end of the day raises some concerns in terms of operational problems. Understanding the factors affecting the submission of Fedwire payments at the end of the day is therefore of particular interest when analyzing potential economic costs and benefits of alternative policy options.

Understanding the factors affecting the submission of Fedwire payments at the end of the day is . . . of particular interest when analyzing potential economic costs and benefits of alternative policy options.

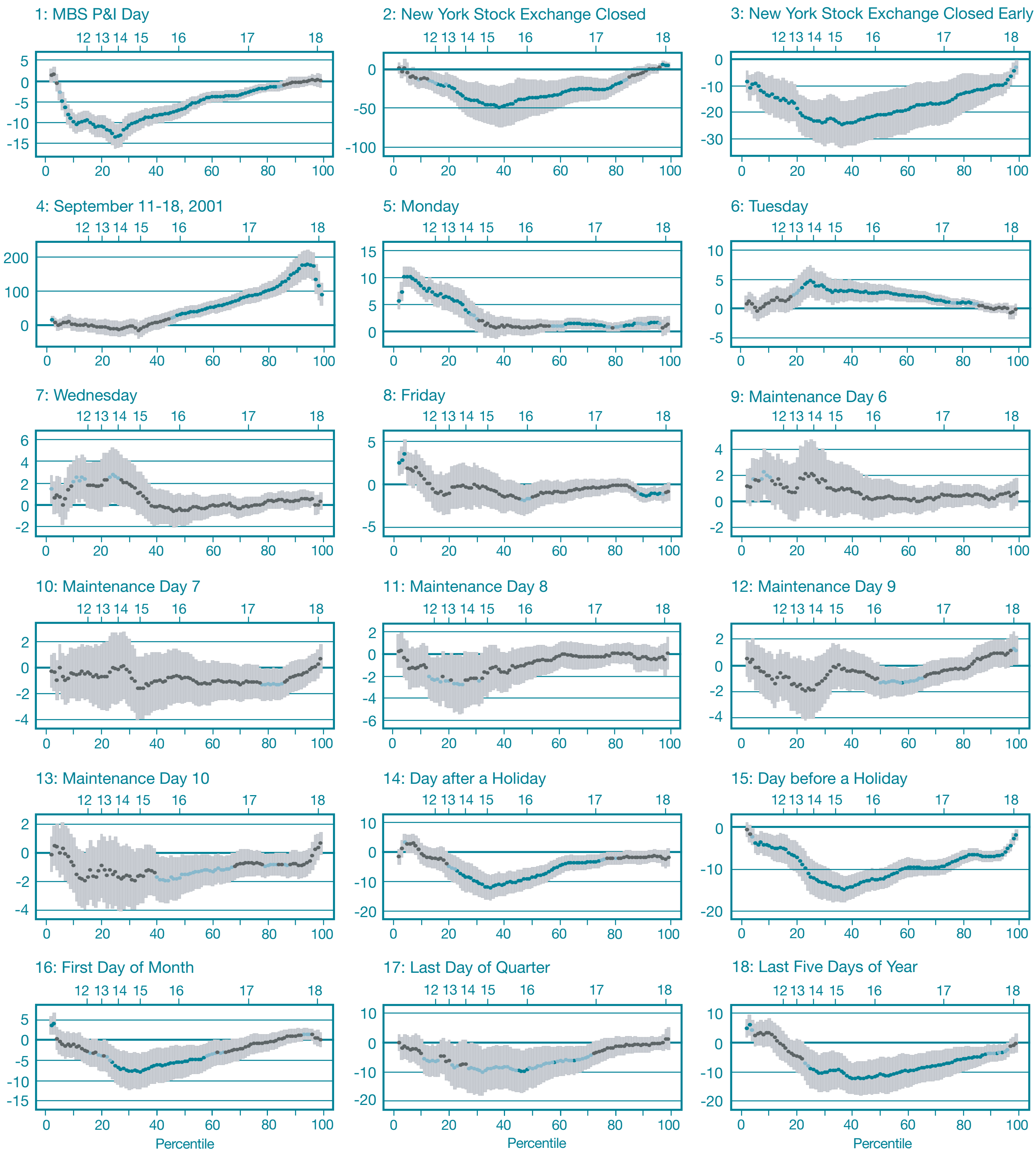
In our first exercise, we take out the effect of long-run growth in most continuous variables, such as volume and value, by removing the trend of each of the continuous covariates (those that vary in number or in value over time). Next, we measure a small day-to-day variation in the level of the variable—namely, a one standard deviation of this adjusted variable. We also consider a typical day by setting all calendar effects equal to zero and without an extension of Fedwire operation. Finally, we ignore the effect of past specific events, such as the May 17, 2004, extension of Fedwire operating hours and the creation of CLS Bank, as these events are not expected to repeat in the variation of activity on Fedwire from one day to the next. We then measure the economic significance of all of our variables in explaining the timing of payments made on Fedwire after 17:00 by multiplying the estimated coefficient by the one-standard-deviation change in the variable.

Surprisingly, we find that only three variables appear to play a significant economic role in explaining the timing of Fedwire

Chart 11

Regressions of the Fedwire Funds Value Time Percentiles

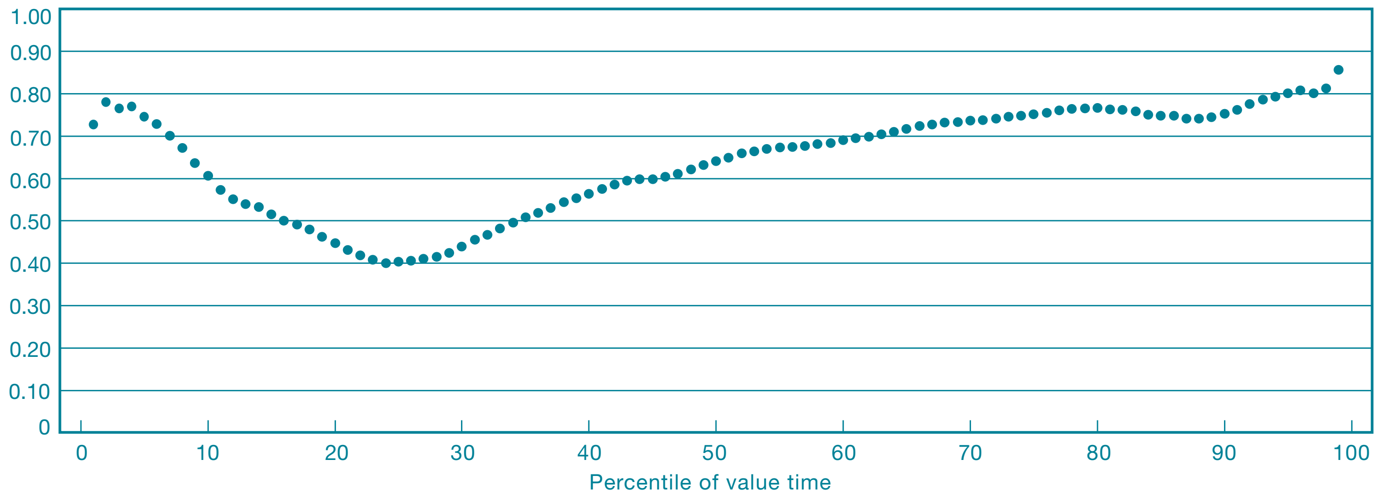
Calendar Effects



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: The upper x-axis displays the mean 2006 time for selected percentiles. The y-axis displays the value of the coefficient. The band represents the 95 percent confidence interval for the point estimates. We use independent ordinary least squares regressions with Newey-West standard errors (maximum lag = 10) for the 2nd to 99th percentiles of value. The color of the point indicates the significance of the coefficient: blue = 1 percent, light blue = 5 percent, dark gray = insignificant. There are 2,200 observations for each regression. MBS is mortgage-backed securities; P&I is principal and interest.

Chart 12
Adjusted R^2 of Regressions



Source: Authors' calculations.

payments submitted after 17:00 from one day to the next. In particular, a typical day-to-day variation in either the value of interbank payments or in the number of daily Fedwire transfers each delays payments submitted after 17:00 by roughly three minutes on average. In addition, we find that an extension of CHIPS operating hours delays virtually all payments submitted after 17:00 by ten minutes on average. We note that although large in magnitude, this effect does not necessarily constitute a major risk, as CHIPS extensions are rare in practice, especially since CHIPS changed its settlement mechanism (for instance, CHIPS extended its hours of operation only once between January 2005 and December 2006). Nevertheless, this effect does illustrate how significantly the operations of settlement institutions are interconnected. To summarize, this first exercise indicates that although numerous factors contribute to the shift in the time of highest Fedwire activity between 16:30 and 17:00, their effects appear to be confined to that period and do not spill over near Fedwire's closing time. This result should be reassuring, as it suggests that very few variables have an influence on payments submitted after 17:00 on a day-to-day basis.

[Our result] suggests that very few variables have an influence on payments submitted after 17:00 on a day-to-day basis.

Our second exercise evaluates the economically significant factors that contributed to the shift in the after-17:00 value time distribution of Fedwire payments between 1998 and 2006. Note, however, that our regression model is not perfectly suited to disentangle the respective contribution of each explanatory variable from the changes in the timing distribution observed during this period. Because most of the covariates varied jointly between 1998 and 2006, our model cannot pin down precisely the contribution of each explanatory variable to the shift in the timing of payments. The results presented here should therefore be interpreted as orders of magnitudes rather than exact measurements. Again, we are more interested in the end-of-day changes and we therefore focus on the changes in the 75th percentile, which roughly corresponds to the time of highest Fedwire activity in 2006.

Our model predicts that compared with 1998, the 75th percentile should have shifted fourteen and a half minutes later in 2006. This shift is slightly more than the thirteen minutes we actually observe in the data, but it is within two standard deviations. If we consider the effect of the time trend as being exogenous, we find the following:

1. When combined, the increase in the number and value of Fedwire payments between 1998 and 2006 accounted for slightly more than 40 percent of this shift in the 75th percentile toward a later time.[10]
2. The change in CHIPS closing time on January 18, 2000, if considered an exogenous event, contributed more than 30 percent by itself. This effect, however, is partially offset by the modification in the CHIPS settlement mechanism, which moved the 75th percentile earlier. As a result, the aggregate contribution of CHIPS to the later settlement

[10] In this measure, we include all variations between 1998 and 2006 in the values of interbank payments, customer payments, and federal fund deliveries and returns. We also include the increase in the share of individual payments greater than $10 million.

of the 75th percentile may be estimated at around 10 percent.

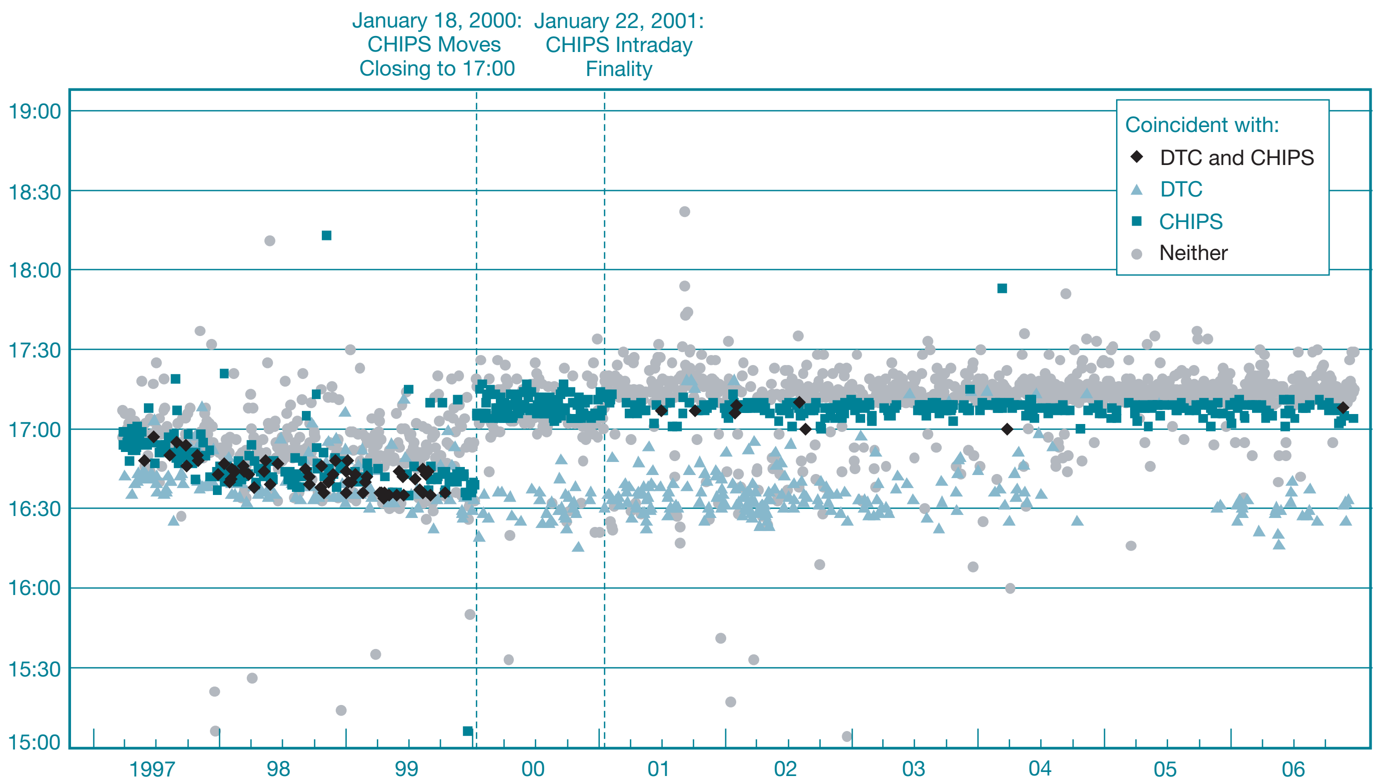
3. Finally, the last major contributor to the shift of the 75th percentile toward a later time is the higher concentration of payment activity among banks. We find that between 1998 and 2006, the increase in the HHI accounted for close to 30 percent of the shift.

5. Influence of CHIPS and DTC on Settlement Time

Our regression analysis points to a few specific variables as highly explanatory of the shift to later settlement of the 70th-90th percentiles of Fedwire activity. In this section, we present other evidence of the influence of the settlement institutions—CHIPS and DTC—on the value time distribution of Fedwire activity. Specifically, we consider the time at which the peak in Fedwire value transfer occurs. Recall Chart 2, which shows that the peak in Fedwire value transfer occurred before 17:00 in 1998 and after 17:00 in 2006.

To illustrate the dependency of the Fedwire value time distribution on the behavior of CHIPS settlement timing, we measure the timing of the peak of Fedwire activity as the midpoint of the ten contiguous minutes of highest value transferred during the day. Chart 13 displays the time of the peak ten minutes of value transferred on Fedwire from 1997 through 2006. Each point represents the time at which the midpoint of the top ten contiguous minutes of Fedwire value settled on each day between 1997 and 2006. We see that prior to January 18, 2000, there was a peak at approximately the same time daily (although that time varied from day to day). Its time trended downward from around 17:00 in 1997 to around 16:48 in early January 2000. After January 18, 2000, however, a distinctly new pattern emerged. Peaks tended to occur at two specific times: 16:30 and 17:11, with the most common peak at 17:11. This is consistent with the observation in Chart 5 that the value time distribution of Fedwire changed from a single-peak distribution to a dual-peak distribution in 2000.

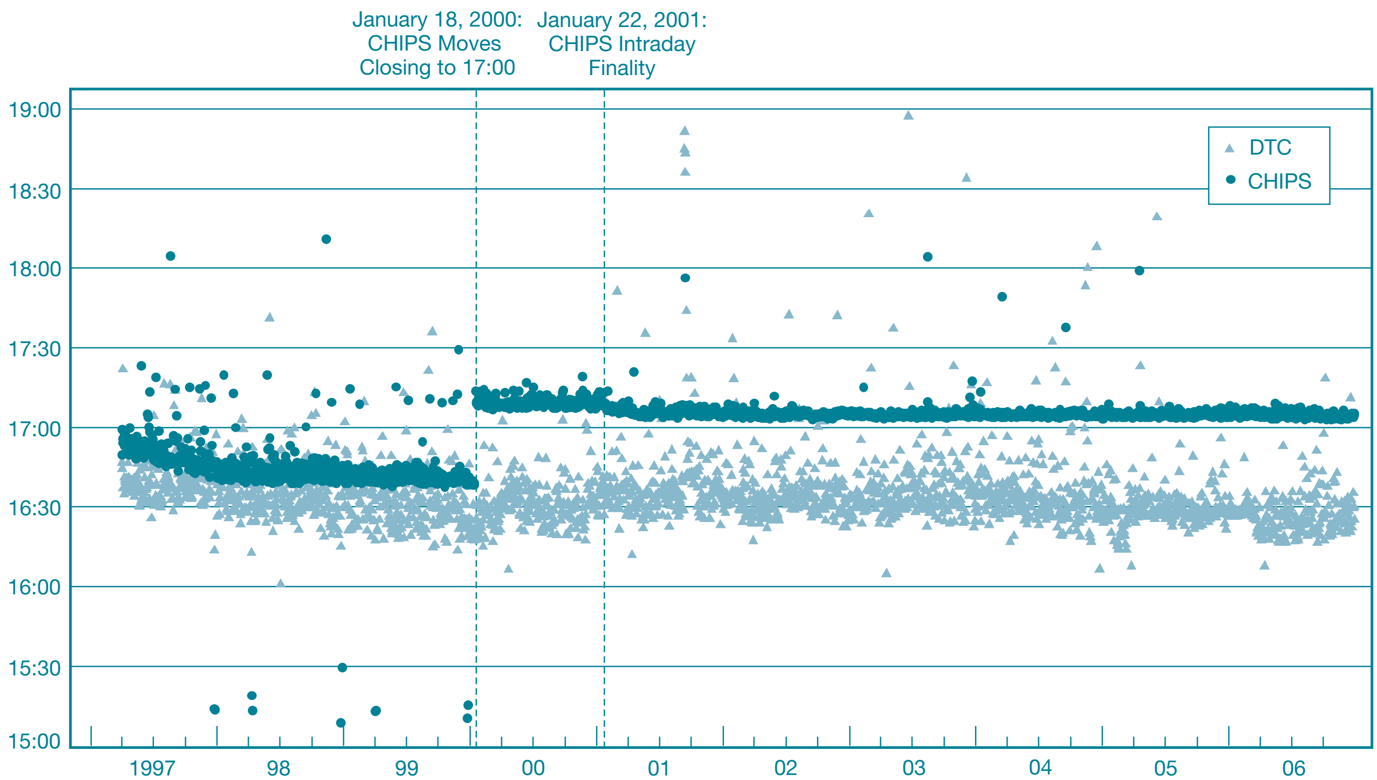
Chart 13
Time of Peak Ten Minutes of Fedwire Funds Value



Sources: Federal Reserve Bank of New York; authors' calculations.

Notes: Seven days settling after 19:00 were excluded. Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding.

Chart 14
End-of-Day Settlement Times of CHIPS and DTC



Sources: Federal Reserve Bank of New York; authors' calculations.

Note: Seven days settling after 18:30 were excluded.

The regression analysis suggests a reason for why the shift occurred. Prior to January 18, 2000, CHIPS effected its final settlement at approximately 16:40; on January 18, 2000 (and thereafter), it moved its settlement of final payouts to approximately 17:10, while the value-weighted time of DTC's settlement remained roughly constant over the period. Examination of Charts 13 and 14 suggests that prior to January 18, 2000, Fedwire's peak of settlement activity occurred simultaneously with the roughly coincident settlement times of CHIPS and DTC. After January 18, 2000, when CHIPS moved to a later settlement time, two peaks of settlement activity emerged on Fedwire. One coincided with DTC settlement time at 16:30, and the other moved more closely to CHIPS settlement time after 17:00. The distinct change in pattern, so closely matching the timing pattern of CHIPS and DTC, as well as the evidence from the regression analysis, points to the timing of the settlement institutions' late-in-the-day settlement as being highly explanatory in the timing shifts of Fedwire's peak and late-day activity over the 1998-2006 period.

Why should Fedwire's peak activity in value transfer coincide so closely with the final payouts of the major settlement institutions? We advance four hypotheses, which are not necessarily mutually exclusive: the bank liquidity cascade, the customer credit cascade, the resolution of uncertainty, and the role played by settlement times as focal points.

First, we advance our bank liquidity cascade hypothesis. Consider the activities of DTC. As we discuss in Appendix B, DTC accumulates balances in its account and releases them back to the banking system during its final payout procedures. That outflow of balances from DTC and the resulting inflow to banks can trigger a cascade of payments made by the receivers of DTC payouts, which triggers further payments made by the receivers of those payments, and so on. The cascade of payments can occur if banks are withholding payments because they face internal constraints attributable to a cost of liquidity or some other limit on their willingness to submit payments earlier. Beyeler et al. (2006) provide a model and simulation of a similar process.

A second reason for such a release of payments is the related customer credit cascade hypothesis. Not only do banks face constraints in making timely payments, but so do their customers. Those customers, or the depositors of banks, also receive funds following payouts on a major settlement system, which provides funds to their accounts. If those depositors had been withholding payments because they could not easily obtain credit to send payments, the inflow of settlement system payments could provide the needed funding for them to execute their withheld payments; the release of those depositor payments could result in the release of payments by the receivers of their payments, and so on.

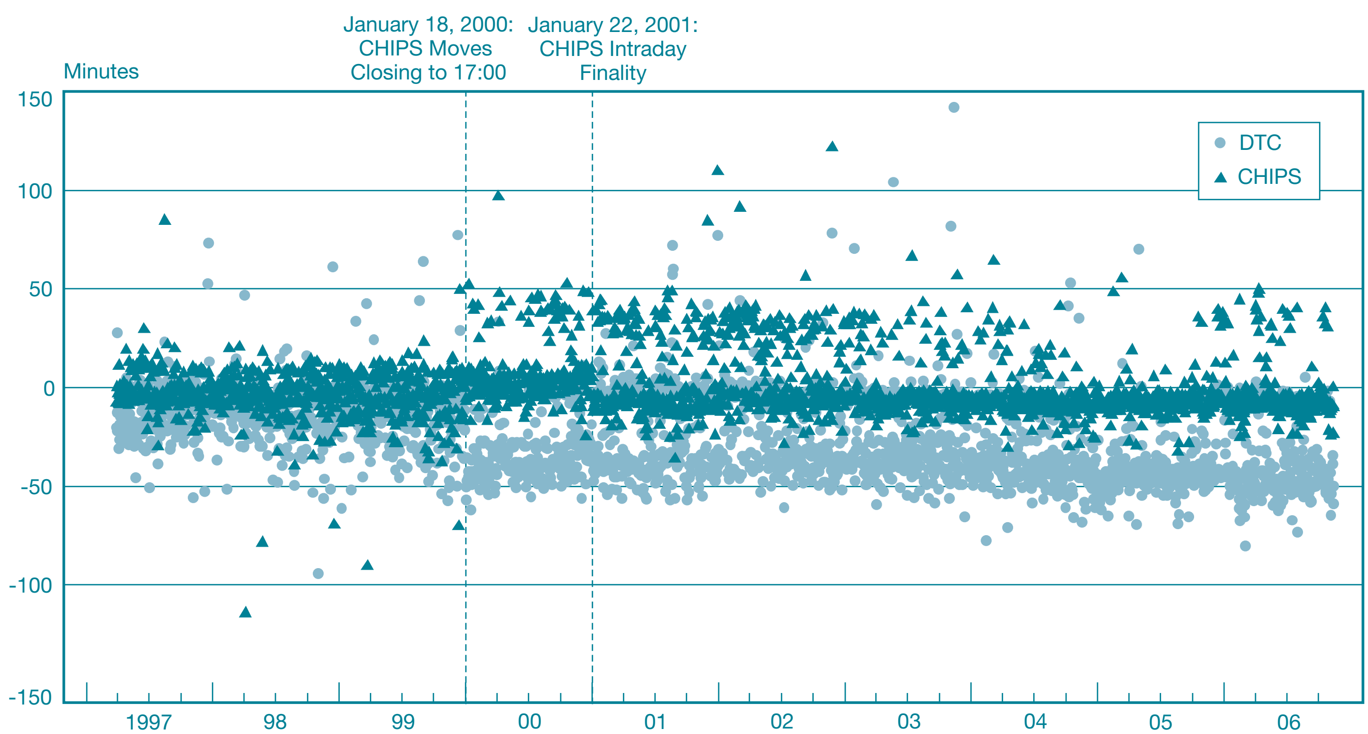
The resolution-of-uncertainty hypothesis is a third reason why banks might release payments after a major settlement system's final payouts. Prior to settling at a major settlement institution, banks could be uncertain of the exact amount of their payout from the system (or their obligation to the system), and there could be some uncertainty about whether all parties in the settlement system will perform as expected. After the uncertainty is resolved, banks might evaluate more precisely the effect of releasing payments, arranging for the purchase or sale of federal funds, and making any other adjustments.

Finally, banks, according to the focal point hypothesis, might coordinate their payment submissions with a settlement institution. Banks might choose to release payments after the final payouts of a major settlement institution because they believe that other banks will do so at the same time. If many banks choose to release payments at the same time, each bank has a higher likelihood of receiving payments during that peak of payment activity; therefore, it is more likely that the banks will have a lower cost of funding their outgoing payments at that time (by incurring fewer daylight overdrafts and avoiding any constraints, such as bumping up against a net debit cap). This same phenomenon could hold true for depositors—a hypothesis discussed in McAndrews and Rajan (2000).

While our analysis cannot clearly distinguish between the alternative hypotheses, it is instructive to consider what the data might imply about the relative weight of the various hypotheses as explanations for the changes observed after January 18, 2000. Chart 15 shows the difference in the settlement times of CHIPS and DTC and the time of the top ten minutes of Fedwire value transferred. A band of five minutes on either side of 0 on the y-axis represents the time of settlement of the top ten minutes of Fedwire value. A circle represents the time of DTC settlement and a triangle the time

Chart 15

Difference between Peak Ten Minutes and Settlement Times of CHIPS and DTC



Sources: Federal Reserve Bank of New York; authors' calculations.

Note: Values exclude payments associated with CHIPS, CLS Bank, DTC, and principal and interest payment funding.

of CHIPS settlement. When a point falls within the five-minute band on either side of 0, the settlements of Fedwire and that system coincide. The clearest pattern that Chart 15 displays is that prior to January 18, 2000, the three systems—CHIPS, DTC, and Fedwire—settled roughly simultaneously; after that date, CHIPS and DTC settlement times diverged. The time of Fedwire peak settlement tends to bounce between the time of DTC settlement and a time immediately subsequent to CHIPS settlement. This is because, prior to January 18, 2000, it was fairly common to have all three systems settle roughly simultaneously, so it is difficult to distinguish among the various hypotheses. These same patterns are visible in Chart 13.

After January 18, 2000, we see two tendencies. First, on days on which the 16:30 peak is the time of the highest value transferred on Fedwire, DTC settlement often falls within the ten minutes of highest value transferred on Fedwire. Second, when the 17:11 peak is the time of highest value transferred on Fedwire, it is usually (and increasingly over time) the case that CHIPS settlement precedes and falls outside the highest ten minutes of Fedwire value transferred.

CHIPS is the last major institution to settle; at that point in the processing day, most uncertainty about bank balances is expected to be resolved and banks might reason that it is a good time to send payments after CHIPS has settled.

These tendencies might suggest that DTC settlement kicks off a liquidity and customer credit cascade, an observation supported by the quickness of value transfer on Fedwire following DTC settlement. This observation is also consistent with the fact that DTC's account balances grow over the day and are then released with DTC settlement to other banks, which effectively increases the short-term supply of liquidity in the rest of the banking system. However, the fact that the peak of activity follows CHIPS settlement with only some delay might suggest that the cause of the activity is more closely associated with resolution of uncertainty than with a liquidity or credit cascade. This explanation is also consistent with the fact that CHIPS settlement is mainly a redistribution of balances among banks, rather than a net release of liquidity back to banks. The focal point hypothesis might also offer a good explanation for the peak that occurs after CHIPS settlement. CHIPS is the last major institution to settle; at that point in the processing day, most uncertainty about bank balances is expected to be resolved and banks might reason that it is a good time to send payments after CHIPS has settled.

6. Conclusion

In its report on large-value payments systems, the Committee on Payment and Settlement Systems noted that the timing of payment submission is a market practice of major importance (Bank for International Settlements 2005). It is not uncommon for stable behavioral conventions to arise around the time when participants submit certain types of payments for settlement. Such conventions can arise endogenously among direct participants, and non-RTGS payments can be an important "exogenous" factor affecting a bank's RTGS liquidity.

Our examination of payment time distribution trends in the Fedwire funds service from 1998 to 2006 finds that the lower percentiles of timing have moved to earlier in the operating day. In addition, the very last percentiles transferred have also moved to earlier over the study period, as extensions of the Fedwire operating day have become more rare. We also observe that while more value settles earlier in the morning and the later percentiles of value settle later in the afternoon, the distribution of value transferred has become more peaked. A greater percentage of value was transferred during the peak of activity in 2006 than in the early part of our sample.

We considered a host of factors affecting changes in the distribution of Fedwire timing. Federal Reserve policies and operations were found to affect settlement times in notable ways. Changes to the Federal Reserve's Payments System Risk Policy in 2002, which expanded net debit caps significantly, quickened Fedwire settlement in the morning and early afternoon, while 2006 policy changes, which lowered extensions of daylight credit overall, tended to slow settlement throughout the afternoon. Both changes are consistent with banks economizing on liquidity costs in their submission of payments. Changes in the values and volumes transferred over Fedwire have increased over the period, as has the proportion of large individual payments. Taken together, these changes explain a large share of the later settlement of payments (after 17:00) over the period. However, we find that a larger value of customer payments tends to quicken Fedwire settlement, possibly because of the higher delay costs of customer payments—another influence noted in the economic literature.

There were numerous changes in settlement institutions over the period. The introduction of CLS Bank operation and the values transferred by CLS Bank appear to have had little effect on the value time distribution on Fedwire. Changes in CHIPS operations had countervailing effects, with a 2000 move toward later settlement time by CHIPS clearly contributing to a later peak in Fedwire activity. The move by CHIPS to intraday finality in 2001, though, tended to speed settlement of the 40th-80th percentiles of Fedwire value. Increased values transferred by CHIPS tended to delay settlement on Fedwire for the 50th-80th percentiles, while on DTC heightened values quickened early settlement and delayed mid-afternoon settlement. A payment market that has grown more concentrated had a significant influence on the later settlement of Fedwire value. These results reflect the various calendar effects and activity in other financial markets included in our regression analysis.

A major contributor to the later settlement of Fedwire payments was the change in CHIPS settlement time from roughly 16:45 to 17:10 on January 18, 2000. The time of the midpoint of the highest ten minutes of Fedwire value transferred moved in a remarkably coincident fashion from 16:48 to 17:11 on that date. This pattern persisted, with the Fedwire value time distribution displaying two main peaks: one remaining at 16:30, nearly coincident with DTC settlement, and one at 17:11, shortly after the time of CHIPS final payouts. Over time, these new peaks of activity have been stable.

Our results also suggest that changes in the value of interbank payments and in the number of daily Fedwire transfers can explain most of the daily variation in the time of value transferred on Fedwire after 17:00. The rare case of an extension of CHIPS operating hours delays virtually all payments submitted after 17:00, a clear illustration of the interdependence between Fedwire and CHIPS.

In addition, we estimate that increases in the number and value of Fedwire payments between 1998 and 2006 contributed slightly more than 40 percent to the long-run change in the 75th percentile of the value time distribution of payments. Of this amount, the aggregate contribution of CHIPS may be estimated at around 10 percent, and the increase in industry concentration accounted for close to 30 percent.

The clear interdependence between payment timing on Fedwire and CHIPS reinforces the points made in the Committee on Payment and Settlement Systems' report on large-value payments systems (Bank for International Settlements 2005). The role of settlement institutions that utilize Fedwire for pay-ins and payouts is a major factor determining system activity. This article has reviewed a number of hypotheses on the possible channels through which settlement systems affect Fedwire activity. Further research on these channels would indeed provide a better understanding of the factors that affect the timing distribution of payments in large-value systems.

Also deserving of further research is the effect of increased industry concentration on the timing of payments, a topic that has not been explored in the theoretical literature. Furthermore, researchers could benefit from conducting similar studies of other payments systems to ascertain the effects of daylight credit policies, system operations, settlement systems, industrial structure, and other determinants of payment timing. Their results could shed light on the robustness of our results.

Appendix A: Data

Our data source is Federal Reserve Bank of New York records of every Fedwire funds service transaction. Unless otherwise stated, data are used to construct the variables below associated with Fedwire funds activity. We have data on all Fedwire funds transfers between April 1997 and December 2006, except for April 1, 1997; December 22-24, 1997; March 1-3, 1999; April 20-22, 1999; October 14-15, 1999; October 18, 1999; and November 9, 1999.

Variables

ith percentile of value time is the time at which *i* percent of the total daily value has settled. We exclude payments to or from CHIPS, CLS Bank, and DTC. We also exclude payments associated with interest and redemption payments of government-sponsored enterprises and international institutions after the Federal Reserve's Payments System Risk Policy change on July 1, 2006. These payments related to P&I (principal and interest) are Fedwire funds payments between two different accounts of the securities issuer, that is, payments from the general account to the funding account and from the funding account to the distribution account.

Foreign Capital Equivalency Policy is a binary variable equal to 1 on and after February 21, 2002, when the Federal Reserve changed the criteria for determining U.S. capital equivalency for foreign banks. This policy change increased the sum of the net debit caps of all Fedwire funds participants by $123 billion, or 12 percent (see Board of Governors of the Federal Reserve System [2001]).

GSE credit policy is a binary variable equal to 1 on and after July 1, 2006. The Federal Reserve changed its Payments System Risk Policy to require GSEs and international organizations to fully fund interest and redemption payments on securities before the funds are sent, and it removed the provision of free intraday credit to these issuers (Board of Governors of the Federal Reserve System 2004; McAndrews 2006).

MBS P&I day, pre-GSE policy is a binary variable equal to 1 on the 15th and 25th of the month, or the first business day thereafter, before the change in GSE credit policy on July 1, 2006. On these days, Fannie Mae and Freddie Mac make interest and redemption payments on mortgage-backed securities (MBSs). These are generally the largest interest and redemption payment days of the month.

MBS P&I day, post-GSE policy is a binary variable equal to 1 on the 15th and 25th of the month, or the first business day thereafter, after the change in GSE credit policy on July 1, 2006.

Fedwire opens at 21:00 is a binary variable equal to 1 for all days on or after May 17, 2005. On that date, the Federal Reserve extended the operating hours of the Fedwire funds service from 18 hours to 21.5 hours by moving the opening time from 00:30 to 21:00 (Board of Governors of the Federal Reserve System 2003).

Operating hour extension is the number of minutes that the Fedwire funds service remains open after 18:30. The Federal Reserve will occasionally extend Fedwire's operating hours at the request of a participant having operational difficulties or if the system is experiencing operational problems (Bank for International Settlements 2005).

Fed funds target rate — Source: <http://www.ny.frb.org/markets/omo/dmm/fedfunds.cfm>.

Interbank payment value is the sum of the payment values of all Fedwire funds transfers that are not fed funds deliveries, fed funds returns, customer payments, or settlement payments for CHIPS, CLS Bank, or DTC, or that are not principal and interest redemptions.

Customer payment value is the sum of the payment values of all Fedwire funds transfers with a business function code of customer payment.

Fed funds deliveries is the total value of new fed funds loans. These loans were identified from Fedwire funds transactions, as in Furfine (1999).

Fed funds returns is the total value of the returns of the fed funds loans. It is equal to the value of fed funds deliveries for the previous business day plus the interest on those loans. These loans were identified from Fedwire funds transactions, as in Furfine (1999).

Payments ≥ $10 mn. is the fraction of daily value from payments greater than or equal to $10 million. This excludes all CHIPS, CLS Bank, DTC, and P&I funding payments. The threshold value of $10 million is the value used in a survey of bank intraday liquidity management conducted by the Payments Risk Committee and the Wholesale Customer Advisory Group (2007).

Number of payments is the daily number of Fedwire funds payments, including interbank, customer, and fed funds transactions, but excluding all CHIPS, CLS Bank, DTC, and P&I funding payments.

CHIPS settlement at 17:00 — CHIPS settlement time is a binary variable equal to 1 for all days on or after January 18, 2000. On that date, the time at which end-of-day CHIPS payouts occurred moved from approximately 16:45 to 17:10.

CHIPS intraday finality is a binary variable set to 1 for all dates on or after January 22, 2001. This is the date when CHIPS moved from an end-of-day multilateral net debit system to a mixed-payments system with intraday finality.

CHIPS final payout value is the value of the end-of-day payouts sent by CHIPS over Fedwire to CHIPS participants with a net credit position.

CHIPS extension is a binary variable for a later-than-normal CHIPS final payout time. This is defined as a CHIPS final payout occurring after 17:00 for days before January 18, 2000, and after 17:15 otherwise.

DTC settlement time is the value-weighted mean time of Fedwire funds payments sent by DTC after 16:00.

DTC net-net credit value is the sum of all Fedwire funds payments sent by DTC after 16:00.

CLS Bank opens is a binary variable equal to 1 for all days on or after September 10, 2002, when CLS Bank International began settling U.S. dollar transactions.

CLS Bank USD value is the daily sum of payments sent by CLS Bank over Fedwire. It is equivalent to the value of all U.S. dollar legs settled by CLS Bank.

Sep. 11-18, 2001, is a binary variable equal to 1 for those dates. This is the period in which the Fedwire payments system was disrupted by the terrorist attacks on September 11 (McAndrews and Potter 2002).

NYSE closures and NYSE early closures — Source: <http://www.nyse.com/pdfs/closings.pdf>.

Reserve maintenance cycle days are binary variables for the days in a reserve maintenance cycle. The maintenance cycle is a two-week period starting on a Thursday (see Federal Reserve Banks [2006] for the starting and ending dates of maintenance cycles). We include dummies for all days of the week with Thursdays—the first day of the reserve maintenance cycle—as the excluded group. To disentangle the effect of the maintenance cycle above from any day-of-week effects, we include binary variables for maintenance days in the second week of the maintenance cycles, that is, days 6-10.

HHI of Fedwire value is the Herfindahl-Hirschmann Index of the value of Fedwire funds payments sent by master accounts.

Fed funds deviation is the difference between the effective fed funds rate and the target fed funds rate. Source: <http://www.ny.frb.org/markets/omo/dmm/fedfunds.cfm>.

CHIPS

CHIPS is a private, large-value U.S. dollar payments system owned and operated by the Clearing House Payments Company (Federal Reserve Bank of New York 2002; Bank for International Settlements 2003b, 2005). As of April 2007, CHIPS had 45 members and settled 329,000 transactions valued at $1.7 trillion per day.[11] From its opening in 1970 until 2001, CHIPS operated as an end-of-day multilateral net debit settlement system: After CHIPS closed at 04:30 (05:00 after January 18, 2000), participants with negative net positions would send payments to CHIPS over Fedwire to cover their positions; CHIPS would then send payments to those participants with net positive positions.

On January 22, 2001, CHIPS adopted intraday payment finality with a continuous offsetting algorithm to optimize liquidity. All CHIPS participants must fund their accounts with a Fedwire transfer to CHIPS between the opening of Fedwire and 09:00 before they can send or receive payments. These balances, totaling about $3 billion, are used to settle payments during CHIPS operating hours. At the close of CHIPS at 17:00, any unsettled payments are multilaterally netted. These net positions are settled over Fedwire via transfers to and from CHIPS.

CLS Bank

CLS Bank is a payment-versus-payment settlement system that settles foreign exchange transactions in fifteen currencies (CLS Bank 2007; Miller and Northcott 2002; Bank for International Settlements 2003a, 2005). CLS Bank is operated by CLS Bank International, a bank-owned Edge Act corporation incorporated in the United States. CLS Bank was founded in response to concerns raised by the G-10 central banks about settlement risk in foreign exchange transactions (Bank for International Settlements 1993). CLS Bank began operation in September 2002; as of December 2006, it had 57 members and settled an average of 290,000 transactions valued at $3.3 trillion per day.[12]

CLS Bank uses a payment-versus-payment method in which funds to settle trades are exchanged simultaneously in different currencies. In order to accomplish simultaneous transfers, CLS Bank is open during the five-hour settlement window—01:00 to 06:00 EST—when real-time gross settlement systems in Europe, the Americas, and Asia are open.

DTC

DTC is a securities settlement system that settles the majority of U.S. corporate securities and commercial paper transactions. It is a wholly owned subsidiary of Depository Trust & Clearing Corporation (Bank for International Settlements 2003a, 2005). DTC has 407 participants and 86 settling banks. On average, it settles 800,000 transactions valued at $896 billion per day (Payments Risk Committee and Wholesale Customer Advisory Group 2007).

DTC participants fund their accounts through Fedwire transfers (via a settlement bank for many) to the DTC Federal Reserve account. Money market instruments represent 62 percent of DTC value. The ability of paying agents to accept maturing securities is limited by the agents' net debit cap. To remove the debit cap constraint, agents will make progress payments to their accounts via Fedwire transfers to DTC. The majority of this activity occurs between 12:00 and 14:00. At 16:00, the DTC settlement process begins. Banks with net debits send the net amount to DTC over the net settlement system at 16:35. At 16:40, DTC sends Fedwire funds transfers to participants with net credits (Payments Risk Committee and Wholesale Customer Advisory Group 2007).

[11] Source: CHIPS (<http://www.chips.org/about/pages/001221.php>).

[12] Source: CLS Bank International (<http://www.cls-group.com/news/article.cfm?objectid=78EA8ED8-EC63-6345-C60967F0ECA7E5C3>).

References

Angelini, P. 1998. "An Analysis of Competitive Externalities in Gross Settlement Systems." Journal of Banking and Finance 22, no. 1 (January): 1-18.

———. 2000. "Are Banks Risk Averse? Intraday Timing of Operations in the Interbank Market." Journal of Money, Credit, and Banking 32, no. 1 (February): 54-73.

Bank for International Settlements. 1993. "Central Bank Payment and Settlement Services with Respect to Cross-Border and Multi-Currency Transactions." Committee on Payment and Settlement Systems, CPSS Publication no. 7, September. Basel, Switzerland.

———. 2003a. "A Glossary of Terms Used in Payment and Settlement Systems." Committee on Payments and Settlement Systems. March. Basel, Switzerland.

———. 2003b. "Statistics on Payment and Settlement Systems in Selected Countries—Figures for 2001." Committee on Payment and Settlement Systems, CPSS Publication no. 54, April. Basel, Switzerland.

———. 2005. "New Developments in Large-Value Payment Systems." Committee on Payment and Settlement Systems, Publication no. 67, May. Basel, Switzerland.

Bech, M. L. 2008. "Intraday Liquidity Management: A Tale of Games Banks Play." Federal Reserve Bank of New York Economic Policy Review 14, no. 2 (September): 7-24.

Bech, M. L., and R. Garratt. 2003. "The Intraday Liquidity Management Game." Journal of Economic Theory 109, no. 2 (April): 198-210.

Becher, C., M. Galbiati, and M. Tudela. 2008. "The Timing and Funding of CHAPS Sterling Payments." Federal Reserve Bank of New York Economic Policy Review 14, no. 2 (September): 113-134

Beyeler, W., R. Glass, M. Bech, and K. Soramäki. 2006. "Congestion and Cascades in Payment Systems." Federal Reserve Bank of New York Staff Reports, no. 259, September.

Board of Governors of the Federal Reserve System. 2001. "Policy Statement on Payments System Risk." Federal Register 66, no. 240: 64419-33. Docket nos. R-1107, R-1108, R-1109, and R-1110.

———. 2003. "Expansion of the Operating Hours for the On-Line Fedwire Funds Service." Federal Register 68, no. 101: 28826-8.

———. 2004. "Policy Statement on Payments System Risk." Federal Register 69, no. 187: 57917-31. Docket no. OP-1182. Available at <http://www.federalreserve.gov/BoardDocs/Press/other/2004/20040923/attachment.pdf>.

———. 2006. "Consultation Paper on Intraday Liquidity Management and Payments System Risk Policy." Federal Register 71, no. 119: 35679-87. Docket no. OP-1257.

CLS Bank. 2007. "CLS Overview."

Federal Reserve Bank of New York. 2002. "Fedpoint: CHIPS." April. Available at <http://www.newyorkfed.org/aboutthefed/fedpoint/fed36.html>.

Federal Reserve Banks. 2006. "Funds Transfers through the Fedwire Funds Service." Federal Reserve Financial Services, Operating Circular no. 6, September. Available at <http://www.frbservices.org/OperatingCirculars/pdf/Oc6.pdf>.

Federal Reserve System. 2006. "Reserve Maintenance Manual." November. Available at <http://www.frbservices.org/Accounting/pdf/rmm.pdf>.

Furfine, C. H. 1999. "The Microstructure of the Federal Funds Market." Financial Markets, Institutions, and Instruments 8, no. 5 (December): 24-44.

Green, E. J. 2005. "The Role of the Central Bank in Payment Systems." Paper presented at the conference "The Future of Payments," Bank of England, London, England, May 19-20.

Heller, D., T. Nellen, and A. Sturm. 2000. "The Swiss Interbank Clearing System." Swiss National Bank, Payment System Subsection, June. Available at <http://www.sic.ch/dl_tkicch_sicartikelsnb.pdf>.

Kahn, C. M., J. McAndrews, and W. Roberds. 2003. "Settlement Risk under Gross and Net Settlement." Journal of Money, Credit, and Banking 35, no. 4 (August): 591-608.

References (Continued)

Koponen, R., and K. Soramäki. 2005. "Intraday Liquidity Needs in a Modern Interbank Payment System—A Simulation Approach." In H. Leinonen, ed., Liquidity, Risks and Speed in Payment and Settlement Systems—A Simulation Approach. No. E: 31 in Bank of Finland Studies, chapter 3, 73-116.

Leinonen, H., and K. Soramäki. 2005. "Optimizing Liquidity Usage and Settlement Speed in Payment Systems." In H. Leinonen, ed., Liquidity, Risks and Speed in Payment and Settlement Systems—A Simulation Approach. No. E: 31 in Bank of Finland Studies, chapter 4, 117-50.

McAndrews, J. J. 2006. "Alternative Arrangements for the Distribution of Intraday Liquidity." Federal Reserve Bank of New York Current Issues in Economics and Finance 12, no. 3 (April).

McAndrews, J. J., and S. Potter. 2002. "Liquidity Effects of the Events of September 11, 2001." Federal Reserve Bank of New York Economic Policy Review 8, no. 2 (November): 59-79.

McAndrews, J., and S. Rajan. 2000. "The Timing and Funding of Fedwire Funds Transfers." Federal Reserve Bank of New York Economic Policy Review 6, no. 2 (July): 17-32.

Miller, P., and C. A. Northcott. 2002. "CLS Bank: Managing Foreign Exchange Settlement Risk." Bank of Canada Review, autumn: 13-25.

Mills, D. C., and T. D. Nesmith. 2008. "Risk and Concentration in Payment and Securities Settlement Systems." Journal of Monetary Economics 55, no. 3, (April): 542-53.

Newey, W. K., and K. D. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." Econometrica 55, no. 3 (May): 703-8.

Payments Risk Committee and Wholesale Customer Advisory Group. 2007. "Report of the Joint Task Force on the PSR Policy Consultation Paper." February.

Christopher Becher, Marco Galbiati, and Merxe Tudela

The Timing and Funding of CHAPS Sterling Payments

- Participants in CHAPS Sterling often use incoming funds to make payments, a process known as liquidity recycling.
- Liquidity recycling can be problematic if participants delay their outgoing payments in anticipation of incoming funds.
- An analysis of CHAPS payment activity shows that the level of liquidity recycling, though high, is stable throughout the day—a condition attributable to three features of the system.
- First, the settlement of time-critical payments in CHAPS supplies liquidity early in the day—liquidity that can be recycled to fund less urgent payments.
- Second, CHAPS throughput guidelines provide a centralised coordination mechanism that essentially limits any tendency toward payment delay.
- Third, the relatively small direct membership of CHAPS facilitates coordination, enabling members to maintain a constant flux of payments during the day.

1. Introduction

The use of real-time gross settlement (RTGS) systems for the settlement of large-value payments offers considerable advantages, the principal one being the elimination of the credit risk that can arise between participants in deferred net settlement systems. However, in comparison with deferred net settlement systems, RTGS systems require relatively large amounts of liquidity to support payment activity. This liquidity can be sourced from the settlement agent (usually a central bank in the case of large-value payments systems)—in the form of intraday overdrafts—or from incoming payments from other participants.

Obtaining intraday liquidity from a central bank is typically costly. In order to minimise this cost and to take advantage of incoming payments as a funding source, participants may choose to delay outgoing payments. However, payment delay may itself prove costly. Participants face a trade-off, therefore, between the cost of borrowing from the central bank and the expected cost of delaying payments. McAndrews and Rajan (2000) explore this trade-off in a study of payment behaviour in Fedwire. They describe how the use of incoming funds to offset outgoing payments allows participants to avoid incurring costly overdrafts from the central bank and hence reduces the liquidity cost of making payments. Such offsetting can be achieved to a greater extent during activity peaks, so banks are induced to coordinate their payments around, and thereby to reinforce, these peaks.

Christopher Becher, a policy officer at the European Commission, was an analyst at the Bank of England at the time this article was written; Marco Galbiati is an economist and Merxe Tudela a senior economist at the Bank of England.

Correspondence: <marco.galbiati@bankofengland.co.uk>

The authors thank Mark Manning, James McAndrews, participants at workshops at the Bank of England, and an anonymous referee for invaluable input. The views expressed are those of the authors and do not necessarily reflect the position of the Bank of England, the European Commission, the Federal Reserve Bank of New York, or the Federal Reserve System.

This article investigates the factors influencing the timing and funding of payments in the CHAPS Sterling system, drawing where appropriate on comparisons with payment activity in Fedwire. In the next section, we discuss theoretical approaches to the study of payment behaviour and their application to CHAPS Sterling. The empirical analysis of the timing and funding of CHAPS Sterling payments follows in Sections 3 and 4, respectively. Section 5 concludes.

2. Theoretical Studies of Payment Behaviour

Several theoretical studies have addressed the incentives facing participants in RTGS systems. Many focus on the aforementioned trade-off between the cost of liquidity and the expected cost of delaying payments.

2.1 Definition of Terms

The measurement of the *cost of liquidity* varies according to the regime employed by the settlement agent (in the cases described in this article, that agent is the central bank). When credit is supplied unsecured, the cost typically takes the form of an explicit overdraft fee. Credit may also be provided against eligible collateral, in which case the cost to the participant is the opportunity cost of posting eligible securities with the central bank and hence forgoing alternative uses for those assets.

The *cost of delay* may take several forms. Financial penalties may be incurred for failure to make time-critical payments by specified deadlines, such as for settlement payments in ancillary systems or repayments of interbank loans. In addition, failure to make customer payments on time, or indeed at all, on the intended settlement date may result in reputational costs and a loss of future business. Also, as we discuss later, the reputation of a participant within a payments system may suffer if it is perceived to be delaying payments in order to "free-ride" on liquidity provided by others.

2.2 Theoretical Approaches

Bech and Garratt (2003) model the trade-off using a game-theoretical approach, analysing the behaviour of two banks, both of which receive random payment requests from customers at the beginning of a morning and an afternoon period. Both banks face a fixed cost of delaying payments and of posting collateral for a morning or afternoon period. The analysis is repeated under priced and collateralised intraday liquidity regimes, as employed in Fedwire and CHAPS, respectively.[1]

Under a collateralised regime, Bech and Garratt find that both early and delayed payments are possible equilibria, depending on the relative costs of liquidity and delay. The efficient equilibrium is for both banks to pay early. However, for certain levels of delay and liquidity costs, the participants are found to be in a prisoner's dilemma, in which the dominant strategy for both is to delay payments until the afternoon, even though both would benefit if payments were made in the morning. This incentive to delay arises because it is possible to avoid the cost of posting collateral in the morning and instead to incur the (cheaper) delay cost. In these cases, there would be a welfare improvement if the participants could be induced to coordinate and to pay earlier.

This article investigates the factors influencing the timing and funding of payments in the CHAPS Sterling system, drawing where appropriate on comparisons with payment activity in Fedwire.

Under a regime of priced credit, Bech and Garratt again find that multiple equilibria are possible. However, in this case, participants stand to benefit from synchronising payments with each other, since no cost is incurred by either participant if payments are "offset" within the time period over which overdraft fees are calculated. The equilibrium outcome will thus depend not only on the relative costs of liquidity and delay but also on the likelihood that the other bank will receive a payment request. In the specific case where the expected cost of delay is lower than the credit fee, and payment flows are skewed toward the afternoon, Bech and Garratt find that the efficient equilibrium involves delay until the afternoon.

In a similar study, Kobayakawa (1997) models the choice of whether to delay payments in RTGS systems under varying intraday credit arrangements. Again, the relative costs of liquidity and delay drive equilibrium selection. Under a system of priced credit, Kobayakawa (like Bech and Garratt) finds that delayed settlement is an equilibrium, since each participant seeks to avoid incurring an overdraft by delaying payments and thereby "free-riding" on the liquidity provided by the other participant. Under a collateralised regime, Kobayakawa finds a unique equilibrium in which both participants pay early.

[1] Bech and Garratt also examine the case of free intraday credit; however, those results are not discussed here.

However, in this case, the result is obtained by assuming that the opportunity cost of collateral is a sunk cost, and so liquidity is in effect free when the game is played. This is unsatisfactory, since it takes no account of the participants' incentives to reduce the cost of liquidity by economising on the value of collateral posted. Consequently, we focus in the following discussion on the Bech and Garratt model.

Mills and Nesmith (2008) adapt the Bech and Garratt model to look at the effect of settlement risk on timing decisions in payments and securities settlement systems, concentrating on the differential impact of overdraft costs on the two types of systems. A main contribution of this paper is to describe a rationale for delays overlooked in the literature: namely, that banks may withhold payments until they receive information on the others' ability to send funds, in order to obtain a better forecast of the costs of funding their own payments. More precisely, in the model, banks choose between paying "early" and paying "late." There are no delay costs, so, in the absence of settlement risk, "early" is a *weakly* dominated strategy,[2] which may still appear in equilibrium (although only in risk-dominated ones). Introduction of settlement risk definitely tilts the balance against the "early" strategy, because in this case "early" becomes a *strictly* dominated strategy ("late" outperforms it against *any* action by the opponent, as settlement risk imposes some overnight overdraft in probability terms). Thus, settlement risk introduces further reasons to delay, eliminating the "early payment" equilibria.

The overarching conclusion of these [theoretical] works is that institutional features and, in particular, intraday credit regimes have a powerful effect on banks' incentives; as a consequence, they largely determine a payments system's performance.

Building on previous work, Martin and McAndrews (2008) construct a model with a continuum of banks, each making a unit payment to one other bank and each having to decide whether to pay "early" or "late." Banks are assumed to face random delay and liquidity costs, determined in turn by bank-specific shocks that drain (or increase) the available liquidity. The paper shows that, depending on the cost parameters (costs of delay and of overdrafts), on the time-criticality of payments, and on the probability and size of liquidity shocks, the resulting equilibria feature different degrees of delay. More specifically, some or all banks, depending on the shocks received, decide to delay their payments. Martin and McAndrews also explore the effect of a liquidity-saving mechanism on the banks' incentives. This is shown to mitigate the strategic complementarity of banks' strategies by allowing banks to release payments conditional on the receipt of payments. The paper shows that, in this case, the extent of delay in equilibrium also depends on the pattern of payments (whether payments can be offset in pairs or multilaterally), which has implications for the system's efficiency.

The overarching conclusion of these works is that institutional features and, in particular, intraday credit regimes have a powerful effect on banks' incentives; as a consequence, they largely determine a payments system's performance. We draw on—and, where necessary, modify—these theoretical predictions to study payment behaviour in Fedwire and in CHAPS Sterling.

2.3 Implications for Payment Behaviour in Fedwire

The Federal Reserve System supplies intraday liquidity to Fedwire members in the form of uncollateralised daylight overdrafts. Subject to net debit caps, participants can incur overdrafts at any time, which incur a charge calculated as the average per-minute overdraft during the day, multiplied by an effective daily rate, less a deductible.[3]

As Bech and Garratt (2003) describe, a corollary of this charging structure is that participants can avoid overdraft charges by synchronising payments. As long as incoming payments of at least equivalent value to outgoing payments are received within a minute, no overdraft will be required and hence no charge incurred. This sets the scene for a pure coordination game, in which participants attempt to synchronise payments in order to minimise the average overdraft position over the course of the day. The theory would also predict that if the overdraft fee is deemed to be high relative to the expected cost of delay, the efficient equilibrium will involve the delay of payments until later in the day.

This theoretical finding is consistent with the empirical results obtained by McAndrews and Rajan (2000) on the timing and funding of payments in Fedwire. Faced with costly intraday liquidity, participants appear to delay payments until an end-of-day activity peak, during which the probability of

[2] "Early" performs no better than "late," and it performs just as well as "late" if the other also pays early, as offsetting payments incur no charge.

[3] The effective rate is currently equivalent to an annual rate of 36 basis points.

receiving funds from other participants is greater. It is argued that this synchronised delay reinforces the activity peak. The "focal points" for this coordination appear to be provided by ancillary system settlement deadlines (in particular, in CHIPS and DTC). As McAndrews and Rajan note, though, the outcome of this apparent coordination may not be socially efficient, since all participants might stand to benefit from reduced liquidity costs if coordination could be improved so as to take full account of liquidity externalities.

Armantier, Arnold, and McAndrews (2008) extend this analysis in their study of recent changes in the timing of Fedwire funds transfers. Among other things, they explore alternative hypotheses for the timing of late-afternoon payment peaks, with particular reference to the change in the timing of late-afternoon Fedwire transfers following a move to a later CHIPS settlement time. The tendency for Fedwire transfers to be made after ancillary system positions are settled may reflect the "focal point" hypothesis described above. However, it may also be that the liquidity released by ancillary system settlement may trigger a "cascade" of payments, to the extent that participants are liquidity-constrained.[4] Along similar lines, the settlement may also release credit lines, thereby permitting more payments to be made. Additionally, the authors suggest that uncertainty surrounding the size of ancillary system payouts may lead to payments being deferred until after the settlement deadline—that is, once uncertainty has been resolved. The data do not allow for a clear distinction to be made between the competing hypotheses; however, there is sufficient evidence to suggest that the coordination described in the earlier paper is only part of the story.

2.4 Implications for Payment Behaviour in CHAPS Sterling

The Bank of England provides intraday liquidity to members of CHAPS Sterling in the form of interest-free overdrafts secured against eligible collateral. The maximum value of liquidity granted is equal to the value of collateral securities posted, less a "haircut" to take account of movements in the value of the collateral securities. In contrast with Fedwire, where the total cost of liquidity is driven by the *average* overdraft incurred, the cost of liquidity in CHAPS Sterling is driven by the *maximum* overdraft position incurred during the day, since the value of collateral posted must be at least equal to this position.

The cost of posting collateral derives from the fact that the securities posted (or the funds used to obtain the required securities) could be used for alternative purposes; participants therefore face an opportunity cost. As described in Box 1, the upper bound to this cost has been estimated to be of the order of 7 basis points per annum, although for domestic banks subject to the Stock Liquidity Regime the opportunity cost may be significantly lower and may even approach zero.

The Bech and Garratt (2003) model predicts that this regime will result in multiple equilibria, with the selection of an equilibrium dependent on the relative magnitudes of the cost of delayed payment and the opportunity cost of posting collateral. The low opportunity cost of posting collateral for many CHAPS Sterling members may thus be expected to favour an early rather than a delayed payment equilibrium. That said, it is difficult to quantify the cost of delay associated with all but a small number of time-critical payments. Anecdotal evidence suggests that costs of delay are low for the majority of payments.

In contrast with Fedwire, where the total cost of liquidity is driven by the average *overdraft incurred, the cost of liquidity in CHAPS Sterling is driven by the* maximum *overdraft position incurred during the day, since the value of collateral posted must be at least equal to this position.*

Certain qualifications are required in applying this model to CHAPS Sterling. In particular, in the Bech and Garratt model, the benefit from delaying payments in a collateralised regime derives from the assumption that it is less costly to post collateral for the afternoon than for the whole day (and hence that there is an incentive to avoid posting collateral in the morning). This in turn rests on the assumption that it is possible to invest surplus liquidity for a fraction of the day—or, in other words, that there exists an intraday market for liquidity. It is not obvious that this incentive applies in CHAPS Sterling since, in the absence of an intraday market, it is probably no cheaper to post collateral for a morning or afternoon than for a full day. Once collateral is committed to the payments system, the cost for the full day is incurred.[5]

It is possible to modify the Bech and Garratt model to incorporate an incentive to delay that does not rely on the existence of an intraday market for liquidity. By delaying

[4] See also Beyeler et al. (2006).

[5] It is nonetheless possible for banks to withdraw liquidity intraday. As we argue here, while it may not be possible to lend in an intraday interbank market, the collateral could in principle be committed to another payments system. In this case, the ability to commit collateral for only part of a day could be considered valuable.

payment and taking advantage of incoming funds, a participant may be able to reduce the maximum overdraft position and hence reduce the aggregate collateral requirement for the day (or avoid posting collateral altogether). It can be shown that a similar prisoner's dilemma outcome emerges from this model, with both participants defecting despite the mutual benefit from paying early, unless they can somehow be induced to coordinate earlier in the day.[6]

However, the finding that some participants may seek to reduce their aggregate collateral requirements by delaying payments must be seen in the context of the empirical observations that participants in CHAPS typically post collateral at the beginning of the day (that is, they do not generally wait to determine whether collateral posting is required) and many post collateral to a value well in excess of liquidity usage (for the system as a whole, maximum liquidity used is only around one-third of the maximum collateral posted).[7] Two factors appear to be particularly influential in explaining this behaviour.[8]

First, as discussed in Box 1, the low opportunity cost of posting collateral (or of maintaining positive reserve account balances) means that, for many banks there appears to be little incentive to delay posting collateral until later in the day, since the potential savings to be made from reducing the aggregate value of collateral posted are small.

Second, the distribution of "time critical" payments, for which the expected cost of delay is high, appears to be skewed toward the morning in CHAPS Sterling. For example, pay-ins to CLS Bank must be made by 11:00 a.m. in order to avoid significant financial penalties, and market convention dictates that overnight interbank loans should be repaid the following morning. Even for those banks for which collateral posting is relatively costly, the expected cost of delay for time-critical payments may be so high as to warrant posting sufficient collateral at the beginning of the day to ensure that liquidity is available to make these payments without the need for recourse to incoming funds. The existence of throughput guidelines may serve to reinforce the incentive to post liquidity "up front" (as we discuss in more detail in Section 2.5).

[6] We are indebted to Peter Gibbard for these insights.

[7] It should be noted that CHAPS Sterling payments are not the only claim on the available liquidity. Participants in the United Kingdom's securities settlement system, CREST, are able to transfer liquidity from CHAPS Sterling settlement accounts to separate accounts designated for the settlement of the cash legs of securities transactions. Liquidity is also available for the settlement of positions in other ancillary systems, such as CLS Bank, retail payments systems (BACS and C&CC), and LCH.Clearnet. Unlike CLS Bank pay-ins, the latter transfers do not take place in CHAPS Sterling and therefore are not recorded in the payment data in this article.

[8] In addition to the factors described, there may be frictions associated with obtaining eligible securities during the day that tend to encourage early posting. Under normal circumstances, intraday repos with the Bank of England are unwound at the end of the day and the securities are held in custody by the Bank of England overnight, to be reposted the following morning.

Box 1

The Cost of Liquidity in CHAPS Sterling

James and Willison (2004) estimate the opportunity cost of posting eligible collateral as the difference between the unsecured interbank rate and the secured-lending repo rate. By posting collateral, the bank forgoes the opportunity to repo the securities and to lend the funds obtained at a higher rate in the interbank market.

As James and Willison acknowledge, this is only part of the story. U.K. banks are subject to the Stock Liquidity Regime (SLR), under which they are required to hold liquid assets sufficient to cover net sight deposit and five-day wholesale cash outflows. These assets cannot be repurchased overnight to generate cash in the interbank market, but they can be posted with the Bank of England to generate liquidity in CHAPS Sterling (since the SLR requirements are measured only at the end of the day). For those banks subject to the SLR, the opportunity cost of posting collateral—and hence the cost of liquidity—may be even lower than the 7 basis point estimate.

Three foreign settlement banks in CHAPS Sterling—accounting for around 14 percent of transactions by value or 11 percent of transactions by volume—are not subject to the SLR. These banks are instead subject to the Maturity Mismatch Regime, which does not require banks to hold eligible liquid assets if committed outflows equal expected inflows. For these banks, the opportunity cost of posting collateral may be higher. However, the use by these banks of cross-border collateral arrangements (including the ability to back sterling payments with euro cash collateral) implies that estimation of the cost of liquidity for foreign banks would require analysis of the cost of generating liquidity in other jurisdictions.

Following reform of the Bank of England's money market operations in May 2006, participants are now able to hold remunerated reserve account balances with the Bank of England. These balances can be used to fund payments, so members can choose to provide liquidity in this way rather than by posting eligible securities. The relative opportunity cost of holding reserve account balances may vary from member to member, although for some foreign banks, particularly those that do not routinely hold sterling collateral, reserve balances may be a relatively attractive source of liquidity.

The value of collateral posted to support these time-critical payments depends on the expectation of the size of time-critical payment flows. Consider a single period in which delay costs are zero up to a certain time (say, the deadline for a time-critical payment) and very high thereafter. If all payment instructions are known at the beginning of the period in which time-critical payments must be made—and therefore all banks are aware of whether they will be net payers or net receivers at the end of the period—it can be shown that each net payer must raise liquidity at least equal to the value of its net payments and that this amount will be necessary and sufficient to settle all payments in the system (Box 2). Net receivers will be able to meet their payment obligations using incoming funds and hence will not need to raise additional liquidity.

Box 2

Liquidity Requirements for Time-Critical Payments

For simplicity, assume that banks receive all payment instructions exogenously from their customers. These instructions are denoted by x_t^{ij}—that is, at time t, bank i is requested by its customer(s) to pay the amount x to bank j. A payment from bank i to bank j at time t is denoted p_t^{ij}. Banks choose whether to settle payment instructions immediately or to queue them internally. So, p_t^{ij} and x_t^{ij} need not be the same.

Consider the case in which delay costs are zero up to a certain time T and subsequently so high that all payments must be settled within T. It follows that:

$$\sum_{t=0..T} x_t^{ij} = \sum_{t=0..T} p_t^{ij} \quad \forall i,j .$$

The payment balance of bank i against bank j at time t is defined as:

$$b_t^{ij} = \sum_{s=0..t} p_s^{ji} - \sum_{s=0..t} p_s^{ij} .$$

Bank i is a net payer for the period if its total payment balance at T is negative—that is, if $b_T^i = \sum_{j \neq i} b_T^{ij} < 0$. Since customer orders are exogenous, banks cannot affect whether they will be net payers or net receivers. We assume, however, that banks know with certainty at the beginning of the period which type they will be.

We define I as the set of net payers. Each net payer i has to raise liquidity to a value at least equal to $L^i = -\sum_{j \neq i} b_T^{ij}$ in order to execute its payment instructions. Hence, $\sum_{i \in I} L^i$ is the minimum liquidity required to settle all payment instructions. This amount will also be sufficient to settle all payments by time T if, first, every net payer raises L_i at time zero and pays it out immediately and, second, at any t, every bank i uses all of its liquidity to make payments up to that value (or less, up to the exhaustion of queued orders).[a] Because delay costs are zero up to T, this pattern is optimal for all banks. We can therefore conclude that, when delay costs are zero up to a time-critical threshold and very high thereafter, the banks' interests are aligned and compatible with the efficient use of liquidity. All payments are settled using only the minimum liquidity L.

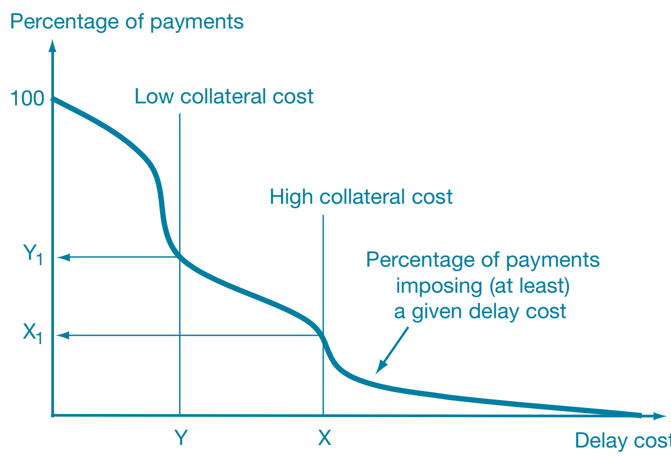
[a] We are abstracting here from 1) the indivisibility of payments—that is, additional liquidity may be required if payments cannot be split and settled in tranches, and 2) the possibility that no bank is a net payer—meaning that all payments net out exactly. In this case, a bargaining process would be required to define who is to provide liquidity, given that liquidity is required, yet no one needs to post collateral if somebody else does.

Banks' incentives are therefore aligned and consistent with the efficient use of liquidity.

However, this result requires that all participants know at the beginning of the period whether they will be net payers or net receivers at the end. In reality, participants will not typically possess full information about payment flows at the time when collateral posting decisions are made, and hence they will face uncertainty about liquidity requirements. Faced with high threshold delay costs, participants will wish to insure themselves against the risk of being net payers and—if the cost of failing to make a time-critical payment is sufficiently high—they may choose to post liquidity at the beginning of the period to a value at least equal to the maximum anticipated gross value of the time-critical payments. For payments with low delay costs, by contrast, participants may be willing to rely on incoming funds rather than post additional collateral. We illustrate this scenario using a simple stylised framework (see exhibit).

Here, the choice of the value of collateral posted at the beginning of the day is determined by the intersection of the expected cost of delay (which varies across payments; in the exhibit, X_1 percent of payments incurs a delay cost of at least X) and the cost of collateral (which, as described above, is fixed once collateral is posted). The value of collateral posted must be sufficient to ensure that time-critical payments—for which the cost of delay is greater than the cost of liquidity—can be made without the need for recourse to incoming funds. By contrast, for those payments for which the cost of delay is lower than the opportunity cost of posting collateral (the proportion of payments 100 percent *minus* X_1 for participant X), participants may be willing to rely on the recycling of incoming funds instead of posting additional liquidity. Such payments—particularly those of high value—will typically be delayed until after time-critical payments are settled, especially when there is uncertainty over the liquidity demands of time-

Delay and Liquidity Costs in CHAPS Sterling System



critical payments.[9] In addition, to the extent that there is reliance on incoming funds, the liquidity released by the settlement of time-critical payments may result in a "liquidity cascade," as queued payments are released. In this way, the early settlement of time-critical payments may serve to catalyse liquidity recycling later in the day.

The proportion of payments to which this applies will vary according to the cost of liquidity. As the exhibit illustrates, banks for which the opportunity cost of posting collateral is relatively low (for example, cost of collateral Y) will post a larger stock of collateral and hence may tend to fund a greater volume of payments (Y_1) from posted liquidity than from incoming funds. However, banks for which liquidity is relatively costly may post less collateral and remain more reliant on the recycling of incoming payments to fund outgoing payments.

Of course, as in the single-period example above, we must also take account of the limited information available to participants when decisions are made. While a proportion of payment instructions may be known at the start of the day (which we would expect to be relatively large when payment activity is driven by proprietary rather than customer business), there may remain considerable uncertainty about the size and distribution of incoming and outgoing payments, including payments with high delay costs. Faced with such uncertainty about aggregate liquidity demands, participants would be expected to hold a buffer of liquidity in excess of the quantity predicted in this framework in order to withstand unforeseen liquidity demands. The tendency to maintain such liquidity cushions—which are indeed observed in practice—will also contribute to the entire system's resilience to liquidity shocks, such as the operational failure of one or more banks to make payments.

2.5 Liquidity Recycling in CHAPS Sterling

We have discussed how the apparently low opportunity cost of collateral for many participants and the high expected delay costs associated with a subset of payments will tend to favour posting collateral at the beginning of the day. This may serve to reduce the incentive for payment delay and hence avoid the prisoner's dilemma outcomes described in the theoretical models. However, we have also seen that for payments for which the cost of delay is relatively low, participants may rely to a greater extent on incoming funds as a funding source in order to avoid squeezing the precautionary buffer of spare liquidity—or indeed to avoid posting additional collateral.

The efficiency with which incoming funds are recycled will depend on the extent to which participants collectively maintain the flow of liquidity around the system, perhaps via proactive payment coordination. McAndrews and Rajan (2000) describe how such coordination is achieved in Fedwire through the delay of payments until an end-of-day activity peak, but suggest that the observed level of coordination may be inefficiently low—and hence liquidity costs inefficiently high—due to collective-action problems, exemplifying the uncooperative outcome in the prisoner's dilemma game described by Bech and Garratt (2003). In principle, CHAPS Sterling members could suffer from a similarly uncooperative outcome in which some members defect and withhold payments, thereby curtailing the ability of others to take advantage of incoming funds.

The efficiency with which incoming funds are recycled will depend on the extent to which participants collectively maintain the flow of liquidity around the system, perhaps via proactive payment coordination.

In practice, the early settlement of time-critical payments will contribute to the recycling of liquidity by ensuring that payments begin to flow early in the day. In addition, certain features of CHAPS Sterling may be particularly conducive to achieving a cooperative outcome and hence to ensuring that liquidity is recycled efficiently. For example, the CHAPS Clearing Company imposes a set of *throughput guidelines*, whereby participants are expected to make 50 percent of payments by value by 12:00 p.m. and 75 percent by 2:30 p.m., as an average over a calendar month, with the explicit intent of improving the efficiency of liquidity usage in the system. While enforcement of the guidelines relies on peer pressure rather than legal compulsion (Box 3), the guidelines are largely observed in practice. As Buckle and Campbell (2003) demonstrate, the existence of such guidelines acts to countervail any tendency toward payment delay and hence serves to promote liquidity recycling earlier in the settlement day, thereby enhancing the efficiency of the payments system.

[9] Empirically, banks tend to settle a large volume of low-value payments early in the day. This might appear to contradict the predictions of this model, since the inherent delay cost associated with any one of these payments is likely to be low. However, expected delay costs for these payments *collectively* may be high, since processing of large volumes later in the day may prove difficult and costly in the event of an operational incident.

Box 3

Enforcement of CHAPS Throughput Guidelines

If a CHAPS Sterling member breaches the throughput guidelines in three consecutive months, that member is required to provide reasons to the CHAPS Clearing Company and to outline the steps taken to ensure that deadlines are met going forward. The participant will be given the opportunity to provide evidence that, over the period in question, failure to meet the guidelines resulted from a lack of payment instructions rather than a shortage of available liquidity.

If the member breaches the guidelines in six consecutive months, or in three consecutive months on two occasions, and has been unable to provide evidence as set out above, it will be obliged to attend a "Star Chamber" hearing. At the Star Chamber, the member's CHAPS board director will be required to explain the steps being taken to resolve the issues and to return performance to acceptable service levels and guidelines.

There is no defined penalty for the breach: As a rule, peer pressure is felt to be sufficient. However, the CHAPS Rules give the company manager the power to suspend or exclude a member "in material breach" of the provisions of the procedural rules, or where, in the opinion of the CHAPS Clearing Company, circumstances have arisen that could be "prejudicial" to the system or represent a threat to its "security, integrity, or reputation."

In addition, the concentrated structure of CHAPS Sterling appears more conducive to coordinated behaviour than does the structure of Fedwire, which has a broader membership. CHAPS has fifteen direct members (including the Bank of England), and the majority of payments are made by a core of four participants. The Fedwire network is more extensive, as around 9,500 participants access the clearings directly.[10] This results in very different network topologies, which in turn has implications for the flow of liquidity around the systems.[11] In particular, the concentration of payment flows among a small group of banks in CHAPS Sterling leads naturally to a higher level of recycling throughout the day than would occur in a more dispersed system, since each unit of liquidity paid out is more likely to be returned quickly if payments are flowing between fewer banks. Furthermore, within a "small club" of participants, the behaviour of each participant is highly visible to others. If one participant defects and fails to provide liquidity to the system, other participants may adopt a punishment strategy, such as delaying their own payments to that member. This cost associated with being perceived as "free-riding" on liquidity provided by peers in a repeated game may thus induce a cooperative outcome.

One specific mechanism possibly enforcing such discipline is the use of *bilateral net sender limits*. This is the simple liquidity management rule whereby bank A ceases to make payments to bank B if the *net* flux of payments from A to B reaches a certain (positive) *limit*; in other words, B is "punished" if it is seen "not to reciprocate." Anecdotal evidence suggests that this mechanism is indeed applied by some CHAPS members. The appendix formalizes the argument, but the logic behind bilateral net sender limits is that they create "interperiod spillovers," increasing the cost of delaying payments by depriving the recalcitrant bank of liquidity in subsequent periods. As a result, banks are encouraged to make payments promptly, to the benefit of the system.[12] As shown in the exhibit, the effect of such limits is to shift the delay curve to the right. In this framework, the result would be to increase the value of collateral posted at the beginning of the day. The effect on liquidity usage would depend on the effect on liquidity recycling over the course of the day.

But even though centralised throughput guidelines and decentralised mechanisms may serve as coordination devices, the use of bilateral net sender limits (in the form described above) would not enforce coordination on a particular time of the payment day and hence would not necessarily overcome a tendency to delay payments until the end of the day. Acting in tandem, however, throughput guidelines and bilateral coordination mechanisms can be expected to both enhance the efficiency of liquidity recycling and to smooth the intraday distribution of payments. We seek evidence of these effects in the empirical analysis that follows.

3. The Timing of CHAPS Sterling Payments

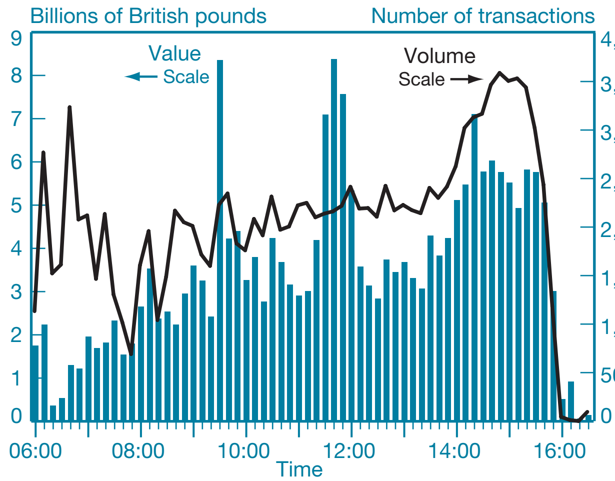
We now turn to an empirical analysis of the timing and funding of payments in CHAPS Sterling. Based on the discussion above, we would expect the low opportunity cost of posting collateral and the high expected delay costs associated with a subset of payments to limit the degree to which members delay payments in order to take advantage of incoming funds. We

[10] Only a small proportion of these banks use Fedwire heavily, however.

[11] The network of payments between settlement banks in CHAPS, however, is underlain by a more extensive network of payments between the originators of payments and the end recipients. The characteristics of this network are similar to those of Fedwire. See Soramäki et al. (2006) and Becher, Millard, and Soramäki (2007).

[12] Bilateral limits also have the important function of reducing the impact of "liquidity sinks," created when a bank is able to *receive* payments but is unable to *release* funds—for example, as a consequence of a technical outage. By restricting flows to the "sink" bank, bilateral limits reduce the amount of liquidity that is syphoned out of the system.

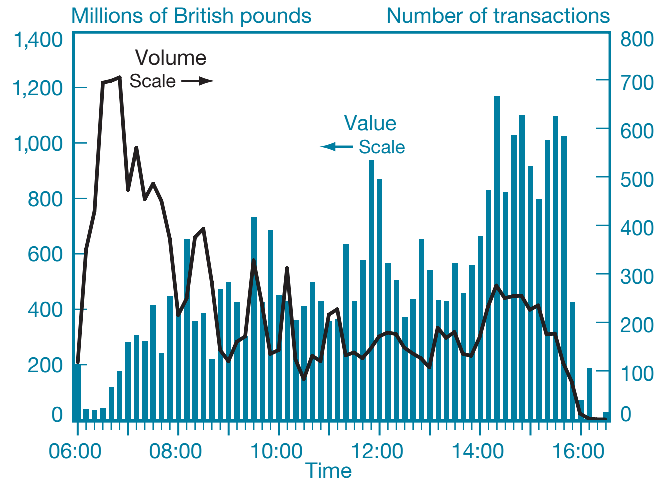
Chart 1
Value and Volume of Payments of All Banks in CHAPS Sterling System



Sources: CHAPS payment database; Bank of England calculations.

Note: We calculate the figures as daily average values and volumes in ten-minute intervals, using data from October 2006.

Chart 2
Value and Volume of Payments of Foreign Banks in CHAPS Sterling System



Sources: CHAPS payment database; Bank of England calculations.

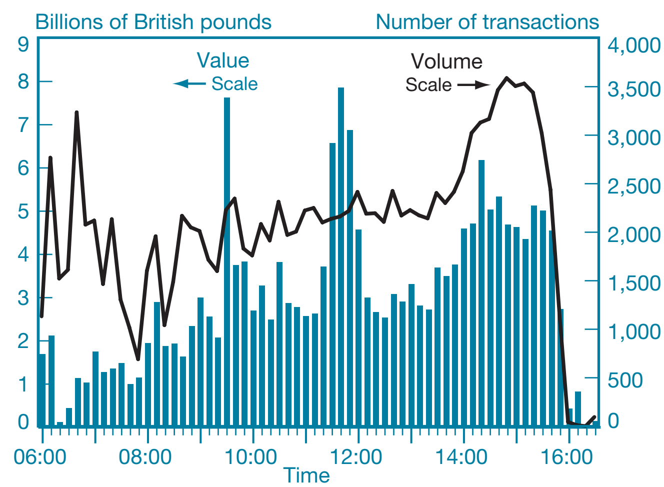
Note: We calculate the figures as daily average values and volumes in ten-minute intervals, using data from October 2006.

would expect to see this reflected in the intraday payment distribution. We also hypothesise that structural features of the CHAPS Sterling system, in particular the imposition of throughput guidelines and the "small club" membership, will promote the recycling of liquidity and smooth the distribution of payments throughout the day.

We first consider the intraday pattern of payments in CHAPS Sterling. The system opens for normal service at 6:00 a.m. and closes at 4:20 p.m. CHAPS settlement banks can initiate transfers on behalf of themselves and their clients normally until 4:00 p.m., although settlement members may make transfers on their own behalf, or on behalf of other credit institutions and certain money market participants, for the purpose of settling their end-of-day positions after this time.

The intraday profiles of payments in CHAPS Sterling are shown in Charts 1-3, alongside the profile of Fedwire payments in Chart 4.[13] The profile of payments by all banks (Chart 1) displays three distinct value peaks: the first around 9:30 a.m.; the second between roughly 11:30 a.m. and 12:00 p.m.; and a third, sustained peak between 2:00 p.m. and 4:00 p.m. The value profiles are similar for foreign and domestic banks (Charts 2 and 3, respectively: on average, 55 percent of foreign bank payments are made by noon, compared with 58 percent

Chart 3
Value and Volume of Payments of Domestic Banks in CHAPS Sterling System



Sources: CHAPS payment database; Bank of England calculations.

Note: We calculate the figures as daily average values and volumes in ten-minute intervals, using data from October 2006.

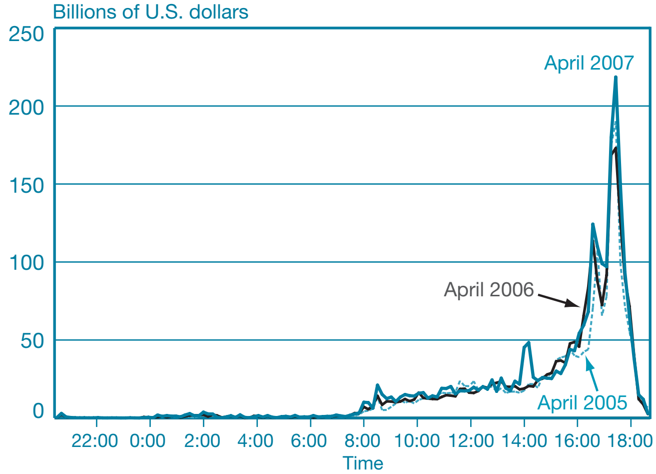
for domestic banks. Both domestic and foreign banks make around 25 to 30 percent of payments by value during the end-of-day value peak. The profile contrasts with that of payments in Fedwire, which exhibits strong concentration of payment value at the end of the day (Chart 4).

[13] The analysis in this section is based on data for CHAPS Sterling payment flows only; other transfers, such as settlement payments for BACS and C&CC, are not included. Our results apply to one month only (October 2006); however, our analysis has been repeated for data from June 2005 and January 2006, with similar results.

Chart 4

Value of Payments Made by Time of Day in Fedwire

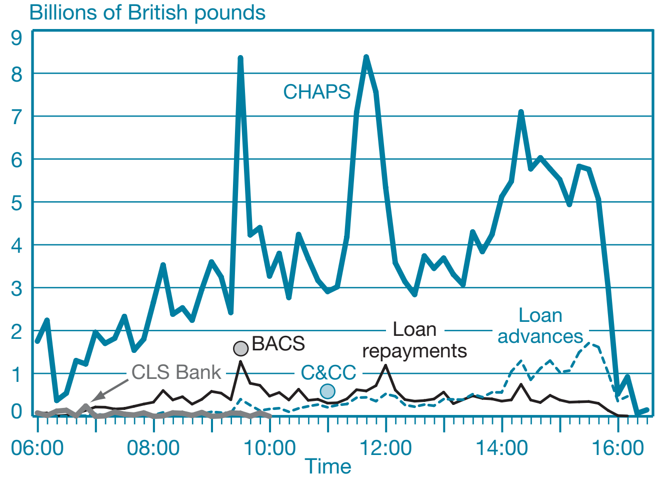
Daily Average by Month



Source: Federal Reserve Bank of New York.

Chart 5

Effect of Payment Events on the Intraday Distribution of CHAPS Sterling System Payments



Sources: CHAPS payment database; Bank of England calculations.

Note: We calculate the figures as daily average values and volumes in ten-minute intervals, using data from October 2006.

Setting aside for the moment the effect of strategic behaviour, we note that the observed *value* peaks correspond well with scheduled payment events, particularly those associated with time-critical payments and throughput deadlines. This result is illustrated in Chart 5.

The first peak (9:30 a.m.) temporally corresponds both with the timetable for pay-ins to CLS Bank, which can be made during a payment window between 7:00 a.m. and 11:00 a.m., and with the settlement of multilateral net positions in the BACS retail payments system.[14] In both cases, the value of the settlement payments involved is small relative to the total value of payments made at these particular times. However, the observed peak may reflect the tendency to delay payments until after the time-critical payments have been made, when any uncertainty around the value of these settlements has been resolved. The settlement payments may also release liquidity for the settlement of subsequent payments. Sharp value peaks also occur ahead of the throughput deadlines, at noon and 2:30 p.m., suggesting that the guidelines do impact significantly on the intraday distribution of payments. In fact, there is prima facie evidence that payments are delayed until the period immediately before the deadlines, which may reflect strategic behaviour.[15]

Finally, the value peaks may also reflect the routine patterns of activity in the overnight interbank market: Late-afternoon value peaks are likely to be reinforced by the creation of overnight loans for the purposes of position-squaring, which must typically be repaid the following morning. Going forward, the effect of overnight markets on payment profiles and liquidity usage is fertile ground for future research.

The *volume* profile of CHAPS Sterling payments is relatively smooth throughout the day for the system as a whole. Volume peaks occur shortly after opening and again late in the day. The volume profile is notably different for the set of foreign banks: Volumes are highly concentrated during the first two hours and fall away sharply thereafter. Approximately 40 percent of foreign banks' payments by volume are made by 8:00 a.m., compared with only around 15 percent for domestic banks. This may in part reflect the settlement of payments queued between the opening of continental European markets and the opening of CHAPS Sterling.

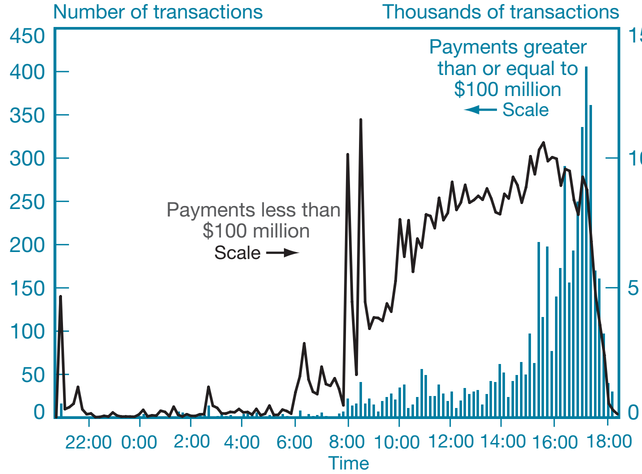
The concentration of payment value in Fedwire appears to be driven by a distinct skew in large-value payments toward the end of the day (Chart 6). This distribution may reflect institutionally imposed timings for certain types of payments, such as for CHIPS and DTC settlement; the creation of overnight loans; and settlement payments in financial markets. This is consistent with McAndrews and Rajan's (2000) observation that this peak existed prior to the imposition of overdraft fees. However, as discussed above, the peak may additionally serve as a focal point for, and be reinforced by, proactive payment coordination.

[14] The plots shown do not include these payments.

[15] The noon peak also follows the settlement of positions in the C&CC. This may be influential, although the very low value of settlement payments in this system suggests that it is unlikely to trigger a liquidity cascade or to generate material uncertainty for participants.

Chart 6

Distribution of Fedwire Payments by Size of Payment



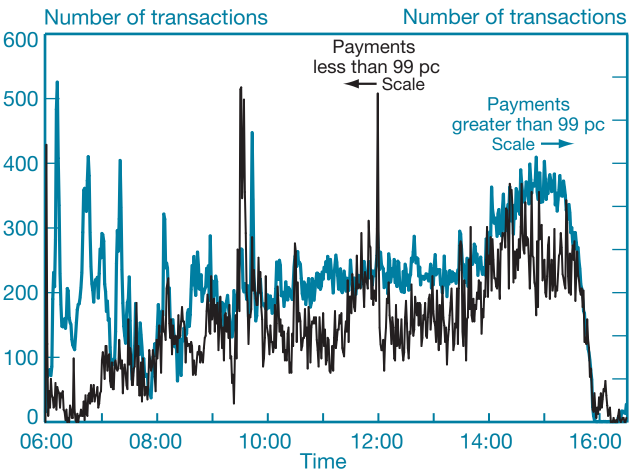
Source: Federal Reserve Bank of New York.

Note: $100 million was the 99th percentile for payment size on March 19, 2007.

Chart 7

Distribution of CHAPS Sterling System Payments by Size of Payment

One-Minute Intervals, Daily Average, October 2006



Sources: CHAPS payment database; Bank of England calculations.

Large-value payments in CHAPS Sterling, defined as those that fall within the 99th percentile of the distribution of payment sizes, account for around 75 percent of daily payment value. As Chart 7 illustrates, the volume of payments per minute in this category is very small, and the intraday distribution is smoother than in Fedwire (Chart 6). Peaks in the incidence of large-value payments occur at 9:30 a.m., 12:00 p.m. (perhaps related to the first throughput guideline), and at the end of the day (consistent with the timing of large position-squaring payments in the interbank market). This suggests that the distribution of large-value payments is driven by the purposes of the payments in question and less by a generalised tendency to delay until the end of the day.

The early concentration of time-critical payments in CHAPS Sterling (in conjunction with other ancillary system settlement payments and the possible need for additional liquidity transfers to CREST) may also help explain why the average value of payments made during the first two hours of opening is low. Participants may be reluctant to commit liquidity to other large-value payments until these time-critical payments have been made. This applies less forcefully to low-value payments; indeed, it is apparent from Chart 7 that low-value payments are released into the system early, perhaps reflecting their low consumption of liquidity.[16]

The contrast between the payment profiles in CHAPS Sterling and in Fedwire—in particular, the observation that payments are much less concentrated at the end of the day in CHAPS Sterling—provides some initial evidence that the incentives for payment delay are weaker in CHAPS Sterling than in Fedwire.[17] The profile of CHAPS Sterling payments is clearly influenced by the existence of time-critical payments and throughput guidelines. To consider whether these patterns are also influenced by the strategic behaviour of members, we now attempt to disaggregate the sources of funding. In particular, we assess whether there is evidence that the use of incoming funds varies by time of day and, following McAndrews and Rajan (2000), whether peaks in liquidity recycling coincide with peaks in payment activity.

4. The Funding of CHAPS Sterling Payments

4.1 Methodology

To decompose the sources of funding of CHAPS Sterling payments, we distinguish between two sources of funding: 1) payments received from other CHAPS Sterling participants within a specified time interval and 2) account balances held at the Bank of England, funded both by collateral posting and the maintenance of positive reserve account balances.

[16] The tendency to settle a high volume of low-value payments early in the day may also reflect the relative complexity of settling high volumes of low-value payments in the event of an operational disruption later in the day. See also footnote 9.

[17] Of course, it is not possible to observe delay directly from the intraday payment profile. This would require knowledge of the timing of payment instructions, as well as of settlement.

Box 4

Measurement of "Offsetting" Payments in CHAPS Sterling

We define the value of payments made from member *i* to member *j* within minute *t* as p_{ij}^t.

The total value of payments made within that minute is therefore $\sum_{i,j} p_{ij}^t$ and the value of net payments for each member, *i*, is $\sum_j (p_{ij}^t - p_{ji}^t) = N$.

The value of payments *not* offset within a minute is equal to the sum of net payments for the set of members for which *N* is positive, or, equivalently, $\frac{1}{2}\sum_i |N|$.

The *value of offsetting payments* is then calculated as the value of gross payments made, less the value of payments not offset: $\sum_{i,j} p_{ij}^t - \frac{1}{2}\sum_i |N|$.

In their study of payment activity in Fedwire, McAndrews and Rajan (2000) consider incoming payments as a funding source for outgoing payments only if those incoming payments offset outgoing payments *within the same one-minute interval*. This definition follows naturally from the overdraft charging structure, as fees are based on the outstanding overdraft at the end of each minute. Provided that all payments are offset by incoming payments within that same minute, irrespective of the ordering of the payments, no charge is incurred.

We choose to adopt the same methodology for the measurement of incoming payments as a funding source in CHAPS Sterling (Box 4). While recognising that payments cannot be "offset" within a minute in the same way as in Fedwire (if outgoing payments are made first, intraday liquidity will be required even if it is subsequently replenished by incoming funds), this measure is useful both as a point of comparison with Fedwire and as an indicator of liquidity recycling in CHAPS Sterling.[18] It should be noted, however, that this particular measure does not capture the recycling of funds hoarded from previous time periods, which is a significant omission. We return to this point in a subsequent discussion.

[18] It is also questionable whether active offsetting of payments within a minute is a realistic representation of members' liquidity management processes. However, as discussed in Section 2.5, the use of bilateral limits may result in such behaviour being observed, since outgoing payments may be released immediately when incoming payments create headroom under a bilateral limit.

4.2 Results

We now decompose gross payments into the constituent funding sources (Chart 8). While the *absolute* value of payments "offset" by incoming payments increases when the value of gross payments increases, the *proportion* of payments funded by incoming payments remains comparatively stable throughout the day. On average, during the day, around 23 percent of payments made are funded using incoming payments; the use of incoming payments peaks at around 42 percent shortly before noon.

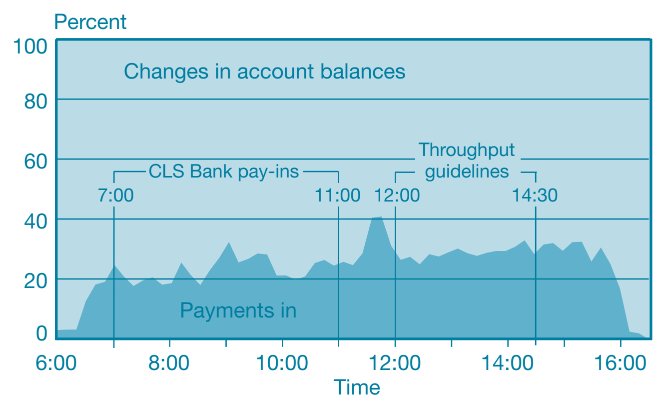
Compared with the results for Fedwire (Chart 9), CHAPS Sterling does not exhibit a pronounced peak in the share of incoming payments as a funding source during the end-of-day value peak, although the proportion of payments funded by incoming funds is above the daily average at this time. It is, however, notable that a distinct peak occurs shortly before the first throughput deadline (at 12:00 p.m.), suggesting that incoming payments are a particularly important funding source at this time. This may ease the liquidity demands of meeting the throughput deadline; indeed, this concentration may be a product of deliberate payment coordination on the focal point(s) provided by throughput guidelines.

Perhaps the most striking observation from Charts 8 and 9, however, is that on this measure, the *average* level of liquidity recycling within each minute is considerably higher in CHAPS Sterling than in Fedwire. In other words, the *level* of liquidity recycling, whether as a result of active coordination or otherwise, appears to be greater throughout the day in CHAPS Sterling. This result is consistent with our hypothesis that the

Chart 8

Shares of Funding Sources of CHAPS Sterling System Payments

October 2006

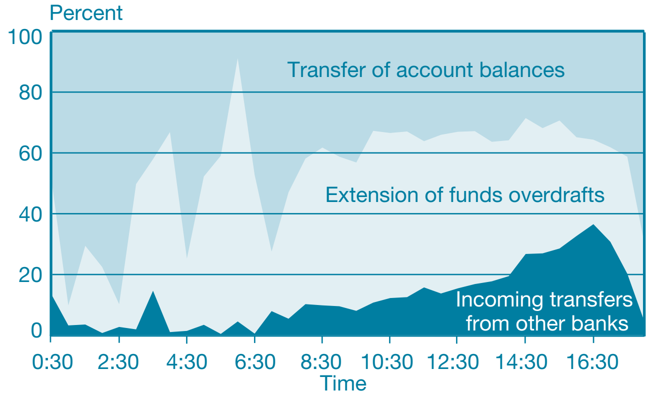


Sources: CHAPS payment database; Bank of England calculations.

Chart 9

Shares of Funding Sources of Fedwire Funds Transfers

Average of Four Days over Half-Hour Intervals



Source: Federal Reserve Bank of New York.

Note: Because few payments are made between 12:30 a.m. and 8:30 a.m., the variation in the shares of funding sources during that period of the day is driven by a small number of payments.

more concentrated structure of CHAPS Sterling is likely to be more conducive to liquidity recycling throughout the day, both as a natural consequence of there being fewer participants and as a result of bilateral coordination resulting from "small club" behaviour. Such behaviour, in combination with the distribution of time-critical payments and the effect of throughput guidelines, may ensure that liquidity continues to flow smoothly through the system.

4.3 Liquidity Recycling and Liquidity Constraints

We noted earlier that the opportunity cost of posting collateral (or holding positive reserve account balances) may not be uniform across all CHAPS Sterling participants. In particular, the cost of collateral, and hence of liquidity, should be higher for foreign banks, since they are not subject to the Stock Liquidity Regime. If this is the case, foreign banks would have a stronger incentive to fund payments using incoming funds, and they would be seen to attain higher recycling ratios. Is this indeed the case?

To answer that question, we consider the relationship between the maximum proportion of liquidity drawn down and the overall level of liquidity recycling achieved by each member during the day, measured as the ratio of the value of total daily gross payments to the maximum value of liquidity used (*the recycling ratio*). This measure is not subject to the critique of the previous section, since it does capture the benefits of liquidity hoarded over multiple periods.

Recycling and Liquidity Usage

Daily Average, October 2006

Recycling Ratio (r)	Number of Banks	Liquidity Used (Minimum-Maximum Range, in Percent)
$0 < r \leq 5$	2	68.0 - 85.1
$0 < r \leq 10$	5	37.0 - 99.7
$r > 10$	5	12.5 - 58.5

Sources: CHAPS payment database; Bank of England calculations.

Notes: We exclude the Bank of England and CLS Bank. The Royal Bank of Scotland and Natwest are treated as a single entity (the Royal Bank of Scotland Group), although they retain separate settlement accounts.

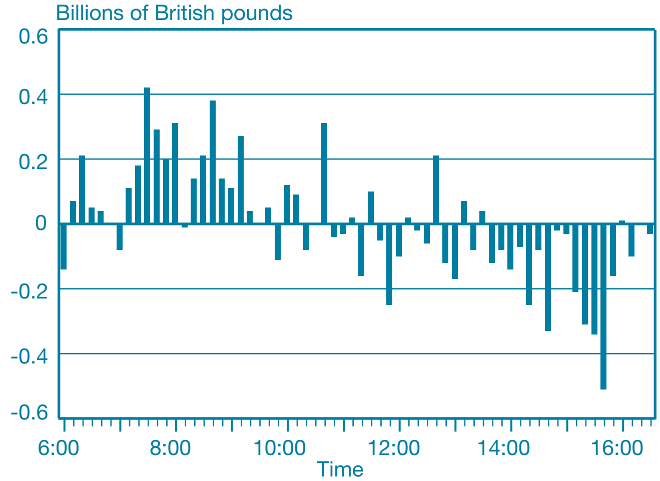
There is a wide variation in the extent of liquidity recycling achieved by CHAPS Sterling members (see table). Only five banks—all domestic—achieve recycling ratios greater than 10, and in each of these cases the proportion of liquidity drawn down is relatively low.[19] The foreign banks achieve lower recycling ratios and draw down a correspondingly large proportion of available liquidity. This implies that foreign banks may indeed face greater liquidity constraints than domestic ones, as suggested earlier in our discussion of the cost of collateral. As noted, we would expect the incentive to delay payments in order to take advantage of liquidity recycling to be correspondingly high, but this expectation is not supported by the data. How might this be explained?

From Section 3, we know that the intraday value profiles of *outgoing* payments made by domestic and foreign banks are similar. However, the ability to recycle liquidity and thereby lower aggregate liquidity requirements during the day also depends on the distribution of *incoming* payment flows. Charts 10 and 11 clearly illustrate that the pattern of net payments is very different on aggregate for domestic and foreign banks, even though all members must comply with the throughput guidelines.

Domestic banks are, on aggregate, net recipients of funds in the morning and net suppliers of funds in the afternoon. Foreign banks exhibit the opposite trend: Net payments are negative until late morning and become positive thereafter. This implies that domestic banks tend to accumulate funds during the morning and then pay these funds out in the afternoon, thereby reducing intraday liquidity usage and increasing the recycling ratio. When the flows are reversed,

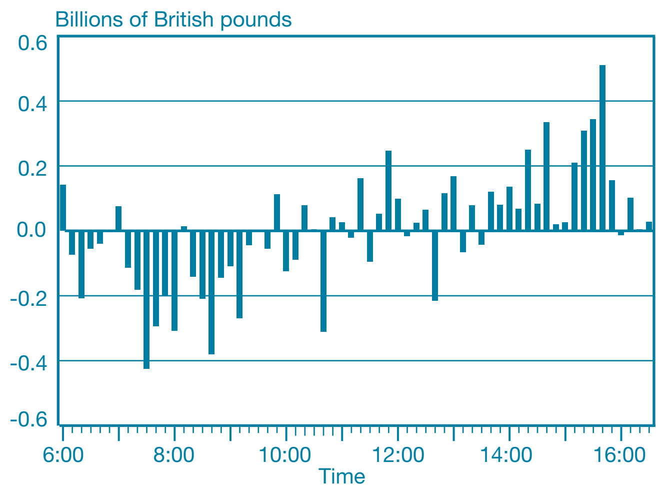
[19] Note that foreign and domestic banks are not differentiated in the table.

Chart 10
Net Payments for Domestic Banks in CHAPS Sterling System
Daily Average, October 2006



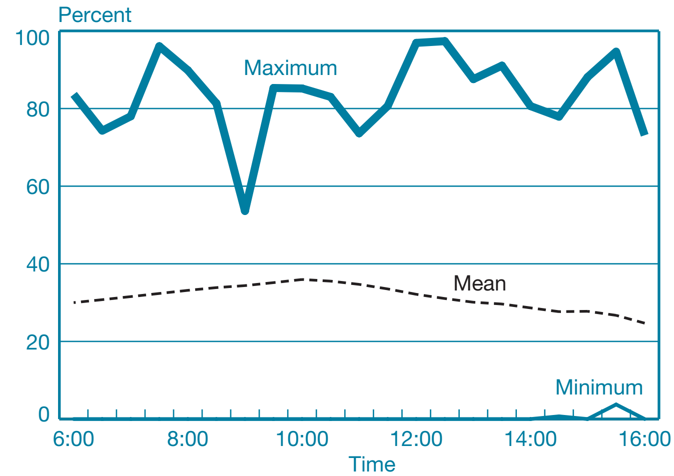
Sources: CHAPS payment database; Bank of England calculations.

Chart 11
Net Payments for Foreign Banks in CHAPS Sterling System
Daily Average, October 2006



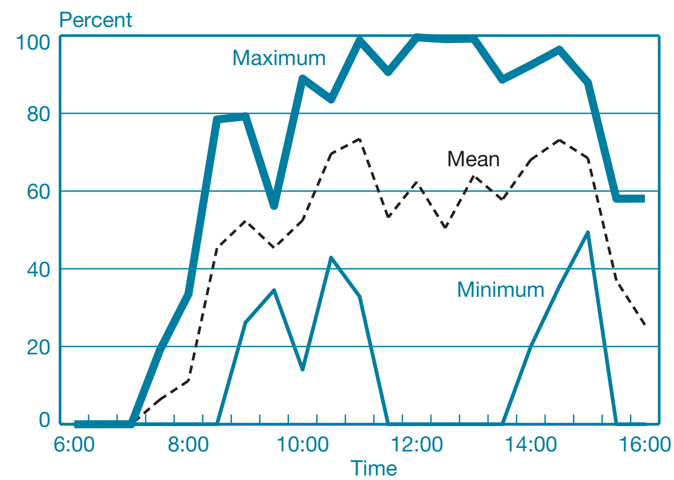
Sources: CHAPS payment database; Bank of England calculations.

Chart 12
Liquidity Usage by Domestic Banks
Daily Average, October 2006



Sources: CHAPS payment database; Bank of England calculations.

Chart 13
Liquidity Usage by Foreign Banks
Daily Average, October 2006



Sources: CHAPS payment database; Bank of England calculations.

as is the case for foreign banks, liquidity must be drawn down to fund the net outgoing payments; hence, one would expect recycling ratios to be lower. However, while the distinction between the payment flows of the domestic and foreign banks at the aggregate level is striking, these results conceal considerable variation within both sets of banks. This is apparent from the intraday patterns of liquidity usage, illustrated in Charts 12 and 13.

While it is clear that the maximum proportion of liquidity used by foreign banks is, on average, higher for much of the day than it is for domestic banks, there is considerable variation within both sets of banks. Some domestic banks are net payers for much of the morning and use a large proportion of their available liquidity early in the day. For these banks, the recycling ratio is relatively low. Other domestic banks, by contrast, use little or no liquidity for much of the day. These banks are more able to take advantage of funds received in the morning and achieve the highest recycling ratios of all CHAPS Sterling members. By contrast, all foreign banks in CHAPS Sterling are net payers during the morning and approach their maximum liquidity usage early in the day.

So while liquidity recycling appears to be relatively efficient in CHAPS Sterling, the extent to which individual banks benefit from recycling varies considerably. For those banks that are net payers early in the day—including all of the foreign banks and some large domestic banks—recycling ratios are much lower. One could argue that this results from the low cost of liquidity, since the incentive to structure payments so as to reduce liquidity costs may be weak. But many of the banks with relatively low recycling ratios do appear to face liquidity constraints, since they also use a large proportion of liquidity posted. This suggests that *some* banks may be unable to recycle liquidity to the extent that they would wish, which may reflect the simple observation that coordination will result only if all (or a sufficient number) of the banks are similarly incentivised by liquidity pressures to cooperate.

> *While liquidity recycling appears to be relatively efficient in CHAPS Sterling, the extent to which individual banks benefit from recycling varies considerably.*

The variation in recycling ratios also reflects the effect of the other influences on payment timing. The observed patterns of net payments in Charts 10 and 11 are likely to reflect structural differences in the underlying businesses of the participants and their customers, resulting in differences in the distribution of payment instructions and deadlines. If, for example, certain participants (or their customers) routinely borrow in the overnight market while others lend, the payment flows of the two groups will be correspondingly different. To the extent that these structural factors limit participants' discretion over the timing of payments, this may explain the observed variation in the distribution of recycling benefits.[20]

5. Discussion and Conclusions

Our analysis indicates that even though the intraday liquidity regime supporting CHAPS Sterling payments does not give rise to the same incentives for minute-by-minute payment coordination as those in Fedwire, the observed degree of liquidity recycling appears to be high. We have also seen that the intraday profile of payments is comparatively smooth. Taken together, these observations reveal that even if collateral posting is perceived to be costly by some banks—and hence a "liquidity incentive to delay" does exist in CHAPS Sterling—other features of the system help avoid a prisoner's dilemma equilibrium in which the majority of payments are delayed until late in the day. This serves to reduce the maximum liquidity required to make a given set of payments and hence the aggregate value of collateral that needs to be posted. The empirical evidence suggests that payment coordination may also play an important role in Fedwire, although in this case coordination—and consequently liquidity recycling—is strongly concentrated around an end-of-day focal point.

Which features of the system support this high and constant level of liquidity recycling? Centralised coordination devices are likely to play a role; in particular, throughput guidelines counteract any generalised tendency to delay payments until the end of the day. Indeed, the spike in the proportion of payments "offset" before noon is evidence that the incoming funds become an increasingly significant funding source at this time of day, reducing the liquidity cost of complying with the deadline.

Other forms of "decentralised coordination" between members may also be significant. The high visibility of payment flows in the concentrated CHAPS system allows members to monitor their bilateral positions and to take action if counterparts fail to make payments in a timely fashion. The prisoner's dilemma may then simply be resolved through the repeated interaction of the small number of participants, whereby recalcitrant participants are "punished" for failing to provide liquidity to the system. It is arguable that these pressures may be less strong in the more diffuse Fedwire system. While not explicitly revealed by the aggregate data, there is also anecdotal evidence that participants apply bilateral net sender limits with respect to other system participants, thereby promoting the recycling of liquidity between each bilateral pair and enhancing the liquidity efficiency of the system. An empirical question remains as to how often these limits "bite" in practice, but a "small club" like CHAPS is a natural environment for the application of such devices, which help generate a smooth payment profile.

The patterns in the timing and funding of CHAPS Sterling payments described in this article would appear to be risk-beneficial, for individual participants and for the system as a whole. The low opportunity cost of posting collateral and the tendency for all members (domestic and foreign) to post collateral at the beginning of the day help ensure that time-critical payments do not fail for want of liquidity. Moreover, the apparently low opportunity cost of collateral results in many banks providing a liquidity cushion in excess of that

[20] The patterns of funding flows in the overnight market are the subject of ongoing research at the Bank of England.

required to make time-critical payments. This not only increases the resilience of that participant to liquidity shocks, but also contributes to the resilience of the system as a whole.

The efficient recycling of liquidity during the day further contributes to a reduction in liquidity risk by reducing the aggregate liquidity required to make a given set of payments. In theory, this is particularly beneficial for those banks that face a relatively high cost of liquidity. However, we have seen that many of the members that draw down a large proportion of available liquidity are unable to take advantage of incoming funds to the same extent as other members, perhaps owing to the distribution of underlying payment orders from customers. For these banks, the likelihood of needing to post additional collateral during the day may be correspondingly high.

Our analysis can take several interesting directions. In this article, we have formulated a number of hypotheses on the determinants of behaviour in a payments system, suggesting some implications for the efficiency of the system itself. However, we have not been able to disentangle fully the effects of the factors identified; this is perhaps an inevitable drawback given our descriptive approach. Formal analysis, supported and complemented by further econometric work on payment data, may help in this direction by shedding additional light on key issues such as the effect of membership size on payment behaviour and the precise way in which banks achieve coordination (using, for example, bilateral net sender limits). Our analysis of the variations in liquidity recycling intensity also makes a strong case for further analysis of the overnight market, particularly its effect on payment behaviour and liquidity usage. Ultimately, such analysis will contribute to an understanding of how a large-value payments system functions and suggest where and how risks to the system may crystallise.

Appendix: Effects of a Bilateral Net Sender Limit on a Bank's Behaviour

We present a simple, formal illustration of a single bank's problem when choosing an optimal level of liquidity. A bilateral net sender limit is shown to incentivise early liquidity provision and early payment, thereby generating high levels of liquidity recycling. No attempt is made here to develop a fully fledged game-theoretical model of payments. That model would be more complex, because incoming payments would need to be modeled as a strategic choice (by the other banks) instead of as an exogenous random variable. Considering a single bank's decisions in isolation allows us to focus on the marginal effect of a bilateral net sender limit on the incentives to post liquidity and to delay payments. We would nevertheless expect to find this effect in a more complex setting.

Setting

Suppose that our bank receives payment orders exogenously from its customers. To execute these orders, the bank requires liquidity, which can be obtained either by posting collateral or by waiting for exogenous (and random) incoming payments. We assume that our bank faces the following sequence of events, all of which occur within a fixed time interval *t* (which can be thought of as a metaphor for a trading day, or part of the day such as "the morning"):

$t.0$	$t.1$	$t.2$	$t.3$
x_t	w_t	y_t	$p_t,\ \delta(Q_t)$.

- At $t.0$, the bank receives payment orders to the value x_t .
- At $t.1$, the bank decides how much liquidity to raise, w_t , at a cost $\lambda(w_t)$.
- At $t.2$, incoming payments provide the bank with additional liquidity y_t , so the bank has total liquidity of $l_t = w_t + y_t$.
- At $t.3$, the bank makes payments p_t . If $l_t < x_t$, the bank can only pay up to $l_t (p_t = l_t)$, so it "queues" an amount of payments $Q_t = x_t - l_t$. If instead $l_t > x_t$, then $p_t = x_t$ and the bank has spare liquidity. To simplify, we assume that the cost of a backlog Q_t is a function $\delta(Q_t)$, with $\delta(Q_t) > 0$ if $Q > 0$, and $\delta(Q_t) < 0$ otherwise.[21] To simplify further, we assume that if $Q_t < 0$, the bank sells the spare liquidity in the market, immediately realising $\delta(Q_t)$; if instead $Q_t > 0$, then $\delta(Q_t)$ includes all costs derived from delaying payments, in particular, the cost of the extra liquidity with which the bank settles or cancels the queued payments.
- The bank then begins the next period $t + 1$ afresh, with no liquidity and no queues (all costs / benefits stemming from Q_t are accounted for by $\delta(Q_t)$).

We now look at the bank's incentives and its optimal choice. We want to show how a bilateral net sender limit incentivises short queues and thus liquidity recycling. To do so, we compare two cases, one in which there is no bilateral net sender limit and one in which a limit is implemented.

The Bank's Problem: I

No Bilateral Sender Limits

Suppose that incoming payments y_t arrive according to some exogenous distribution $f(.)$, which is independent of *t* and of the bank's choices. In this case, the bank's problem is actually a single-period maximisation,[22] so we are able to eliminate all time indices. By borrowing *w*, the bank reduces the expected queue Q , thus abating the expected delay costs $\delta(Q)$ (and possibly transforming them into a gain, if $Q < 0$). However, liquidity is costly, and the bank may hope to make use of receipts from the other bank (via *y*). In general, the bank will not raise a full $w = x$, and it will rely partly on incoming payments.

[21] A negative queue is a positive amount of liquidity whose positive value is a negative cost. Note that, to rule out the existence of "money-making machines," δ and λ (the cost of liquidity) are mutually restrained. See the example below.

[22] There are no spillover effects between t and $t+1$ because we assume that the bank realises $\delta(Q_t)$ from any queue or spare liquidity, beginning period $t+1$ afresh.

To find the optimal w, the bank looks at the total cost from raising w:

$$C(w) = \lambda(w) + E\delta(Q)$$
$$= \lambda(w) + \int_0^r f(y)\delta(x - w - y)dy,$$

where r is an upper bound on the incoming payments that the bank can expect to receive. Provided that some technical convexity conditions hold, the optimal amount of w^* is determined by solving the standard condition $C'(w) = 0$.

Bilateral Net Sender Limit

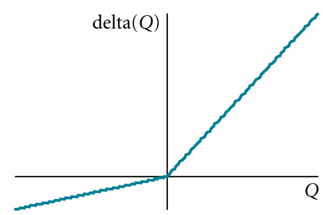
When a bilateral net sender limit is in place, the distribution of incoming payments y is no longer exogenous, but depends on previous payments. As a trivial example, if too few payments are sent out, the bilateral limit is hit and no further incoming y can be expected. More generally, the effect of choosing a particular w_{t-1}, and hence a volume of payments p_{t-1}, spills over to the next period and influences the expected amount of incoming funds y_t. As a consequence, compared with the case in which no bilateral net sender limits exist, every unit of liquidity made at $t-1$ now carries an extra benefit in terms of a liquidity saving at time t (although, of course, only in expected terms and only if it actually allows additional payments to be made at $t-1$). This interperiod spillover tilts the balance in favour of posting more w_t, increasing payments and reducing the queue. If, in addition, the liquidity cost λ increases in time, this mechanism is reinforced: The liquidity saved at $t+1$ by posting more w_t at t is even more valuable. Similarly, if the bilateral limit itself depends on i's past payments (for example, a "bad" payment record may induce the other banks to tighten prudentially their bilateral limit toward i), the incentives to pay early will be even stronger.

The Bank's Problem: II (Example)

We now solve analytically the bank's problem, under particular assumptions about the cost functions and the distribution of y. Suppose the delay cost is

$$\delta(Q) = \begin{cases} CQ & \text{if } Q \geq 0 \\ cQ & \text{if } Q < 0 \end{cases},$$

with $0 < c < C$. The corresponding graph is therefore:



In this case, the costs of a positive queue Q grow faster than the gains from spare liquidity (which are simply the negative costs from a negative queue).

We also assume that liquidity costs are linear: $\lambda(w) = \lambda w$ with $\lambda > 0$. To make the problem interesting, we first impose $\lambda < C$, implying that it is better to post liquidity and make a payment than not to post it and fail to make the payment, and then impose $c < \lambda$, which ensures that it is not optimal to post infinite liquidity. Indeed, if it were $c > \lambda$, then any pound of liquidity would be worth more to the bank than the cost λ, independently of whether it is used to shorten a queue (which gives a benefit $C > \lambda$) or if it results in spare liquidity (whose benefit is precisely c, which must therefore be assumed to be smaller than λ).

No Bilateral Net Sender Limit

Suppose y_t is uniformly distributed in $[0,\ r]$, so its probability density function is $f(y) = \frac{1}{r}$. Then, the bank's cost function is:

(1)
$$C(w) = \lambda(w) + \int_0^r f(y)\,\delta(x-w-y)\,dy =$$
$$= \begin{cases} \lambda w + \frac{1}{r}\left[C\int_0^{x-w}(x-w-y)dy + c\int_{x-w}^{r}(x-w-y)\,dy\right] & \text{if } x-w \le r \\ \lambda w + \frac{1}{r}\,C\int_0^r (x-w-y)\,dy & \text{if } r < x-w \end{cases}$$
$$= \begin{cases} \lambda w + \frac{1}{r}\left[C\frac{1}{2}(w-x)^2 - c\frac{1}{2}(r+w-x)^2\right] & \text{if } x-w \le r \\ w(\lambda - C) - \frac{1}{2}C(r-2x) & \text{if } r < x-w \end{cases}.$$

We now find the optimal w, to be called w^*.

- Suppose w^* is such that $x - w^* \le r$. In this case, the optimality condition would be

(Opt)
$$\frac{d}{dw}\left(\lambda w + \frac{1}{r}\left[C\frac{1}{2}(w-x)^2 - c\frac{1}{2}(r+w-x)^2\right]\right)$$
$$= \frac{1}{r}(r\lambda + Cw - Cx - cr - cw + cx) = 0$$

yielding an optimal liquidity $w^* = x + \frac{r(c-\lambda)}{C-c}$ and thus a total cost equal to

$$\frac{1}{2(C-c)}(-r\lambda^2 + 2Cx\lambda + 2cr\lambda - 2cx\lambda - Ccr) = C^*.$$

- Suppose instead $r < x - w^*$. Because $\lambda - C < 0$, the total cost decreases in w as long as $r < x - w$ (see equation 1). Hence, we would have a corner solution at $w^* = x - r$, which yields a cost $x\lambda - r\lambda + \frac{1}{2}Cr = C^{**}$.

Now, the difference $C^{**} - C^* = \frac{1}{2}\frac{r}{C-c}(C-\lambda)^2$ is always positive. Hence, the cost-minimizing w is the one found in the first case, supposing $x - w^* \le r$:

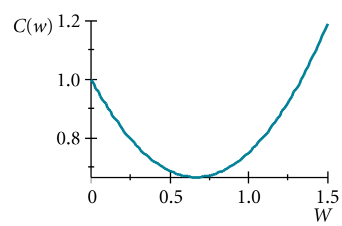
(2)
$$w^* = x + \frac{r(c-\lambda)}{C-c} < x.$$

It should be noted that w^* has the anticipated properties: It increases in x (the amount of payments to make) and in C (the cost of queues), and it falls with $\lambda - c$, that is, with the difference between the cost of liquidity and its benefits as end-of-day spare liquidity.

Finally, substitution of equation 2 into equation 1 yields the optimal (minimised) cost:

(3)
$$C^* = \frac{1}{2(C-c)}(-r\lambda^2 + 2Cx\lambda + 2cr\lambda - 2cx\lambda - Ccr).$$

If we set, for example, $c = \frac{1}{2}$, $\lambda = r = x = 1$, and $C = 2$, the graph of $C(w)$ is:



In this case, the liquidity posted is 66 percent of the payments due ($x = 1$).

Bilateral Net Sender Limit (BNSL)

Suppose again that y_t is uniformly distributed. This time, however, imagine that incoming payments are drawn from $[0, r_t]$, with r_t determined by a bilateral net sender limit:

$$r_t = r_{t-1} + (p_{t-1} - y_{t-1}).$$

In this case, an increase in p_{t-1} due to higher liquidity w_{t-1} pushes up r_t, which in turn affects the minimised costs (see equation 3). What is the value of such a spillover, from $t-1$ liquidity into time t costs? The answer is

$$\frac{dC_t^*}{dw_{t-1}} = E_{t-1}\left[\frac{dC_t^*}{dr_t}\frac{dr_t}{dw_{t-1}}\right],$$

which we calculate term by term.

The term $\frac{dC_t^*}{dr_t}$ immediately derives from equation 3, which reveals that the expectation term is irrelevant here (λ , c, and C are constant over time):

$$(4) \qquad \frac{dC_t^*}{dr_t} = -\frac{1}{2(C-c)}(\lambda^2 + Cc - 2c\lambda) = Z < 0 .$$

To prove the last inequality, recall that $c < \lambda < C$. This implies both $\lambda c < \lambda^2$ and $\lambda c < Cc$, which, summed member by member, gives $2\lambda c < \lambda^2 + Cc$.

Expectations do matter on the second term $\frac{dr_t}{dw_{t-1}}$. In fact, an extra pound worth of w_{t-1} translates into one more payment only if the available liquidity $w_{t-1} + y_{t-1}$ turns out to be less than the payment orders x_{t-1} , with y_{t-1} unknown at $t-1$. Formally,

$$\frac{dr_t}{dw_{t-1}} = \begin{cases} 1 & \text{if } w_{t-1} + y_{t-1} < x_{t-1} \\ 0 & \text{otherwise} \end{cases} .$$

Hence, the expectation is calculated as:

$$E_{t-1}\left[\frac{dr_t}{dw_{t-1}}\right] = \int_0^{x_{t-1}-w_{t-1}} 1f(s)ds = \frac{1}{r_{t-1}}(x_{t-1} - w_{t-1}) = W > 0 ,$$

where the last inequality comes from equation 2. Combining the equation above and equation 4, we finally have

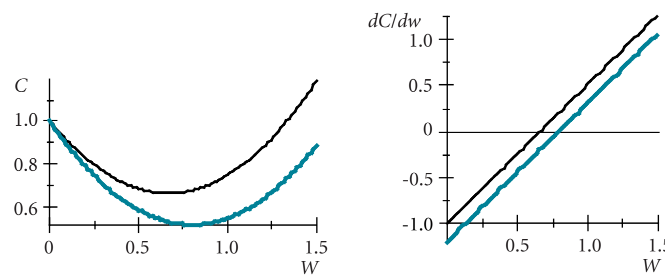
$$\frac{dC_t^*}{dw_{t-1}} = ZW < 0 .$$

As anticipated, the spillover is negative, corresponding to an abatement of costs. When the bank internalises these spillovers, these gains (perhaps discounted by a factor β) are added to the optimality condition (Opt), which therefore becomes

$$\frac{dC_{t-1}}{dw_{t-1}} + \beta\frac{dC_t^*}{dw_{t-1}} = \frac{1}{r}(r\lambda + Cw - Cx - cr - cw + cx) - ZW = 0 .$$

Hence, the BNSL shifts down the marginal cost schedule. This clearly brings about a higher level of liquidity posting in $t-1$ (in the example below, from 66 percent to 80 percent).

Total and marginal cost of w, with (blue) and without (black) a BNSL:



References

Armantier, O., J. Arnold, and J. McAndrews. 2008. "Changes in the Timing Distribution of Fedwire Funds Transfers." Federal Reserve Bank of New York Economic Policy Review 14, no. 2 (September): 83-112.

Bech, M. L., and R. Garratt. 2003. "The Intraday Liquidity Management Game." Journal of Economic Theory 109, no. 2 (April): 198-219.

Becher, C., S. Millard, and K. Soramäki. 2007. "The Network Topology of CHAPS Sterling Payments." Unpublished paper.

Beyeler, W., R. Glass, M. Bech, and K. Soramäki. 2006. "Congestion and Cascades in Payment Systems." Federal Reserve Bank of New York Staff Reports, no. 259, September.

Buckle, S., and E. Campbell. 2003. "Settlement Bank Behaviour and Throughput Rules in an RTGS Payment System with Collateralised Intraday Credit." Bank of England Working Paper no. 209.

James, K., and M. Willison. 2004. "Collateral Posting Decisions in CHAPS Sterling." Bank of England Financial Stability Review 17, December: 99-104.

Kobayakawa, S. 1997. "The Comparative Analysis of Settlement Systems." Centre for Economic Policy Research Discussion Paper Series, no. 1667, July.

Martin, A., and J. McAndrews. 2008. "Liquidity-Saving Mechanisms." Journal of Monetary Economics 55, no. 3 (April): 554-67.

McAndrews, J., and S. Rajan. 2000. "The Timing and Funding of Fedwire Funds Transfers." Federal Reserve Bank of New York Economic Policy Review 6, no. 2 (July): 17-32.

Mills, D. C., and T. D. Nesmith. 2008. "Risk and Concentration in Payment and Securities Settlement Systems." Journal of Monetary Economics 55, no. 3 (April): 542-53.

Soramäki, K., M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler. 2006. "The Topology of Interbank Payment Flows." Federal Reserve Bank of New York Staff Reports, no. 243, March.

Risk Management in Payments Systems

Articles by:

Michele Braun, James McAndrews, William Roberds, and Richard Sullivan

Antoine Martin and David C. Mills

Michele Braun, James McAndrews, William Roberds, and Richard Sullivan

Understanding Risk Management in Emerging Retail Payments

- The retail payment landscape is shifting increasingly from paper to electronic form as the number of ways to make noncash payments grows.
- Payment products, services, rules, and technologies are changing at a rapid rate—as are the tools for perpetrating fraud, illicit use, and breaches of data security.
- Providers of emerging payment methods now face the same risks as providers of more established methods; failure to control these risks can lead to rejection in the market.
- By limiting access to payment networks, monitoring for compliance with risk mitigation standards, and enforcing penalties for noncompliance, emerging as well as established providers can contain many of the risks associated with fraud, illicit use, and data security breaches.

1. Introduction

Electronic checks, cell phones, and speed-through lanes at toll booths are just a few examples of new payment methods recently introduced to the market. Based on computer technology, online commerce, and telecommunications, these new payment methods rely on electronics for most or all of their functions. Many products based on these methods have failed, some have struggled to grow, and a few have become well accepted in routine commerce. All face a variety of risks.

Reflecting these risks, news reports of data breaches, identity theft, and fraud have become a part of the electronic payment landscape. Novel characteristics associated with "emerging" payments include low-cost ways to store and transmit data. These technologies can reduce risk, but they can also lead to new risks. It is timely now to develop a structure and vocabulary for examining how new payment technologies affect risk, particularly as the number of ways to make noncash payments grows and as payments shift from paper-based to electronic form.[1]

Understanding the structure of risk is useful, although assessing losses and mitigation efforts in a new payment

Michele Braun is an officer and James McAndrews a senior vice president at the Federal Reserve Bank of New York; William Roberds is an economist and policy advisor at the Federal Reserve Bank of Atlanta; Richard Sullivan is a senior economist at the Federal Reserve Bank of Kansas City.
Correspondence: <james.mcandrews@ny.frb.org>

This article began as a Federal Reserve System staff research project. The authors acknowledge those who participated in that research, each contributing significantly to the concepts and substance of this article: Eugene Amromin, Peter Burns, Marianne Crowe, Jim Cunha, Dan Littman, Mark Manuszak, Antoine Martin, James McGrath, and John Yanish. They thank Christine Cumming and Claudia Swendseid for sponsoring and supporting the research and for valuable comments, as well as Stephanie Heller, Ariel Stern, Nathan Halmrast, and Daniel Meyer for their assistance. The views expressed are those of the authors and do not necessarily reflect the position of the Federal Reserve Banks of Atlanta, Kansas City, and New York, or the Federal Reserve System.

product can be difficult. Low levels of fraud losses, for example, could imply that: 1) risk is low, 2) current mitigation practices are effective, or 3) weaknesses have not yet been discovered. However, high levels of losses demonstrate that risks are high, and it takes time to know whether mitigation efforts can succeed. In either case, only time and the monitoring of problems will reveal whether risk can be controlled sufficiently. In this article, we consider whether, in this period of uncertainty, the sponsor of an emerging payment method has enough incentives and tools to control risk before the harm from fraud or operational problems becomes widespread.

Our analysis suggests that the sponsors and providers of successful emerging payment methods must be aware of potential fraud risk and operational risk. Moreover, they must mitigate these risks or face rejection in the payment market. Service providers can contain risks by limiting access to their payment networks, monitoring for compliance with risk mitigation standards, and enforcing penalties for noncompliance. While much of this containment activity is voluntary, some is enforced by public authorities that can help coordinate activities as well as define and enforce standards.

It is timely now to develop a structure and vocabulary for examining how new payment technologies affect risk, particularly as the number of ways to make noncash payments grows and as payments shift from paper-based to electronic form.

This article explores in several ways the structure and vocabulary of emerging payments system risks and their mitigation. We begin by recounting several incidents of fraud and losses associated with emerging payment methods. We then describe an economic framework for understanding risk control in retail payments. Next, we apply the framework to the risk experiences of three new payment types. These approaches—both deductive and inductive—are complementary ways to understand risk and its mitigation in emerging payment methods. Finally, we discuss some general observations derived from integrating the economic concepts and actual experiences, then offer conclusions.

[1] In 2003, the number of electronic payments exceeded the number of check payments for the first time. See Federal Reserve System (2004).

2. True Accounts of Fraud and Operational Risks in Payment Innovations

The following accounts illustrate fraud and operational problems that exploited the novel characteristics of new payment methods. These incidents include a telemarketing scheme, a complex online fraud, and two data security breaches. The crimes that underlie these incidents—fraud, con artistry, and theft of money, property, or someone's good name—are not themselves new. The operational problems are also not necessarily new, but the potential scale and speed of the disruptions are of a magnitude untypical of their paper-based counterparts.

2.1 Telemarketing Fraud

In 2003, the Federal Trade Commission (FTC) announced that it had closed down the Assail Telemarketing Network and its affiliates. The FTC alleged that the Assail companies ran telemarketing activities from so-called boiler-room operations that offered credit cards to consumers with poor credit records.[2] Under the guise of charging membership fees, these firms persuaded consumers to provide the bank and account information from their checks.[3] The telemarketers then used this information to create electronic debits to consumers' checking accounts as payment for the "membership" fees. These credit cards appear to have been rarely, if ever, delivered. The consumers found, however, that they had also been signed up for expensive and dubious products (so-called upsell programs) such as auto club memberships, the fees for which were directly charged to their bank accounts. When consumers called to complain, the companies used elaborate scripts to avoid repayment or cancellation of the membership. The FTC alleged that Assail and its principals engaged in deceptive marketing activities that totaled more than $100 million.[4]

The particular type of electronic transaction that Assail used, a debit through the automated clearinghouse (ACH), must be processed, collected, and paid through participating banks. These banks are supposed to monitor the companies for

[2] See Federal Trade Commission, "International Telemarketing Network Defendants Banned from Telemarketing," press release, January 24, 2005, available at <http://www.ftc.gov/opa/2005/01/assail.htm>, as well as other FTC press releases.

[3] Consumers provided the encoded information that runs across the lower edge of a check, which is also known as magnetic ink character recognition (MICR) information.

[4] ConsumerAffairs.com, "Bogus Credit Card Marketers Settle Federal Charges," January 26, 2005.

which they provide this ACH origination service. In this case, First Premier Bank admitted that it had failed to perform due diligence on the activities and legitimacy of its customers, but it then helped identify the telemarketers and supplied information to the investigative agencies. The bank later paid $200,000 to Iowa, South Dakota, and Minnesota as part of a wider settlement and agreed to engage vigorously in know-your-customer practices and ongoing monitoring of customer activity.[5]

Before the particular ACH transaction type used by Assail was introduced, this type of fraud was often perpetrated by creating a "remotely created check"—a check that contains a text legend in lieu of the payer's signature. This approach is still used to commit fraud, but it does not offer the speed and scale this fraudster achieved using automation.[6]

2.2 Transaction Fraud and Data Security Breach

The U.S. Department of Justice reported that, in 2000, two Russian men, Vasiliy Gorshkov and Alexey Ivanov, used unauthorized access to Internet service providers in the United States to misappropriate credit card, bank account, and other personal financial information from more than 50,000 individuals.[7] They allegedly hijacked computer networks and then used the compromised processors to commit fraud through PayPal and the online auction company eBay.

According to the Justice Department's press releases, the fraudsters developed elaborate programs to establish thousands of anonymous e-mail accounts at websites that, at the time, did not have the sophisticated tools required to distinguish human intervention at set-up. Gorshkov's programs created accounts at PayPal that were based on random identities and stolen credit card numbers. The programs then transferred funds from one account to another to generate cash and to pay for computer parts purchased from vendors in the United States. Additional computer programs allowed the conspirators to control and manipulate eBay auctions so that they could act as both seller and winning bidder in the same auction and then effectively pay themselves using the stolen credit cards.[8]

This was a case of fraudsters hacking into databases, stealing payment-related and other information, using the stolen identities to create fictitious accounts, manipulating online auctions, and using machine-based tools to proliferate their thefts and confound the transaction/audit trail.

Ultimately, the FBI used an undercover operation to lure the two hackers to Seattle, Washington, where they had been invited under the pretext of a job interview with "Invita," a fictitious computer security company. In October 2002, the two men were sentenced to three years in prison.

2.3 Unsecure Data

In 2005, the president and chief executive officer of CardSystems Solutions, Inc., a transaction processor, testified before a Congressional committee that, in September 2004, an unauthorized party had placed a clandestine computer program on the company's transaction processing system (Perry 2005). CardSystems reported that, on May 22, 2005, it suffered a "potential security incident." Records on 263,000 transactions were stolen—including account holders' names, account numbers, expiration dates, and security codes. Forty million records were potentially at risk.

CardSystems disclosed the breach to its bank as well as to MasterCard, Visa, and American Express. The three credit card companies determined that CardSystems had violated the credit card industry's prevailing security and data retention standards. Visa and American Express announced that they would not permit the firm to process their transactions after October 31, 2005. On October 15, Pay by Touch announced its acquisition of CardSystems Solutions because of the latter's network connections to 120,000 merchants, despite the demise of its card transaction processing business.[9]

More recently, in early 2007, the TJX Companies, which operate retail stores in the United States, Canada, Ireland, and the United Kingdom, reported that data security breaches from mid-2005 until late 2006 might have compromised more than 45 million customer records.[10] Company investigations also revealed breaches in 2003 and 2004, as well as compromised driver's license numbers and addresses. The Massachusetts Bankers Association reported fraudulent use of debit and credit cards issued by its members as a result of that breach. The

[5] This was the first time that the Federal Trade Commission tried to hold a bank responsible for the deceptive practices of its customer.

[6] To help reduce the potential for fraud in the use of remotely created checks, the Federal Reserve Board amended its Regulation CC effective on July 1, 2006, to create transfer and presentment warranties under which any bank that transfers or presents a remotely created check warrants that the check is authorized by the person on whose account the check is drawn. See Federal Reserve Board press release, November 21, 2005, available at <http://www.federalreserve.gov/boarddocs/press/bcreg/2005/20051121/>.

[7] U.S. Department of Justice, "Russian Computer Hacker Sentenced to Three Years in Prison," press release, October 4, 2002.

[8] Physor.com describes some of the techniques used by criminals to perpetrate fraud through online auction sites. See <http://www.physorg.com/news84545784.html>, December 5, 2006.

[9] Pay by Touch, "Pay by Touch to Acquire CardSystems Solutions, A Leading Provider of Integrated Payment Solutions," press release, October 15, 2005.

[10] TJX Companies, Inc., "The TJX Companies, Inc. Victimized by Computer Systems Intrusion; Provides Information to Help Protect Customers," press release, January 17, 2007, available at <http://home.businesswire.com/portal/site/tjx/>.

Association's press releases recounted that fraudulent card data had been used to make purchases in many U.S. states, Hong Kong, Sweden, and other countries.[11]

The *Wall Street Journal* reported that hackers first tapped into data transmissions from handheld equipment used to manage store inventory and prices.[12] Reportedly, they used these captured data to crack encryption codes and to steal employees' user names and passwords at company headquarters. With the resulting access to TJX's network, they stole credit and debit card numbers and even left messages for each other. Stolen card numbers were then allegedly sold on the Internet. Press reports traced losses to banks across the country. In addition to direct purchases with stolen credit and debit card numbers, the thieves or their customers also purchased prepaid cards, which were in turn used to purchase goods and services.

3. Definitions and Economic Insights

The examples just offered illustrate some risks of financial loss that are present in payment methods. We now turn to an economic examination of these risks and their mitigation, beginning with three general observations. First, the risks present when new or still-emerging payment methods are used are not wholly different from those present in long-established methods of payment. Nonetheless, our analysis suggests that certain risks are more salient in *emerging* retail payments than elsewhere in the payment marketplace.

Second, new payment methods are generally based on, or emerge from, existing payment products. To focus this discussion, we define *established* payments to include paper checks, recurring transactions transferred through the ACH, credit card and debit card transactions made with magnetic-stripe cards, and wire transfers. To this base, enhancements, innovations, and rules are added to address newly identified market opportunities or to take advantage of expanding technical capabilities. Sometimes innovations are sufficient to yield a distinguishably new payment method. Thus, we define *emerging* retail payments as those newly introduced payment methods that differ from established payments in a significant way—that is, technologically, contractually, legally, or conceptually.

Third, every payment method involves risk. The Bank for International Settlements' Committee on Payment and Settlement Systems identifies five major categories of risk associated with payment transactions: fraud, operational, legal, settlement, and systemic.[13] Generally, other types of risk are subcategories of these five broad types. Emerging payment methods may be particularly susceptible to fraud and operational risks. They may also carry enhanced legal risk simply because case law is less well developed or because the drafters of established laws and regulations may not have foreseen some of the ways in which payments are initiated, processed, and settled. Definitions of the three risks mainly associated with emerging payments are presented in the box.

A payment method may also carry risks not directly associated with the success or failure to transfer value. Instead, indirect problems may arise that appear ancillary to the financial transaction. For emerging retail payment methods, two risks of this type are notable: data security risk and risk of illicit use. In these cases, the payment methods function and transfer value correctly, but something underlying the transaction is "bad."

Data security risk is a form of operational risk involving unauthorized modification, destruction, or disclosure of data used in or to support transactions. For example, a data security breach may facilitate identity theft, which could trigger later harm to a party in a transaction or an otherwise uninvolved party elsewhere in the system.

The risks present when new or still-emerging payment methods are used are not wholly different from those present in long-established methods of payment.

Risk of illicit use is the risk that a payment method may be used for illegal purposes, for example, money laundering, terrorism financing, or the purchase of illegal goods and services such as drugs or child pornography. Similarly, the ease with which criminals can launder stolen funds or finance terrorists with legitimately earned funds affects not only the victims of the crimes that give rise to the "dirty" funds, but society as a whole.

[11] Massachusetts Bankers Association, "Massachusetts Banks Now Reporting That Fraud Has Occurred Due to the TJX Data Breach," press release, January 24, 2007, available at <http://www.massbankers.org/pdfs/TJXfraudNR.pdf>. Also see "Massachusetts, Connecticut Bankers Associations and the Maine Association of Community Banks and Individual Banks File Class Action Lawsuit Against TJX Companies Inc.," press release, April 24, 2007.

[12] Joseph Pereira, "Breaking the Code: How Credit-Card Data Went Out Wireless Door: Biggest Known Theft Came from Retailer with Old, Weak Security," *Wall Street Journal*, May 4, 2007.

[13] Bank for International Settlements (2000).

Major Risks in Emerging Payments

Type of Risk	Definition
Fraud	Risk of financial loss for one of the parties involved in a payment transaction arising from wrongful or criminal deception. The risk that a transaction cannot be properly completed because the payee does not have a legitimate claim on the payer.
Operational	Risk of financial loss due to various types of human or technical errors that disrupt the clearing and settlement of a payment transaction. The risk that a transaction cannot be properly completed due to a defective device or process that precludes the completion of all the steps required in a transaction.
Legal	Risk that arises if the rights and obligations of parties involved in a payment are subject to considerable uncertainty.

Source: Bank for International Settlements (2000).

3.1 Some Insights from Economic Theory

Risk Containment as a Good

Economic theory offers some useful concepts for understanding risk in payments systems. All payments systems are systems for managing valuable information: They keep records of transactions and communicate transaction data. Any information stored and transmitted by a payments system can be described as an economic *good*, an item having value in exchange.

Thanks to modern information technology, emerging payment methods can offer tremendous efficiency gains over traditional methods of making payments. Electronic data can be easily stored at a few locations and then shared among payments system participants at very low cost. Payment data thus meet Varian's (1998) description of a *digital good*, a good that can be stored and transferred in digital form.

Varian argues that digital goods are different from standard, physical goods (such as cornflakes, sneakers, and minivans) in that they are *nonrival goods.* A nonrival good is one whose value does not diminish with any one individual's use or consumption of it. A textbook example of a nonrival good is broadcast television: One's consumption of a TV show does not diminish the quantity available for consumption by another individual. Other examples of digital goods that are nonrival goods are recorded music, video, and computer software. The data managed by modern payments systems are another example of this type of good: The use of a credit card in one electronic transaction does not diminish the ability to

Any information stored and transmitted by a payments system can be described as an economic good, an item having value in exchange.

use it in another transaction so long as the credit limit is not exceeded. (Credit, cornflakes, and sneakers are not nonrival goods; they get used up.)

Central to the value of any digital good is data integrity—garbled music or video is useless, for example. The usefulness of payment data can be diminished by fraud and security breaches or by operational disruptions that make it difficult to transmit data. Consequently, we argue that the integrity of payment data is also a nonrival good. If a payments system participant secures a facility against operational disruptions and fraud, it creates an environment conducive to smooth operation of the payments system, generating benefits for other participants as well.

Nonrival goods are classified as *club goods* or *public goods* according to whether access to the good can be limited. A club good is a nonrival good that a group or individual can be stopped or excluded from consuming. For example, cable television firms exclude nonsubscribers from their service by encoding their signals and giving decoders only to paying subscribers. A public good is a nonrival good for which access cannot be limited. National defense, for example, is a nonrival public good because everyone in a country is covered and no one can be excluded from the benefits.

In the case of actions to contain fraud and operational risks in emerging payments, the club good description is perhaps the most appropriate. Successful private sector payment providers (for example, credit cards, debit cards, and ATM networks) have by and large managed to contain fraud.[14] They also maintain operating procedures and auditable controls to limit operational risk. Participation in these systems is limited by membership rules, and participants (individuals, merchants,

[14] Reported fraud rates for credit card transactions are about 5 basis points of value, and similar fraud rates are reported for checks (Nilson Report). Industry representatives report that actual rates may be a little higher (Green Sheet). Visa reports an operational "reliability rate of 99.999 percent" ("Securing Payments: Building Robust Global Commerce," 2005, available at <http://whitepapers.zdnet.com/whitepaper.aspx?&scname=Bank+Management&docid=152783>).

banks, and processors) associated with high levels of fraud or operational snafus can be expelled.

There are natural limits to the power of exclusion, however. Since every payments system is a type of communications network, excluding too many network participants lowers the network's value for those parties that remain. There will always be a trade-off between security and inclusiveness.

Why Containing Fraud and Operational Risks Is Difficult

Hirshleifer (1983) describes a model of a nonrival good that is particularly applicable to data integrity in electronic payments systems. He describes the problem of people living in a "polder," a low-lying patch of land protected from flooding by a system of dykes. Each resident of the polder is responsible for maintaining the portion of the dyke that abuts his or her property. The dyke clearly provides a nonrival communal good: flood protection for all residents of the polder.

In this example, the degree of flood protection provided depends exclusively on the height of the lowest portion of the dyke. In other words, the degree of protection will not be determined by the *total* flood-mitigation efforts of everyone living inside the dyke, but rather by the one resident who exerts the *least effort* in maintaining the dyke. The analogy with emerging payments is straightforward: The risk mitigation effort of each party in the particular payments system to maintain data integrity prevents fraudulent data from circulating in the system, and a commitment to operational excellence allows others in the system to complete their transactions effectively.

A data breach or operational disruption in one portion of a payments system can open the metaphorical floodgates to problems throughout the entire system.

There are obvious parallels between flood protection in Hirshleifer's polder model and the mitigation of fraud and operational risks in payments systems. The 2005 data breach at CardSystems Solutions, for example, resulted in problems not only for CardSystems, but also for numerous other users of card payments systems—cardholders, merchants, banks, and processors. A data breach or operational disruption in one portion of a payments system can open the metaphorical floodgates to problems throughout the entire system. The potential for rapid propagation of fraud and operational disruptions is the flip side of the efficiency of electronic payments.[15]

Varian (2004) points out some difficulties in the provision of such nonrival goods. Because the amount of mitigation depends crucially on the participant that exerts the least effort, and because different system participants have different amounts at stake, there is a significant risk that participatory incentives will not be uniform. Participants with a lot at stake—that is, those with high net benefits from more mitigation activity—will prefer a higher level of protection from the risk in question than those with lower net benefits are willing to support. However, because overall protection depends on the participant that exerts the least effort, the latter group determines the overall level of risk mitigation.

The problem of nonuniform risk management incentives crops up regularly in payment situations. Various stakeholders in payments systems will naturally prefer different levels of mitigation in the system. The longer the supply chain or the larger the network for a given payment technology, the greater is the potential for disagreement about the appropriate level of mitigation.

Many different providers of services are integral to the processing of electronic payments. These providers include encryption firms, processors that route transaction data, and Internet service providers, among others. However, because minimizing fraud and operational risks requires effort from all participants, some mechanism is needed to give all participants the right incentives to "maintain the dyke." Private contracts, laws, and regulations can each play a beneficial role in creating such incentives.

Confronting Fraud and Operational Risks

Despite the difficulties outlined above, experience has shown that all successful payments systems have learned to keep fraud and operational risks at fairly low levels. Competition among payments systems gives important incentives to service providers to mitigate these risks. Systems that fail to contain risks do not survive in the payment marketplace.[16]

Service providers have developed three broad approaches to managing various kinds of payment risk: pricing, insurance, and containment.

[15] For a formal exposition of this point, see Kahn and Roberds (2005).

[16] "Thinking Like a Criminal," *Arizona Republic,* August 24, 2006, recounts how an entity that tried to compete with PayPal in the mid-1990s was closed down by Visa because as many as "three out of five . . . transactions turned out to be fraudulent."

- *Pricing* means that a party that bears a risk is compensated appropriately. Pricing is extremely important in allocating credit risk—banks that issue credit cards charge higher prices, in the form of higher interest rates on borrowing and higher annual fees for cards, to subprime cardholders who they believe are less likely to pay their balances. Issuing banks willingly bear a high level of credit risk on these cards because the higher interest earned compensates for the greater risk taken.[17]
- *Insurance* is an agreement between two parties as to who will bear a loss when one occurs. Thus, for instance, a merchant that receives a credit card payment is insured against the risk that the cardholder will not be able to pay the balance.
- *Containment* is a catchall term for activities that tend to deter or suppress risk. In the case of fraud risk, examples include swiping a credit card through a card reader to verify that the card is valid and asking for extra identification.

For fraud risk in particular, the effectiveness of the pricing and insurance approaches is limited by factors known as *adverse selection* and *moral hazard*.[18]

Adverse selection refers to situations in which undesirable outcomes result from asymmetric information among various parties to a transaction. Pricing works best to offset risks that are known and can be quantified in advance. When the payee and payer are anonymous to each other, the payee cannot know if the payer poses a bad risk and is likely to make a fraudulent payment. Correspondingly, the payer cannot know if the payee is selling legitimate goods. Particularly when commerce is conducted remotely (for example, over the Internet or by telephone), adverse selection undermines incentives to play by the rules. "Bad actors" can optimize their own malign incentives, undermining the confidence of legitimate merchants and consumers.

Moral hazard describes the effect of insurance on the incentives and thus behavior of an insured party. The availability of insurance can lead to opportunistic behavior on the part of the insured at the expense of the insurer. For example, a merchant that accepts payments via cards branded by a major network like Visa, MasterCard, or American Express is insured against credit risk (and sometimes fraud risk) and consequently may not have an incentive to make sure that a payment is legitimate and within a cardholder's credit limit. The card networks and their issuing banks, which provide the insurance, contain the risks by imposing on merchants authorization and authentication procedures that create appropriate incentives and guard against fraudulent card use.

[17] About 4 percent of balances are never paid off.

[18] These problems generally plague information security; see Anderson and Moore (2006).

Of course, moral hazard can arise on the payer's side, too—for example, when the right of a credit card holder to dispute a transaction may tempt the cardholder to claim that fraud was committed when it was not. Authentication procedures, particularly the collection of signatures at the point of sale, are designed to contain this form of moral hazard.

> *Experience has shown that all successful payments systems have learned to keep fraud and operational risks at fairly low levels.*

Moral hazard can also lead to opportunistic behavior that magnifies operational risk in payments systems. A payment processor might fail to spend the resources to maintain sufficient backup facilities—in the case of, say, a natural disaster knocking out a key data center—because the negative consequences of failing to maintain backup data do not accrue fully to the processor, but rather to thousands of other individuals and businesses as well. Card networks impose backup and resiliency standards to offset the lack of private incentives and contain this particular risk.

But pricing and insurance alone are not sufficient risk management techniques: Credit card issuers do not seek out cardholders who are likely to commit fraud, then attempt to recover the costs through differentially higher fees or interest rates; ACH operators do not offer two fee schedules, one for reliable and another for unreliable originating banks; and providers of payment services are generally reluctant to give unknown buyers and sellers guarantees against loss.[19]

Containment Techniques

Containment of fraud and operational risks requires cooperation among payments system participants. All need to have incentives to undertake actions that will keep fraud and operational risks down to acceptable levels. These incentives can be provided by monitoring system participants and then imposing penalties for inadequate risk controls that can lead to significant losses or disruptions.

Monitoring is the foundation of containment: Checking on participants will reveal whether they are engaging in appropriate levels of risk mitigation. But monitoring is unlikely to be

[19] Provisions in the Federal Reserve's Regulations E and Z, which implement the Electronic Funds Transfer Act and the Truth in Lending Act, respectively, impose some insurance requirements.

effective without some system of penalties for noncompliance. Monetary fines serve as deterrents. Contracts and laws assign legal liability for failures, which can be costly if breached, while some regulations establish performance standards and impose penalties when they are not met. Varian's (2004) theoretical analysis of the polder model, described earlier, suggests that relatively severe penalties—beyond the economic cost of a security lapse—may be necessary to ensure compliance.

Limitations on liability mean that penalties cannot do the whole job. In cases of fraud, the party most deserving of punishment, the fraudster, is usually long gone by the time the fraud is discovered. Even in cases where liability for a fraud or operational incident can unambiguously be assigned to a known party, there may be no practical level of penalty that could cause the guilty party to internalize the consequences of its inadequate risk controls. Sometimes the threat of the ultimate penalty—exclusion—may be the most effective deterrent: Payments system participants that fail to maintain adequate operational standards or fraud controls may be barred or expelled from the system.

[There are] a variety of techniques—pricing, insurance, and containment—for creating incentives for participants in retail payment transactions to mitigate fraud and operational risks.

Thus, we have a variety of techniques—pricing, insurance, and containment—for creating incentives for participants in retail payment transactions to mitigate fraud and operational risks. Underlying structural aspects of many electronic retail payments—particularly their nonrival nature—and the concomitant ability to limit access to the payment networks make containment techniques especially useful for creating deterrence tools.

3.2 Special Concerns for Emerging Payments Systems

Any viable payments system must find ways to maintain the integrity of payment data, but certain concerns are unique to emerging payment methods.

First, there is a "newness factor." The novelty of emerging payment methods implies that various problems may not be anticipated and therefore adequate safeguards and procedures may not be in place to address them. Emerging methods face a learning curve when confronting these issues. As evidenced by their survival and success, established payment methods have devised ways to mitigate these risks. The key question regarding emerging payment methods is whether their providers have the incentives and means to overcome the risks that could otherwise hinder widespread adoption.

Competition gives important incentives to payment method providers to mitigate many of these risks. Users can choose from many payment methods, and their choices reflect the extent to which the methods best facilitate smooth, low-risk transactions. In competition with payment methods less susceptible to fraud or operational failures, providers of new payment methods have clear incentives to address those risks. Failure to do so jeopardizes a method's viability. As in other markets, competition among payment methods is an important mechanism to induce providers to address these problems.

New payment technologies can improve economic welfare by allowing diverse participants—consumers, merchants, banks, and nonbank service providers—to exchange payment data in ways not previously possible. The value of these technologies hinges, of course, on data integrity. Successful payments systems will find ways of coordinating the behavior of diverse parties to facilitate data exchanges that serve their mutual best interests.

Data Integrity and Privacy

Integrity of payment data is important not only as a safeguard against fraud and operational interruptions, but also for maintaining participants' privacy. Privacy issues have come to the fore in recent months. A group known as the Privacy Rights Clearinghouse reports that more than 165 million records have been compromised by data security leaks since February 2005.[20] Such data breaches create potential for fraud, identity theft, and general loss of privacy.

Similar to other aspects of payment data integrity, the maintenance of participants' privacy constitutes a nonrival good. By preserving the privacy of its legitimate participants, a payments system encourages widespread participation and enhances the value of the system to all users. But as discussed above, nonrivalness can make it difficult to reach agreement among payments system participants on the necessary level of privacy protection.

[20] See <http://www.privacyrights.org/> (accessed September 7, 2007).

Maintaining privacy is tricky because, by nature, it runs counter to the payment function: Every type of payment requires the exchange of some information, which under the wrong circumstances can be subject to misuse. For a consumer to use a credit card to buy something from a merchant, for example, he or she must give the credit card information to the merchant. The consumer's surrender of credit card information is essentially a compromise between the merchant's need to identify the consumer and the consumer's desire to remain anonymous to prevent misuse of his or her personal information. The merchant obtains enough information about the consumer to determine that the transaction is legitimate, but no more. Under some circumstances, maintaining privacy can conflict with the goal of preventing fraud, as Stigler (1980) points out. Moreover, Katz and Hermalin (2006) discuss efficiency reasons for privacy that suggest that the full sharing of private information within a payments system could be inefficient, even in the presence of fraud risk. Every successful payments system has to reach a workable compromise between these two facets of transaction privacy.

Illicit Use

Unlike many risks associated with payments systems, the use of a payment method for illicit purposes (such as money laundering, financing of terrorism or crime, or the purchase of illegal goods) rarely involves direct risk of financial loss to a participant in the payment transaction. Thus, unlike many operational risks, the use of a payment method for illegal activities does not pose a risk as such for other users of that payment method. In this case, the payment method works as designed, but individuals use the method for nefarious purposes external to the payments system itself. Rather than creating financial risk to direct participants in a transaction, illicit use introduces or carries broader societal risks. Since monetary gains are one determinant of the level of criminal activity, erecting obstacles and deterrents to these activities supports an important public good.

Unfortunately, many of the features that provide value for legitimate transactors can also make them susceptible to misuse by individuals engaging in money laundering and terrorism financing. Features that suggest the potential for a payment method to be used or misused for illicit purposes include speed of value transfer, transportability, intermediation, anonymity, quantity limits, network connectivity, and ease of interface. A common feature of many of these methods, especially electronic methods, is the speed with which value can be transferred. While the relative speed of the transactions is generally a desirable feature in the general market—it reduces certain types of fraud—it can also make it difficult to identify and preempt illicit transactions.

Similarly, some emerging payment methods involve highly transportable stores of value, either in physical or electronic form. Diverse participation and a high degree of privacy, both of which are features that make a payments system attractive for legitimate users, can make it easier to mask illicit use. Some emerging payment methods operate with little or no involvement of conventional financial intermediaries such as banks, making it difficult for authorities to monitor and identify illicit use. Network connectivity addresses the breadth of uses of a payment method and may alter its attractiveness as a store of value. The interfaces through which transactions are initiated may alter the ability to identify illicit transactions. In practice, it may be hard to distinguish between "user-friendly" and "illicit-user-friendly" platforms.

Rather than creating financial risk to direct participants in a [payment] transaction, illicit use introduces or carries broader societal risks.

Like other types of nonrival goods, payments systems can guard against illicit use through the use of monitoring and penalties (including criminal penalties) and through the exclusion of miscreants. But the high degree of similarity between the needs of legitimate and illegitimate users of payment technologies, as well as the need to balance societal costs and benefits, suggests that some amount of criminal use and other socially undesirable activity will always slip through. Society's determination of what constitutes an acceptable threshold of illicit use is a complex and thorny issue that goes beyond the scope of this article.

4. Three Examples of Risk and Its Management in Emerging Payments

We present three informal case studies to illustrate how characteristics of new payment methods affect potential risk, how key participants act to mitigate those risks, and how participants' actions demonstrate the economic principles described above. The three payment methods are: 1) general-purpose prepaid cards, 2) e-check payments through the ACH system, and 3) proprietary online balance-transfer systems such as PayPal. Each incorporates new technologies, new networks, and new rules to create an entirely new payment method. These examples are not intended to demonstrate the full range of payment options. They are used in different venues, employ different means for initiating payments, and clear and settle transactions differently; yet they employ similar risk mitigation strategies. (The appendix describes our selection of the case studies.)

While the payment methods in these case studies are not immune to all types of risk, we concentrate on fraud, operational, and illicit use risks because emerging methods appear particularly susceptible to these problems. The case studies focus on those areas of emerging payments that differ from established payment types. To the extent that an emerging payment is initiated using new technology but clears and settles through an established settlement network, our discussion examines the new front-end mechanism but excludes the clearing and settlement portion.

4.1 General-Purpose Prepaid Cards

General-purpose prepaid cards, branded by a payment network such as Visa, MasterCard, American Express, or Discover, can be used by all merchants that accept that network brand. Introduced in the 1990s, the cards function similarly to credit and debit cards at a point of sale: A customer swipes a plastic card through a standard reader, and the transaction is authorized and settled through a card network. In addition, some cards can be used to withdraw funds from ATMs and to make remote purchases or pay bills, similar to debit cards.[21] Cardholders can often check the balances available on their cards through a website or telephone response system.

General-purpose prepaid card programs differ in price, product functionality, customer identification requirements, value limits, and levels of cardholder protection. Their distinguishing characteristic is that they require cardholders to turn over funds in advance for future purchases of goods and services. Frequently, the funds on these cards can be reloaded at a variety of outlets, such as at merchants, over the Internet, or through ATMs. This feature allows a cardholder to use a single card without replacement or interruption, thus increasing the card's value as a potential substitute or complement to a formal banking relationship.[22]

[21] Payroll cards are a similar application, but differ dramatically in terms of the business model used for marketing and distribution and in terms of regulatory coverage. Payroll card programs are not discussed here.

Risk Analysis

The advance-payment feature substantially mitigates credit or nonpayment risk in general-purpose prepaid card products, allowing such cards to be marketed widely and distributed directly to consumers by nonbank third parties, referred to as card sponsors. Although every payment card must be issued by a bank, a nonbank sponsor's logo often appears as the most prominent brand name on the card. The broad involvement of nonbank institutions in the distribution of general-purpose prepaid cards stands in contrast to the common practices of traditional debit and credit card programs.

Since general-purpose prepaid cards use the credit and debit card infrastructure for transactions, clearing, and settlement, they share the risks inherent in these more mature financial products. These cards also exhibit a number of new risks, including a complex supply chain that often involves nonbank third parties at vulnerable stages of delivery and an increased susceptibility to money laundering and illicit transactions.

For general-purpose prepaid cards, nonbank institutions often stand between the cardholder and the bank that issues the card. In many cases, the nonbank institutions maintain the primary relationship with the cardholder. This prominent role for a third party in initiating and maintaining customer relationships can complicate the regulatory treatment of cards and introduce credit risk for the bank issuers and, potentially, the cardholders. The third-party entities could go bankrupt or be subject to various operational failings that would be less likely to impact accounts at a supervised and FDIC-insured financial institution. The involvement of the major card associations and the fraud detection that they bring to bear appear to deter illegal activity. News reports recount instances of fraud, however, such as using stolen credit cards to purchase prepaid cards at a self-serve checkout counter.[23]

Since general-purpose prepaid cards use the credit and debit card infrastructure for transactions, clearing, and settlement, they share the risks inherent in these more mature financial products.

[22] See McGrath (2007) for further discussion of the functionality and market position of prepaid cards.

[23] David Hench, "Savvy Thieves Use Gift-Card Scam to Fool the System," *Portland Press Herald/Maine Sunday Telegram*, February 21, 2007.

Additionally, third-party nonbanks may not have the same level of data security that banks have, potentially exposing consumer data to greater risk of theft. In particular, third-party distributors may fail to impose uniform data security standards for their retailers, a security lapse that increases risks for data gathered and stored at the point of sale.

A downside of the flexibility provided by cards able to facilitate nearly anonymous transactions is that they are attractive vehicles for abuse by illegal enterprises.[24] The Drug Enforcement Administration, Immigration and Customs Enforcement, and Internal Revenue Service–Criminal Investigation each allege that prepaid cards are used in bulk cash smuggling. They explain that drug dealers load cash onto prepaid cards and send them to their drug suppliers outside the country who use the cards to withdraw money from a local ATM.

This potential for illicit use is exacerbated if card issuers or sponsors operate offshore because it makes it harder to enforce relevant regulatory requirements. In fact, some general-purpose prepaid card products are openly marketed as a convenient way to circumvent law enforcement and tax authorities. For example, a prepaid card called the Freedom Card used to promise, among other things, "a fully anonymous ATM debit card . . . requiring no phone numbers or IDs . . . no daily cash withdrawal or loading limits . . . real-time card funding with any e-currency, PayPal, Western Union or bank wires."[25] The card was originated by an offshore financial institution, but it could be used to obtain funds throughout the world. This product appears not to exist any more. While such offerings are often short-lived, dubious new products emerge regularly.

Mitigation

Efforts are also under way to deal with the illicit use and data security risks. An industry task force says it is in the process of creating "AML [anti-money laundering] best practices guidelines" in response to anticipated regulations from the U.S. Treasury's Financial Crimes Enforcement Network aimed at thwarting money laundering and terrorist financing through prepaid cards.[26] The major card networks have issued guidelines to the issuing banks that are intended to reduce the attractiveness of prepaid cards for money laundering.[27] These include capping the stored dollar amount per card, limiting the frequency with which and the value of funds that can be reloaded, obtaining and confirming certain customer data prior to approving card applications, and providing liability protection for consumers in the event of card loss or fraudulent usage.

The operational and fraud risks of general-purpose prepaid cards are evidenced by: 1) a more complicated supply chain for providing the cards, often involving nonbank third parties in primary customer relationships, and 2) the potential for illicit transfer of funds. Domestic and international law enforcement officials are particularly interested in mitigating the latter risk.

The major card networks have issued guidelines to the issuing banks that are intended to reduce the attractiveness of prepaid cards for money laundering.

4.2 ACH e-Checks

Over the last decade, the National Automated Clearing House Association (NACHA), which sets rules for ACH transactions, has gradually developed rules and formats for six new electronic debit transactions, referred to as *e-checks*. These e-checks allow banks and their clients to convert checks or information from checks into ACH debits.[28] The following discussion describes two types of e-check transactions—accounts-receivable conversion (ARC) and telephone-initiated (TEL) transactions—to illustrate the risk factors and mitigation trade-offs associated with these new transaction types.

ARC rules permit businesses to transform checks mailed by bill-paying consumers into ACH debits. In the fourth quarter of 2006, the 613 million ARC transactions initiated accounted

[24] See Money Laundering Threat Assessment Working Group (2005), Financial Action Task Force (2006), and Sienkiewicz (2007).

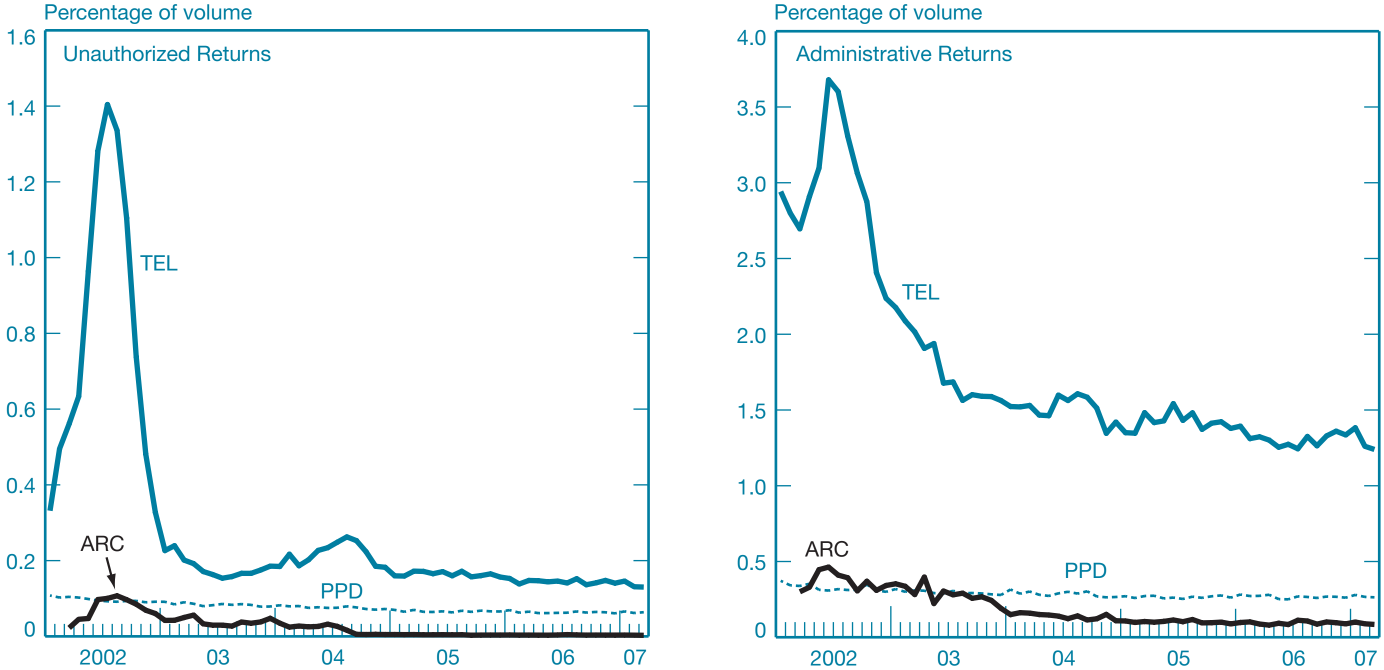
[25] See <http://www.freedom-cards.com> (accessed mid-2006).

[26] See <http://www.cardassociation.org> for information on Network Branded Prepaid Card Association efforts.

[27] As reported in Money Laundering Threat Assessment Working Group (2005).

[28] NACHA members include financial institutions and regional clearinghouse associations. NACHA manages the development, administration, and governance of the ACH system. Its rules provide more than fifteen worktype codes for different types of payments—such as corporate-to-corporate payments, recurring payments, point-of-sale payments, and e-check payments—as a means to identify specific rules, formats, and uses. Historically, the ACH has typically been used for direct deposit of payroll and Social Security payments and to collect recurring monthly mortgage, insurance, student loan, and business-to-business payments. For more information, see <http://www.nacha.org/About/default.htm>.

Return Rates for Selected Automated Clearinghouse (ACH) Applications, 2002-07



Source: Federal Reserve System.

Notes: Return rates are calculated from monthly Federal Reserve ACH data; ARC, PPD, and TEL are accounts-receivable conversion, preauthorized payment and deposit, and telephone-initiated transactions, respectively.

for more than half of e-check volume. Use grew by about one-third over the prior year: Roughly 6 percent of checks written are now being converted to electronic debits under ARC rules.[29]

TEL transactions are debits to consumers' accounts authorized by the account holder via telephone to a merchant, vendor, or service provider. These transactions make one-time ACH payments available when written authorizations are not feasible.[30] The 76 million TEL transactions processed during the fourth quarter of 2006 reflect a 16 percent increase over the previous year. TEL transactions account for about 7 percent of e-check volume.

ACH transactions that fail to clear (because of, for example, insufficient funds, errors in processing, or suspected fraud) are returned along with a code indicating the reason for the return. Return rates are useful indicators of risk because the standardized return reason codes indicate what type of problem caused the clearing failure. Typically, high levels of specific return codes indicate high levels of specific risks. For example, various processing-problem codes identify administrative returns that can indicate operational problems, whereas returns of transactions not authorized by an account holder (known as unauthorized transactions) can indicate fraud problems.

The chart presents historical rates for ARC and TEL for the types of returns most likely to suggest administrative or fraud problems. It also shows parallel return rates for preauthorized payments and deposits (called PPD), which is the most widely used recurring, non-e-check debit transaction and serves as a useful comparison.

There are similarities and differences in the records of ARC and TEL returns. In the fourth quarter of 2002, ARC return rates for both unauthorized and administrative returns were similar to those for PPD debits, but by the fourth quarter of 2006 they had fallen below the rates for PPD debits. In the fourth quarter of 2002, TEL return rates were some six to eight times the rates for PPD debits, and although they have since fallen, TEL return rates remain at least twice those for PPD debits. Overall, ARC and TEL each had high return rates in their introductory periods and the rates declined over time. ARC return rates today are very low, whereas those for TEL remain relatively high.

[29] Bank for International Settlements (2006, p. 157) reports that 33.1 billion checks were paid in 2005. ARC transactions are written as checks but paid as electronic debits.

[30] NACHA rules restrict TEL transactions to prevent their use for "cold-call" telemarketing, but they can be used when there is a preexisting relationship between merchant and consumer or when the consumer initiates the call.

Risk Analysis

ARC and TEL transactions share some risks with other ACH debit transactions, but differ in risks that are driven by the location at which the payments are initiated and the relationships among the parties to each transaction.[31] For ARC, retail lockbox processors convert checks sent by consumers to billers. Lockbox staff use high-speed equipment to capture coded information from the remittance slips and checks. The lockbox business is highly concentrated, mature, and controlled. In many cases, these processors operate as subcontractors to the originating banks, supporting the banks' cash management product offerings. In contrast, TEL transactions rely on customer input of account information via telephone, a context in which the data and customer's identity cannot easily be verified.

For ARC, the largest risk is operational. ARC rules initially made business checks ineligible for conversion to an ACH debit. Early problems included inadvertent conversion of business checks, particularly those that are the same size as consumer checks. Banks on which these checks were drawn often returned the transaction.[32] Another early problem was that processors could not properly match the ARC payment to the appropriate checking account: During a pilot program, one bank reported its associated administrative returns were as high as 10 percent.[33]

> *Inadequately researched bank relationships can undermine the "gatekeeper" function in [the telephone-initiated] payments system.*

For TEL transactions, fraud is a larger risk, perhaps augmented by operational risk caused by various participants in its supply chain. TEL is designed for ad hoc transactions between merchants and consumers, some of whom do not have a preexisting relationship. A long-standing business relationship is thus often absent, which increases the likelihood of either seller or buyer fraud. Adding to this risk is the fact that TEL opened the ACH network to new merchants and businesses, including some telemarketers and bill collectors, that may not have received sufficient scrutiny or monitoring from the banks through which they originate their transactions. Inadequately researched bank relationships can undermine the "gatekeeper" function in this payment method, making it difficult to deny dishonest originators access to the ACH network. Use of third-party service providers for TEL can compound the difficulty of identifying illegitimate initiators by adding an intermediary between the payment-originating bank and its ACH debit-originating clients.

Mitigation

Many of NACHA's rules and procedures aim to control and mitigate these risks. NACHA defines the rights and responsibilities of ACH participants, including originators (merchants, lockbox operators, and other businesses that initiate ACH payments) and originating banks (banks that provide ACH services to originators). Originators are required to follow NACHA rules and procedures when preparing and submitting ACH payments. Originating banks warrant certain aspects of ACH transactions and are financially liable for returned transactions. To help control this liability, originating banks typically use contract language to shift risk to originators. Originators are thus given financial incentives to correct and avoid processing problems.

When the problem arose of business checks being converted inappropriately, ARC originators reconfigured processing equipment to improve separation of business checks and worked to change NACHA rules to permit conversion of the business checks that were hardest to identify.[34] Originators also reduced administrative returns by building databases to match ARC payments and checking accounts.[35] NACHA requires a lockbox processor or its bank to keep check images for two years, but to destroy the physical check within fourteen days. Such measures help mitigate fraud risk and simultaneously decrease the risk of processing a check twice.

NACHA rules also require originating banks to gather sufficient information to understand the background and business of any new originator that may be given access to ACH services. This gatekeeping function generally keeps dishonest originators out of the ACH network, but it proved inadequate for TEL transactions. As illustrated by the Assail example

[31] Shared infrastructure and processes can contribute to risks in certain ACH payments. The ACH network does not use real-time authentication and authorization, and there are no centralized databases of originators accused of fraudulent use of the ACH system. These are mitigation techniques used by other payment networks.

[32] Daniel Wolfe, "Dealing with the Accidental Conversions," *American Banker*, December 8, 2004.

[33] Steve Bills, "Pilot Done, Wells to Widen Lockbox Conversion Effort," *American Banker*, October 18, 2002.

[34] Daniel Wolfe, "Dealing with the Accidental Conversions," *American Banker*, December 8, 2004. Note that effective September 15, 2006, business checks that do not carry an indicator in the auxiliary on-us field of the MICR line can be converted to ACH debits.

[35] Steve Bills, "Pilot Done, Wells to Widen Lockbox Conversion Effort," *American Banker*, October 18, 2002.

described earlier, telemarketing was one source of fraud that resulted in high return rates for TEL transactions.

As evidence of problems with TEL transactions mounted, NACHA intervened directly with originating banks and outside of its normal processes. NACHA and its member banks identified the specific TEL originators responsible for initiating many of the transactions that were subsequently returned, and these originators were shut down.[36]

Various participants in the ACH system have taken steps to improve the effectiveness of originating banks as gatekeepers. These steps include the introduction of NACHA rules requiring originating banks to screen and monitor originators and to execute appropriate contracts.[37] Additionally, in June 2003, NACHA instituted a monitoring process to flag originators with TEL returns exceeding 2.5 percent. Outside of the NACHA framework, ACH operators have introduced risk monitoring services and rule changes, and federal regulatory agencies increased the attention given to these transactions in their guidelines on controlling risk in retail payments.[38] These actions were followed by a rapid decline in returns, suggesting that monitoring, enforcing rules, and limiting access to the ACH network have been successful strategies for risk mitigation.

Although these mitigation efforts reduced TEL return rates, the return rates remain higher than NACHA would like. Thus, NACHA is pursuing additional proposals to make monitoring return items and resolving problems more effective. To address risk issues beyond those of TEL more broadly, NACHA also reorganized its risk management infrastructure, creating a Risk Management Advisory Group to help implement a new risk management framework.[39] Subsidiary work groups are attempting to address three areas of risk mitigation: 1) control of access to the ACH system, 2) the monitoring and control environment, and 3) enforcement activity. Additionally, to increase the visibility of risk management at ACH originating banks, the Office of the Comptroller of the Currency issued a guidance document in September 2006 requiring that key ACH statistics be reported to banks' boards of directors and senior officials.[40]

4.3 Proprietary Online Balance-Transfer Systems

Among the payment options that arose for Internet commerce are proprietary online schemes to transfer balances of funds between accounts. In this type of scheme, customers establish an account with a service provider, such as PayPal, and use e-mail messages to initiate payments.[41] If both parties to a payment have accounts with the same service provider, the service provider simply transfers monetary balances between their accounts. At PayPal, most customers are buyers and sellers (small businesses and individuals) involved in online transactions, usually at an auction site. The service is also used by small online companies and by individual customers who value the ability to transfer funds from person to person. Neteller, a similar service provider, is widely used for payments

Among the payment options that arose for Internet commerce are proprietary online schemes to transfer balances of funds between accounts.

to online gaming sites.[42] Other online person-to-person payment providers that follow a proprietary balance-transfer or similar model include GreenZap, StormPay, and eGold.[43] We call these providers *proprietary online balance-transfer systems.*

PayPal, the largest and most well-known online payment service provider, uses the proprietary online balance-transfer approach and intermediated almost $23 billion in transactions during the first half of 2007. PayPal is larger and more sophisticated than any of its competitors. eBay, the huge online auction business, acquired PayPal in 2002, and eBay

[36] Wells Fargo, "Waging War on ACH Fraud," <http://www.nacha.org/ACHNetwork/ACH_Quality/WellsFargo_DB.doc> (accessed January 12, 2007).

[37] This includes establishing limits on ACH transactions and on return items, conducting audits, and making ad hoc contact to verify that the originator has represented its business appropriately in terms of the products it is marketing, its financial strength, and so on.

[38] For information on Reserve Bank services, see <http://frbservices.org/Retail/fedachRisk.html>. For information on Electronic Payments Network services, see <http://www.epaynetwork.com/cms/services/processing/value/001477.php>. See also Federal Financial Institutions Examination Council (2004, pp. 43-4).

[39] See *NACHA Risk Management Newsletter* 2, no. 2, pp. 1-2 (2006).

[40] See OCC Bulletin no. 2006-39, "Automated Clearing House Activities: Risk Management Guidance," available at <http://www.occ.treas.gov/ftp/bulletin/2006-39.pdf>.

[41] This discussion is our interpretation based on information from public sources; it is not based on conversations with anyone at PayPal. For detailed descriptions of PayPal's processes, see Bradford, Davies, and Weiner (2003) and Kuttner and McAndrews (2001).

[42] Neteller describes "the online gaming industry" as its "main market" ("President and CEO's Report for the Six Month Period Ended 30 June 2006," available at <http://investors.neteller.com/neteller/upload/1NLRInterims2006releaseFINAL11sep062.pdf>). As of early 2007, Neteller, based in the United Kingdom, did not permit U.S.-based customers to make gambling payments. See <http://content.neteller.com/content/en/member_businessupdate.htm> (accessed February 2007).

[43] Companies that have tried but failed to provide online services for consumer-to-consumer and consumer-to-business payments include Citibank, Yahoo!, and eBay, with their respective products C2it, PayDirect, and BillPoint.

transactions currently generate almost 70 percent of PayPal's dollar volume.

These service providers act as agents by accepting deposits from customers and allowing money to be transferred from one in-house account to another. Although specific arrangements vary, in-house account balances typically are funded from a bank account by ACH transfer or a buyer's/sender's credit or signature debit card. Frequently, funds can be withdrawn by check or by co-branded debit/ATM card, transferred to the user's individual bank account by ACH credit, or used for future transfers within the network.

Risk Analysis

Although the volume of activity suggests that this type of payment meets a market demand for rapid online payments, it remains an emerging payment method accompanied by a variety of risks. Examples of fraud, operational, and illicit use risks include: 1) fraud associated with simple enrollment and anonymity; 2) operational errors and malicious attacks, such as "phishing" and "pharming";[44] 3) operational risk associated with technological complexity and a complex supply chain; and 4) susceptibility to illicit use. Specific rules, processes, controls, and screening capabilities vary across providers, yielding different levels of unmitigated risk and affecting the availability of mitigation options.

The core philosophy of the proprietary online balance-transfer model is to permit easy, quick entry and 24-hour availability. Under this system, an unknown, possibly anonymous, seller can be positioned to perpetrate fraud or simply fail to live up to his or her side of a transaction. Such a dishonest seller could take the money and not ship the product. The buyer would then have to try to recover funds under the rules of the payment provider's user agreement or protection policy. To be covered under PayPal's Buyer Protection Policy, the seller must enroll in the verification program and the buyer must comply with other eligibility requirements.[45] In contrast, the user agreement for GreenZap, a smaller but similar service provider, indicates that it is not liable for any purchases or services and does not issue refunds for a product or service if the seller does not fulfill on commitments. GreenZap also states that members send funds to third parties at their own risk.[46]

Online businesses are also vulnerable to the risk of outages, and businesses with high visibility seem to be most attractive to those seeking to disrupt services and overcome security features. The size of PayPal (about 133 million accounts as of year-end 2006) and the speed at which technical changes are made to support its growth have, indeed, led to some significant system downtime and made PayPal subject to hacker attacks.[47] In addition, in October 2004, a site redesign crippled some of its operations, leaving the website unavailable for two days and subject to intermittent outages for several days thereafter.[48] Moreover, PayPal and eBay were the top phishing targets in 2005, representing 62 percent of all attacks, according to Netcraft, a company that tracks and blocks phishing sites.[49]

Proprietary online balance-transfer payment methods depend on complex, multistep processes. For the user, the tasks are kept simple. Behind the scenes, however, many parties . . . are involved in completing a transaction.

Proprietary online balance-transfer payment methods depend on complex, multistep processes. For the user, the tasks are kept simple. Behind the scenes, however, many parties (including individuals, merchants, third-party service providers, the buyer's and seller's banks, and the ACH, debit card, and credit card networks) are involved in completing a transaction. As is the case generally in complex networks, the large number of digital "hands" and handoffs increases the difficulty of identifying and assessing risk severity and the exposures that can vary by user, channel, or product.

Intentional user anonymity makes these services susceptible to illicit use, such as money laundering or payments for illicit purposes. Only the service provider has information about user identities. While this structure protects users from fraud and

[44] Phishing employs social engineering and technical subterfuge to generate "spoofed" e-mails that appear to be from a legitimate company. It uses the company's logo and style to lead consumers to counterfeit websites designed to trick them into divulging private data such as account user names and passwords. In contrast, a pharming attack redirects visitors from a legitimate website to an unofficial location by exploiting technical and procedural security weaknesses that compromise the domain-name server.

[45] Ralph F. Wilson, "Assessing Criticism of PayPal," *Web Commerce Today*, March 15, 2002. Available at <http://www.wilsonweb.com/wct5/paypal_assess.htm>.

[46] GreenZap claimed 777,600 users at year-end 2006. See <http://www.greenzap.com/newz/Company_Update_Q3_Q4_2007.pdf>. See also the GreenZap User Agreement, available at <www.greenzap.com>.

[47] eBay, Inc., "eBay Inc. Announces Fourth Quarter and Full Year 2006 Financial Results," press release, January 24, 2007. PayPal does not disclose how many of the accounts are active or have been used recently.

[48] Jim Wagner, "PayPal Scrambling to Fix Site Glitch," *Internetnews.com*, October 13, 2004. Available at <http://www.internetnews.com/ec-news/article.php/3421031>.

[49] Sean Michael Kerner, "eBay, PayPal Rank High on Phish Lists," *Ecommerce*, January 6, 2006.

identity theft, it can also make it easier for users to transfer funds illegally because traditional enforcement authorities do not have identifying information. In addition, theft of identities outside of the network could provide criminals with sufficient information to set up false accounts that can be used for illegal funds transfers. Further, if a service provider allows international transfers, the payment process might be used to launder funds between domestic and offshore accounts.

Mitigation

As the leading service provider, PayPal has an incentive to invest in good risk management tools and oversight to protect its payment method. Its risk management, in turn, protects legitimate users and establishes standards for other online payment service providers. The following examples illustrate that PayPal, in conjunction with eBay, appears to have learned from its losses and risk exposures, creating systems, technologies, and rules that help control the risks that emerged in its early years. As a result of its efforts, PayPal says that its loss rate is four-tenths of 1 percent, well below that of the credit card industry.[50]

> *Online payment service providers have addressed the risk of illicit use and international exposure by placing limits on transfers and account balances for unverified accounts.*

To combat machine-based attacks, PayPal developed an account creation process that requires manual human input, which has blocked unmanned computer "bots" from opening accounts. It also created multiple levels of service, in which higher levels of account service require additional identity confirmation. The verified member program, for example, protects PayPal and creates a product it markets to customers. PayPal also retains the right to terminate service to any participant it suspects of not complying with its rules.[51]

In addition, PayPal developed background computer monitoring programs (named Igor and Ilya) to search for transaction patterns consistent with suspicious buyer or seller behavior. While these efforts have not totally eliminated fraud, they appear to have had some success: Statistics reported in the press show that merchants using PayPal have loss rates due to fraud that are noticeably below the e-commerce average.[52]

To prevent the risk of a data breach, PayPal says that it collects, encrypts, and stores sensitive customer information on servers not connected to the Internet. Additionally, to counter phishing and pharming, PayPal provides clear instructions on what to do if customers suspect they have received a phishing e-mail. When notified of a phishing attack, PayPal attempts to close down the perpetrator's site within twenty-four hours.[53]

PayPal limits its own risk inherent in its complex supply chain by specifying its own rights and responsibilities as well as those of its users in cases where errors, disruptions, or unauthorized transactions occur. The user agreement is complex, and it is updated as needed. Information on how PayPal establishes contracts or manages relationships with its suppliers is not publicly available.

Online payment service providers have addressed the risk of illicit use and international exposure by placing limits on transfers and account balances for unverified accounts.[54] PayPal relaxes these limits for its verified accounts, but the verification process exposes would-be criminals. As a result, PayPal may have become less useful for money laundering. It does appear possible, however, to launder large sums of money by sending small increments to many accounts using a mass-pay type of function.[55] To counter the above risks, eBay and PayPal have established a joint fraud investigation team to track down problem transactions and users. Moreover, within the context of its legal obligations, PayPal has a strong history of cooperating with law enforcement agencies.[56]

Ultimately, the proprietary online balance-transfer model is a self-contained, closed payment method, albeit one open to a wide range of potential participants. All payment account activity occurs within a single entity, which can make it easier for a service provider to internalize and control risks. By operating as a closed system, a service provider can manage

[50] *Computer World*, "Q&A: PayPal Fights Back Against Phishing," February 12, 2007.

[51] PayPal's user agreements can be accessed at <http://www.paypal.com/cgi-bin/webscr?cmd=p/gen/ua/ua-outside>.

[52] Paul Cox, "PayPal and FBI Team Up," *Wall Street Journal,* June 22, 2001; Ralph F. Wilson, "Assessing Criticism of PayPal," *Web Commerce Today*, March 15, 2002; Rob Garver, "eBay and Banking: Is PayPal a Serious Rival?" *American Banker*, November 15, 2005.

[53] Similarly, Neteller's annual report describes significant expansion in its fraud, security, and IT capabilities. See <http://investors.neteller.com/neteller/upload/1AR2005_0406.pdf>.

[54] See, for example, <http://www.PayPal.com>.

[55] The mass-pay feature allows PayPal Premier or business account holders to pay up to 10,000 recipients in varying amounts at one time.

[56] See Paul Cox, "PayPal and FBI Team Up," *Wall Street Journal,* June 22, 2001. Also see Dawn Kawamoto, "PayPal Charged with Breaking Patriot Act," CNET News.com, March 31, 2003.

fraud by denying or restricting access to users who do not meet its membership eligibility requirements or who fail to provide the required authentication. It can temporarily or permanently block users who do not comply with its rules or who are suspected of fraudulent or unauthorized activities. PayPal's experiences illustrate that a provider must aggressively battle new operational and fraud threats with vigilant monitoring of payment transactions.

5. Lessons Learned

The foregoing case studies offer many useful lessons for managing the problems that arise in emerging payment methods. Although each case is unique, there are common themes, which can be organized into three basic lessons.

5.1 Recognize the Problem

The very features that contribute to the efficiency of new payment forms—their scalability, speed, and relative anonymity—can also enable the rapid proliferation of various types of payment risk. As information moves more easily among payments system participants, more intensive management is needed to safeguard this data flow. Moreover, the more widespread and successful the system becomes, the bigger the potential for disruptions.

The incident reported earlier concerning two Russian men scamming PayPal offers a striking illustration of this principle. The perpetrators first breached the security arrangements at Internet service providers, gaining an initial cover of anonymity. They then used electronic means to create anonymous e-mail accounts, which in turn were used to create bogus accounts at PayPal. The speed and extent to which this was possible relied fundamentally on computers and the Internet.

To date, most innovative payment methods still have relatively low volumes of transactions. So even if risks are not well controlled, the overall risk of loss is limited.[57] Complacency, however, would be irresponsible. Significant flaws or fraud risks to ACH products have the potential to reach more institutions and individuals than most emerging payment products. And, as demonstrated by the Assail, CardSystems, and TJX incidents, even interruptions of low-value payments can result in large losses and disruption of business for many participants.

[57] The volume and value of e-money payment transactions in the United States are negligible. In contrast, ACH e-check transactions grew more than 40 percent last year, totaling about 2 billion transactions for the first half of 2006. See Bank for International Settlements (2006, Tables 7 and 8 and pp. 145-6).

5.2 Maintain a Perimeter

All legitimate payments system participants—consumers, merchants, banks, and other service providers—share a common interest in risk mitigation. The nonrival nature of risk mitigation means that all these participants operate behind the same common protective perimeter of security and reliability. Successful payment methods find ways to encourage an appropriate buy-in of all participants in terms of contributing to this shared resource. As the case studies illustrate, wrongdoers need to be kept outside this perimeter—even in the most inclusive payment methods.

PayPal offers a good illustration of this principle. A key aspect of PayPal's market positioning is its openness, inclusiveness, and ease of use. It claims that all anyone needs to participate in PayPal is an e-mail address. However, as PayPal has become more sophisticated and has placed increased value on avoiding fraud and operational losses, it has accordingly tightened its perimeter and imposed participation standards.

All legitimate payments system participants—consumers, merchants, banks, and other service providers—share a common interest in risk mitigation.

Today, PayPal screens each participant, requiring not only an e-mail address but also some identifying information as well as credit card, debit card, or bank account information (all of which can be independently verified) before a participant is permitted to send funds.

Telephone-initiated ACH transactions offer another example of adaptation to new risks. The highly decentralized nature of the ACH, in which debit transactions are created by a wide variety of entities, has facilitated a relatively high fraud rate in the case of TEL transactions. Recent and proposed changes to NACHA rules are meant to encourage buy-in from banks in controlling this problem. They do so by imposing monitoring of problematic originators and, under some proposed rules, penalties for violators. This process is necessarily more complicated than it is in a proprietary system such as PayPal, given the diverse composition of the ACH. Yet there now seems to be widespread acceptance of the idea that stringent rules—such as exclusion—are required to keep fraud rates down to manageable levels.[58]

Prepaid cards are something of an intermediate case, being neither purely proprietary like PayPal nor as decentralized as the ACH. On the one hand, anyone can purchase a prepaid card at a retail outlet and anyone, not necessarily the same individual, can make a purchase with the card at a participating retailer. On the other hand, the card association whose name is on the card (for example, MasterCard or Visa) screens issuing banks and binds them contractually to particular provisions. The card associations also screen the card-selling merchants for a variety of risks, including the effectiveness of their security to prevent large-scale theft. Finally, the card associations impose contractual and monitoring provisions on merchants that accept their cards.[59]

For prepaid cards, the card associations serve as enforcers to ensure the integrity of the network, for example, by minimizing operational and fraud risks.

For prepaid cards, the card associations serve as enforcers to ensure the integrity of the network, for example, by minimizing operational and fraud risks. Thus far, this control appears to be effective, even though the nominal issuers of general-purpose prepaid cards—merchants and various third parties—are neither typical card issuers nor regulated financial institutions. In a broader context, the aftermath of the May 2005 data breach at CardSystems illustrates the efficacy of such control:[60] Visa and American Express subsequently barred CardSystems from participating in their networks, forcing the firm out of business.[61]

The CardSystems case also highlights the difficulties posed by lengthening supply chains in the payment industry. Again, tensions can arise between efficiency and security. Specialization along the payment supply chain represents a source of efficiency, but the heavy involvement of nonbank or third-party participants means that the defensive perimeter for data integrity cannot be monitored by the banking system alone.

Historically, the role of third-party processors was limited to back-office services, such as lockboxes. In conjunction with emerging payment methods, some third-party entities have moved into the more prominent position of maintaining primary relationships with customers. Conversely, in some cases, banks have moved from maintaining primary relationships to becoming back-office service providers. This role reversal for bank and nonbank institutions has raised policy concerns and is a topic that warrants additional study.[62]

5.3 Trust the Marketplace—but Not Blindly

Producing a nonrival good is always a difficult and often a controversial business. Computer software, recorded music, and video, three common examples, are frequent objects of public controversy, regulation, and litigation. But somehow, the market finds innovative ways to provide these goods fairly—though rarely without growing pains along the way.

Electronic payment services also demonstrate both market-driven discipline and creativity, including for their security and reliability components. New payment products are immediately subjected to the forces of a market's "invisible hand," including ramifications of exposure to operational, fraud, and data security risks. As a result, operators are forced to learn about previously undetected operational problems. Outages of almost any sort can rapidly undermine user confidence in the reliability of a product, a particular service provider, or a new form of payment generally. New products also seem to attract the attention of fraudsters eager to exploit flaws before they are rectified. Only if a payment provider can address such problems quickly and effectively can it stay in business. Thus, for many of these risks, market mechanisms provide significant incentives for service providers to see that they are addressed promptly and thoroughly.

New products in their early stages repeatedly show patterns of operational or fraud problems and unmitigated risk, after which containment efforts follow. When PayPal faced fraud losses early on, it took steps to reduce those losses. It also implemented new authentication techniques and introduced innovative technology. PayPal continues to revise

[58] NACHA has recently approved a code of conduct that establishes standards of behavior and "*specifies NACHA's right to disassociate itself* from any organization that, in NACHA's opinion, fails to meet the standards and principles stated in the code" (emphasis added). See Elliott C. McEntee, "Open Letter," NACHA, April 13, 2006.

[59] See, for example, BankInfoSecurity.com, "Visa Takes Aim at Data Companies," August 8, 2006.

[60] See Perry (2005).

[61] CardSystems was purchased by Pay by Touch for its merchant network, according to a company press release dated October 15, 2005.

[62] Concerns about the role of nonbank third parties in the payments system have been raised, but they remain unresolved. See, for instance, Hoenig (2000) and Sullivan (2007).

its contracts and participation agreements to increase controls and limit risk.

Some providers of similar online payment services have failed, at least in part because of fraud losses, and others have run into trouble with law enforcement authorities over illicit payments.[63] The service providers that survive are those that are able to identify and mitigate losses quickly. When NACHA introduced the TEL product in 2001 and return rates began to soar, it took steps to identify the source of the problems. As a result, return rates fell to more acceptable levels. The WEB transaction, another recently created ACH e-check application useful for Internet transactions, had a return rate of 0.68 percent in 2002, but it fell to 0.08 percent in 2004.[64]

As payments systems grow and flourish, however, so too does the potential for disruption. Recent developments in the payment card industry provide an illustration. Card networks, historically quite vigilant in the protection of their data integrity, have nonetheless been subject to significant data breaches. Increasing volume and a more diffuse supply chain have posed new difficulties. The card networks have responded by putting more pressure on merchants to comply with data security standards, but this effort remains a work in progress.[65]

> *As payments systems grow and flourish . . . so too does the potential for disruption.*

The vitality of the market for payment services does not rule out a role for public policy. Well-designed regulations can help coordinate industry efforts and maintain industry standards. Laws and criminal penalties can serve as deterrents to activities such as fraud. In addition, the importance of confidence in the overall payments system—a public good—should not be underestimated. Policymakers will always have an interest in ensuring that disruptions in one method of payment, however unlikely, do not spill over into other segments of the payments system.

In contrast to other risks considered here, the steps needed to reduce the risk of illicit use are not always fully supported by general market incentives. The federal government and many states respond to this risk by enacting laws and regulations to prohibit the use of payment methods for such purposes and by creating incentives for payment providers to screen out prohibited transactions. A measured policy response again seems appropriate, as the risk of illicit use must be balanced against the costs of compliance.

6. Conclusion

Innovative payment mechanisms, such as the ones described in this article, are making transactions less expensive and easier, while opening new commercial venues for payment transactions. As with more established forms of payment, however, the ultimate success of these inventive arrangements will depend on their ability to control risk.

For retail payments, the predominant risks are operational, fraud, illicit use, and data security risks. Providers mitigate these risks through techniques such as pricing, insurance, and containment. In the growing market of electronic transactions, these techniques have shared value that does not decline with additional use and can be enhanced with additional contributions—in other words, they are nonrival.

This article examined three emerging payment methods to draw some lessons from their operation and markets. The payment methods explored here carry transactions that are relatively low in value, and, during their start-up phases, most had a small number of users. However, some ACH-based transactions quickly reached substantial volume levels. With low values and generally limited breadth, the payment methods do not currently pose systemic risks or demonstrate substantial policy gaps. We note, however, that the risks discussed here are not confined to emerging payments.

All payment processes have risks that must be controlled. Fraudsters seem especially drawn to new technologies, becoming early adopters in their attempts to exploit any identifiable weaknesses. But fraudsters can also perpetrate innovative attacks against established systems. Moreover, even low-value retail payment providers can be the targets of machine-based attacks that can cause substantial damage; the speed of corruption and potential for proliferation of damaging problems are certainly shared by all payment methods that use electronic and networked technologies.

An important lesson to be taken from this study of emerging payment methods and their risks is that the products, services, rules, and technologies are all changing—and doing so at what appears to be an accelerating rate. So, too, are the tools for perpetrating fraud and data breaches as well as the techniques for mitigating them. This study provides a new structure for

[63] See, for example, Neteller Lawrence Complaint: *United States of America v. Stephen Eric Lawrence*, Southern District of New York. January 16, 2007, available at <http://www.casinocitytimes.com/news/article.cfm?contentID=163591>. Also see Neteller Lefebvre Complaint: *United States of America v. John David Lefebvre*, Southern District of New York, January 16, 2007, available at <http://www.casinocitytimes.com/news/article.cfm?contentID=163594>.

[64] Furst and Nolle (2005, p. 37).

[65] Robin Sidel, "Credit Firms Push to Thwart Fraud: Merchants Face a Penalty If Steps Aren't Taken to Curb Identity Theft; Visa Misses Own Security Deadline," *Wall Street Journal*, September 25, 2006.

considering risk and mitigation strategies that can be used to analyze new as well as established payment methods.

Our analysis of the risk mitigation techniques used by payments system providers concludes that containment is the dominant means of controlling risk. Generally, market mechanisms appear to encourage providers to mitigate risks appropriately: Most private-sector providers have the tools to manage many of these risks, particularly because they treat the integrity of the network as a club good; in other words, they retain the option to exclude any party that fails to comply with the network's safeguards. The applicability of this approach to the risk of illicit use, however, appears less certain.

More cooperative, open systems, which derive some of their utility from their universality, have less ability to exclude particular users and thus face greater risk mitigation challenges. Nonetheless, the problems, risks, and gaps in processes can be addressed only if the providers and participants remain vigilant while applying the lessons we described.

Appendix: How the Case Studies Were Selected

In general, new payment methods are built on top of existing products. Enhancements, significant innovations, and various levels of rule changes are added to these products to address newly identified market opportunities or to take advantage of expanding technological capabilities. To identify the extent to which a payment method is new rather than more established, we grouped the components of a generic payment process into two broad categories: the *access channel* and the *payment method*.[a] An access channel is used at the beginning of the transaction process; it provides the user interface or front end (for example, a plastic card with a magnetic stripe) and may or may not include verification of the identity of the involved parties and validation of the payment instrument. The payment method includes the remaining parts of the payment process governed by applicable laws, regulations, and contracts.

These various factors—new versus established components of access channels and payment methods—can be organized into a 2 x 2 matrix, as shown in Exhibit 1. Payment methods that have the fewest changes from established methods fit into the upper-left quadrant, although rule changes or new combinations of established characteristics can yield a new payment method. The lower-right quadrant includes emerging payment methods that incorporate the greatest number of new characteristics in terms of both access channels and payment methods. The remaining two quadrants, upper right and lower left, are hybrids of new and established components.

Exhibit 2 provides examples of payments that might be found in each of these four cells. For the case studies, we selected one payment method from each quadrant (shaded).

- ACH payments initiated via telephone (TEL) fall in the upper-left quadrant, since neither the telephone access channel nor the ACH clearing and settlement portions are new.
- General-purpose prepaid cards use established card-swipe technology to create a new payment and therefore fall in the upper-right quadrant.
- Accounts-receivable conversion (ARC) uses the new access channel of scanning technology and software to read paper checks and create transactions that flow over the established ACH network, as represented in the lower-left quadrant.
- Proprietary balance-transfer systems meld a new access technology—the Internet—with new transaction methods—e-mail and balance transfers—and therefore fall into the lower-right quadrant.

The TEL and ARC payments are types of ACH e-checks that share a clearing and settlement network and many rules (these are addressed jointly in the analysis above).

---

[a] See Bank for International Settlements (2000) for a description of the components of payment processes.

Exhibit 1

Access Channels and Payment Methods

Access Channels (Front End)	Noncash Payment Methods (Transaction, Clearing, and Settlement Processes): Established	Noncash Payment Methods (Transaction, Clearing, and Settlement Processes): New
Established	Well-known technologies initiate commonly known types of payments or use new rules to create new types of payments.	Existing access technologies initiate a new type of payment.
New	New technologies or networks access established payment method.	New technologies and networks initiate new types of payments.

Exhibit 2

Examples of Payments

Access Channels (Front End)		Noncash Payment Methods (Transaction, Clearing, and Settlement Processes): Established	Noncash Payment Methods (Transaction, Clearing, and Settlement Processes): New
Established	Established	Credit card payments Paper checks in general PPD ACH debit PPD ACH credit PIN debit Signature debit	General-purpose prepaid cards Closed network and gift cards
Established	New combinations of established components	ACH debit card ACH TEL Micropayment aggregators Check 21 substitute check	
New		ACH POP ACH ARC ACH WEB Check image presentation (not IRD) Cell-phone payment Highway toll booths Contactless card payments (debit or credit) Charge to phone bill Biometic authentication License ID to create ACH debit ACH credit push	Proprietary balance transfer Secured proprietary balance transfer Proprietary balance transfer via cell phone Prepaid wallet Instant credit

References

Anderson, R., and T. Moore. 2006. "The Economics of Information Security." Science 314, no. 5799 (October 27): 610-3.

Bank for International Settlements. 2000. "Clearing and Settlement Arrangements for Retail Payments in Selected Countries." Basel, Switzerland.

———. 2006. "Statistics on Payment and Settlement Systems in Selected Countries." Preliminary version, November. Basel, Switzerland.

Bradford, T., M. Davies, and S. E. Weiner. 2003. Nonbanks in the Payments System. Federal Reserve Bank of Kansas City. Available at <http://www.kansascityfed.org/publicat/psr/BksJournArticles/NonBankPaper.pdf>.

Federal Financial Institutions Examination Council. 2004. Retail Payments System: IT Examination Handbook. March. Available at <http://www.ffiec.gov/ffiecinfobase/booklets/Retail/retail.pdf>.

Federal Reserve System. 2004. "The 2004 Federal Reserve Payments Study: Analysis of Noncash Payments Trends in the United States: 2000-2003." Available at <http://www.frbservices.org/Retail/pdf/2004PaymentResearchReport.pdf >.

Financial Action Task Force. 2006. "Report on New Payment Methods." October 13. Available at <http://www.fatf-gafi.org/dataoecd/30/47/37627240.pdf>.

Furletti, M., and S. Smith. 2005. "The Laws, Regulations, and Industry Practices That Protect Consumers Who Use Electronic Payment Systems: ACH E-Checks and Prepaid Cards." Federal Reserve Bank of Philadelphia, Payment Cards Center Discussion Paper no. 05-04, March. Available at <http://www.philadelphiafed.org/pcc/papers/2005/ConsumerProtection.pdf>.

Furst, K., and D. Nolle. 2005. "What's Your Risk with the Growing Use of ACH Payments?" Office of the Comptroller of the Currency Quarterly Journal 24, no. 4 (December): 21-43.

Hirshleifer, J. 1983. "From Weakest Link to Best Shot: The Voluntary Provision of Public Goods." Public Choice 41, no. 3 (January): 371-86.

Hoenig, T. M. 2000. "Payments and Settlement Systems: Future Players and Issues." Remarks delivered at the BAI Money Transfer 2000 Conference, Chicago, Illinois, November 9. Available at <http://www.kansascityfed.org/SPCH&BIO/chicago.htm>.

Kahn, C. M., and W. Roberds. 2005. "Credit and Identity Theft." Federal Reserve Bank of Atlanta Working Paper no. 2005-19, August.

Katz, M., and B. E. Hermalin. 2006. "Privacy, Property Rights, and Efficiency: The Economics of Privacy as Secrecy." Quantitative Marketing and Economics 4, no. 3 (September): 209-39.

Kuttner, K., and J. McAndrews. 2001. "Personal On-line Payments." Federal Reserve Bank of New York Economic Policy Review 7, no. 3 (December): 35-50. Available at <http://www.newyorkfed.org/research/epr/01v07n3/0112kutt.html>.

McGrath, J. C. 2007. "General Use Prepaid Cards: The Path to Gaining Mainstream Acceptance." Federal Reserve Bank of Philadelphia, Payment Cards Center Discussion Paper no. 07-03, March. Available at <http://www.philadelphiafed.org/pcc/papers/index.cfm>.

Money Laundering Threat Assessment Working Group. 2005. "U.S. Money Laundering Threat Assessment." Report Prepared by an Inter-Agency Working Group and Issued by the United States Treasury Department, December. Available at <http://www.ustreas.gov/press/releases/reports/js3077_01112005_MLTA.pdf>.

Perry, J. M. 2005. Statement of John M. Perry, President and CEO, CardSystems Solutions, Inc., Before the U.S. House of Representatives, Subcommittee on Oversight and Investigations of the Committee of Financial Services. July 21.

Sienkiewicz, S. 2007. "Prepaid Cards: Vulnerable to Money Laundering?" Federal Reserve Bank of Philadelphia, Payment Cards Center Discussion Paper no. 07-02, February. Available at <http://www.philadelphiafed.org/pcc/papers/index.cfm>.

Stigler, G. J. 1980. "An Introduction to Privacy in Economics and Politics." Journal of Legal Studies 9, no. 4 (December): 623-44.

Sullivan, R. J. 2007. "Risk Management and Nonbank Participation in the U.S. Retail Payments System." Federal Reserve Bank of Kansas City Economic Review 92, no. 2 (second quarter): 5-40. Available at <http://www.kansascityfed.org/publicat/econrev/PDF/2q07sull.pdf>.

Varian, H. R. 1998. "Markets for Information Goods." Remarks prepared for a Bank of Japan conference, June 18 and 19. Available at <http://www.sims.berkeley.edu/~hal/Papers/japan/>.

———. 2004. "System Reliability and Free Riding." Available at <http://www.ischool.berkeley.edu/~hal/people/hal/papers.html>.

Antoine Martin and David C. Mills

An Economic Perspective on the Enforcement of Credit Arrangements: The Case of Daylight Overdrafts in Fedwire

- The Federal Reserve's extension of daylight overdrafts to banks exposes the central bank to some credit risk during the day.
- The Fed manages this exposure through a combination of tools, including monitoring, an awareness of banks' reputations, and collateral requirements.
- Under a proposed policy change, the Fed would supply intraday balances to healthy banks through collateralized and uncollateralized overdrafts; banks would be allowed to pledge collateral voluntarily to support intraday overdrafts.
- An analysis of the increased use of collateral resulting from the change points to potential benefits—as well as costs—for the Federal Reserve, banks, and the financial system.

1. Introduction

Credit arrangements between a borrower and a lender are a prevalent part of the economy. A fundamental concern for any lender is the risk that the borrower fails to fully repay the loan as expected, a type of risk called *credit risk.* Thus, lenders want credit arrangements that are designed to compensate them for—and help them effectively manage—credit risk.

In certain situations, central banks engage in credit arrangements as lenders to banks. For example, the Federal Reserve offers certain banks overnight loans at the discount window. Additionally, it provides liquidity to many banks during the day whenever those banks must overdraw on their Federal Reserve accounts in order to make payments and settle securities. This extension of *daylight overdrafts* by the Fed can be interpreted as very-short-term credit, so the central bank is exposed to credit risk that it must manage.

This article discusses how the Federal Reserve manages its credit risk exposure from daylight overdrafts. We first present a simple economic framework for thinking about the causes of credit risk and the possible tools that lenders have to help them

Antoine Martin is a research officer at the Federal Reserve Bank of New York; David C. Mills is an economist at the Board of Governors of the Federal Reserve System.
<antoine.martin@ny.frb.org>
<david.c.mills@frb.gov>

The views expressed are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve System, the Federal Reserve Bank of New York, or the Federal Reserve System.

manage it. We then apply this framework to the Federal Reserve's Payments System Risk Policy, which uses a variety of tools to manage credit risk. Finally, we discuss a possible increase in the use of collateral as a credit risk management tool, as presented in a recent policy proposal published by the Board of Governors of the Federal Reserve System (hereafter, the Board) that considers changes to its Payments System Risk Policy (Board of Governors of the Federal Reserve System 2008).

2. A Framework for Thinking about Contractual Relationships

Economists have developed a framework for thinking about contracts in general and credit arrangements in particular. We now summarize and illustrate the main elements of this framework. The emphasis is on first principles, an approach that provides a helpful basis for policy analysis.

2.1 Bad Luck versus Opportunistic Behavior

A borrower may not fully repay a lender for one of two reasons: bad luck or opportunistic behavior. By "luck," we mean all random factors that affect borrowers' and lenders' actions and that are independent of their behavior. For example, weather is a random factor that can influence a farmer's yield of corn independent of the amount of effort the farmer exerts. The effect of luck can typically be priced into a contract.

In contrast, opportunistic behavior—a privately beneficial action that increases costs to the other party in the transaction—typically cannot be priced into a contract. Opportunistic behavior occurs when borrowers may not have sufficient incentive to do all they can to repay their debts. In the example of the farmer, the lender wants the farmer to put forth great effort to yield a large amount of corn and would like to be assured that the farmer will do so. The farmer is opportunistic if he does not work very hard in the field. By not working very hard, he may not yield enough corn to fully repay his debt to the lender.

Why do borrowers have an incentive to engage in opportunistic behavior? At the time a credit arrangement is made, all borrowers promise to repay their debt. Otherwise, lenders would refuse to lend. Once the loan is made, however, borrowers have an incentive to renege on their promise and default. The economic decision of the borrower is *time inconsistent*. In other words, the best decision at a given time (the promise made at the beginning of the credit arrangement to repay the loan) may no longer be optimal later because of the consequences of the original decision (once the loan is obtained because of the promise to repay, the borrower no longer wants to repay it). Anticipating this outcome, the lender may choose to forgo making the loan in the first place.

2.2 Enforcement

To achieve a good outcome, borrowers would like to be able to *credibly commit* to not reneging on their promise. A strong enough commitment can sufficiently address the time-inconsistency problem.[1] Experience shows, however, that this kind of commitment is difficult to make. Institutions, formal or informal, that help economic agents make credible commitments are said to provide *enforcement.* Courts are an example of such institutions, but many other examples exist.

If enforcement were costless to lenders, they could adequately control opportunistic behavior and it would not affect the decision to lend and the determination of the interest rate to charge. Lenders would typically charge an interest rate sufficient for them to cover the risk of bad luck in the credit relationship.

2.3 Information Problems

Enforcement is rarely, if ever, costless. In particular, a lender may have inadequate information about some actions or traits of the borrower. Economists distinguish between two types of information problems: *moral hazard* (or hidden actions) of the borrower and *adverse selection* (or hidden types) of borrowers.[2] Consider this example of moral hazard. It is well understood that if a bicycle is insured against theft, its owner is less likely to protect it as carefully as if it were not insured. The hidden action here is how carefully the owner protects the bicycle. Since the insurance company is unable to observe this action, it cannot make the insurance contract dependent on it.

An example of the adverse selection problem is found in the health insurance industry. The hidden type here is an

[1] A classical example of solving a time-inconsistency problem with credible commitment is found in the Greek myth of Ulysses and the sirens. Ulysses would like to hear the song of the sirens, but knows that once he does, he will be compelled to change the course of his ship and crash it against the rocks on which the sirens are standing. To enforce his commitment to not change course, Ulysses asks his sailors to bind him to the mast of the ship and to put wax in their ears. The wax will prevent the sailors from hearing the sirens and from hearing Ulysses when he asks his sailors to change course. Such elegant solutions to an enforcement problem are, unfortunately, not always available.

[2] Moral hazard and adverse selection are terms from the insurance markets, where these problems were first studied.

individual's health. Someone who seeks health insurance knows his or her health better than the insurance company does. Individuals who believe they are likely to need a lot of medical attention, for instance, will want to choose insurance with better coverage.

In the context of a lending relationship, the lender may not be able to observe what the borrower does with the loaned funds. In the farming example, will the farmer buy equipment that will allow for a greater yield of corn, increasing the likelihood that the loan will be repaid? Or will the farmer instead buy a big-screen television, leading to lower effort and making it less likely that the loan will be repaid? Similarly, the lender may not be able to observe the borrower's type. A lender will be more reluctant to lend to a farmer who has previously defaulted on other loans, which would suggest that this borrower is not a good type compared with a farmer who has never defaulted.

Borrower actions or types may be hidden from the lender, in which case they are called *unobservable.* Alternatively, they may be observable to the lender but hidden to parties outside the lending relationship, such as courts of law, in which case they are called *unverifiable.* In the previous example, the farmer's use of fertilizer is observable and verifiable if the lender is able to ascertain whether fertilizer was used and if, in addition, a court is able to establish that fact. However, the quality of the fertilizer used may be observable but not verifiable if, for example, the lender can analyze the fertilizer but cannot prove to a court, or any other third party, that the farmer used a particular fertilizer of a certain quality. Finally, whether the farmer used the correct amount of fertilizer may be unobservable and unverifiable because neither the lender nor a third party can determine how much fertilizer was used.

> *Enforcement is rarely, if ever, costless. In particular, a lender may have inadequate information about some actions or traits of the borrower. Economists distinguish between two types of information problems:* moral hazard *(or hidden actions) of the borrower and* adverse selection *(or hidden types) of borrowers.*

Information that is either unobservable or unverifiable is typically called *private*—as opposed to *public* information, which is both observable and verifiable. Economists therefore classify contractual situations as either of two types: 1) those in which perfect enforcement is possible because all relevant information is publicly available and 2) those in which there is only imperfect or costly enforcement because at least some relevant information is private.

2.4 Enforcing Contracts through Reputation, Monitoring, and Collateral

The information frictions described above can create credit risk over and above the risk that might come only from bad luck. Because of information frictions, a more sophisticated policy than simply charging an interest rate for a loan might be necessary. As we observed, when credit risk arises only from bad luck, no additional policy is necessary because nothing can be done to affect the probability that the loan will be repaid.

In principle, there are several ways to alleviate enforcement problems, and each method is costly. Because enforcement is not perfect, a trade-off always exists between better enforcement of contract terms and more costly means of ensuring enforcement. Among the ways of enforcing contracts in situations of imperfect information are *reputation, monitoring*, and *collateral.* We consider each of them in turn and provide an example of how a loan for a construction project uses all three.

Reputation

In cases of repeated interactions, the terms of a contract can depend on past actions. Borrowers can obtain better terms by establishing a reputation for good behavior. Reputation is achieved by showing a willingness to refrain from short-term opportunism. Reputation can be thought of as a way to make private information about one's type more public. In particular, it signals to a potential lender that a borrower is more interested in long-term outcomes (possibly because he or she wants to avoid punishments that restrict access to future loans) than any short-term gains achieved by defaulting on a loan. Reputation, therefore, can typically alleviate problems associated with adverse selection. In the case of a construction loan, a building contractor who wants to finance a new project may rely on reputation in negotiating terms for a new loan. A solid credit history increases the contractor's chances of securing a new loan and allows him to negotiate favorable terms.

Monitoring

Lenders can prevent opportunism by closely monitoring borrowers' actions, by screening and certifying their quality and that of their project, or, after a default, by verifying the quality and amount of their assets and operations. Monitoring can be thought of as a way to acquire information that would otherwise be private. As a result, monitoring can typically help alleviate the incentive issues associated with both moral hazard and adverse selection. In monitoring construction loans, for example, lenders conduct periodic inspections and require status reports from the contractor or independent third parties as a way to keep track of the project's progress.

Collateral

By posting collateral, the borrower offers a type of guarantee to the lender. Collateral may be something that has value to the lender so that the lender is at least partially compensated in case of default. In that particular case, the collateral plays an insurance role and need not have any value to the borrower. Collateral may also be something that has value to the borrower so that its loss punishes the borrower in case of default. In that case, the collateral plays an incentive role and need not have any value to the lender. In practice, collateral typically plays both roles in that it usually has some value to both the borrower and the lender. It is the incentive role that is most important from the perspective of reducing information frictions. Thus, collateral typically helps alleviate the incentive problems associated with moral hazard.

The effect of improvements in information technology has likely reduced the costs of reputation, monitoring, and collateral, making these tools more effective at reducing information problems.

Various assets can be pledged as collateral. For example, loans for such durable goods as houses, cars, and boats are often secured by the goods themselves. In the financial sector, securities and other financial assets can be used as collateral for various types of loans. In our construction loan example, once a project is complete and a building is ready for sale, the contractor can convert the loan into a standard mortgage, which requires that the new building be pledged as collateral. This conversion can provide the borrower with more favorable terms, such as a lower interest rate.

Depending on the circumstances, some of these ways to alleviate enforcement problems may be more or less costly or efficient. Reputations may be costly or impossible to maintain if there are not enough opportunities to signal one's type—for example, if relationships are short lived or if the economic environment evolves quickly and in unpredictable ways. Monitoring can be difficult or costly because it may require very specific and technical knowledge or because it may be possible to misrepresent the true state of a project. The use of collateral, too, is not without cost; there are costs involved in valuing and managing it. The collateral may have more value to the borrower than the lender, which implies that, in the case of default, the collateral is transferred from one agent that gives it a higher valuation to another agent that assigns it a lower valuation. This reallocation results in a loss to society. There may also be a cost associated with rationing credit if the collateral is insufficient.

Finally, technological advances can also change the relative costs and benefits of the various ways of alleviating enforcement problems. For example, innovations in information technology have improved recordkeeping and the transmission of information. The effect of improvements in information technology has likely reduced the costs of reputation, monitoring, and collateral, making these tools more effective at reducing information problems. The ability to keep better records enables borrowers to signal information about their reputations. It also allows lenders to gather and evaluate information quickly, which reduces the cost of monitoring. Furthermore, better information technology can improve lenders' evaluations of certain assets that can be pledged as collateral, reducing some uncertainty regarding the collateral's value.

3. The Case of Daylight Overdrafts on Federal Reserve Accounts

We now turn to the specific case of the Federal Reserve's policy regarding daylight overdrafts on accounts that banks have at the Fed.[3] Most, but not all, of the value of overdrafts arises from banks' Fedwire activity.[4] Fedwire is a large-value payments system and a securities settlement system that banks use to send

[3] See Board of Governors of the Federal Reserve System (2007).

[4] Overdrafts can also arise from check clearing and settlement via the Automated Clearing House services provided by the Federal Reserve. The Federal Reserve's overdraft policy applies to the net account balance resulting from activity over all Federal Reserve services to banks. Here, we focus on Fedwire because most of the value of overdrafts is the result of Fedwire activity.

each other funds and government securities on behalf of their customers and their own accounts. Transactions are sent over Fedwire one at a time with finality, which means that the Federal Reserve guarantees that the funds or securities a bank receives will not be revoked.[5] Because transactions are processed one at a time, banks must have access to enough funds to complete each transaction. This need for available funds generates various frictions that banks face in the settlement of transactions, such as search frictions, timing frictions, and incentive frictions (see box).

The Federal Reserve alleviates the impact of these frictions by providing intraday liquidity,[6] which allows qualifying banks to overdraw on their Fed accounts in order to make payments via Fedwire. Banks can acquire overdrafts throughout the day to make payments, but must ensure that their accounts are not in a negative position at the end of the day. The Federal Reserve's provision of liquidity through daylight overdrafts can be interpreted as very-short-term credit.

This exposure is something the Federal Reserve must manage to protect itself from moral hazard or adverse selection problems that may arise from the type of information frictions described earlier. For example, because the Fed does not observe all the actions of banks, it may be concerned that some banks could use daylight overdrafts to finance excessively risky bets. Similarly, the Reserve Banks may not have full information regarding a bank's risk of default on daylight overdrafts. The Fed currently manages its exposure to this form of credit risk with a combination of overdraft fees, reputation, monitoring, and collateral. We now turn to some specifics of the policy to make this connection clearer.

The Federal Reserve charges an explicit price for daylight overdrafts, currently a twenty-four-hour rate of 36 basis points less a deductible. This price, though small, is meant to provide an incentive for banks to minimize their use of daylight overdrafts. But even though this fee may help constrain the size of daylight overdrafts, and accordingly the Federal Reserve's credit exposure, it does not address the information frictions of adverse selection and moral hazard. Thus, other aspects of the policy address those issues.

The daylight overdraft fee provides some incentive for banks to constrain the size of their daylight overdrafts. In addition, the Fed uses a *net debit cap*, which is the maximum dollar amount of daylight overdrafts that an institution may incur in its Federal Reserve account. Each bank that has an account and that is also eligible for intraday overdrafts has a net debit cap. The policy on net debit caps is based on a set of specific guidelines and some degree of banking supervision. The policy allows for one of six ratings for a bank. For most

[5] Although transactions cannot be revoked, that does not mean that they cannot be reversed. Reversals, however, are conducted by initiating a second irrevocable transaction.

[6] Note that some of these frictions are attributable to imperfect information and the absence of commitment. However, we focus here on the incentive problems arising from the provision of intraday liquidity by the Reserve Banks.

Frictions in the Payments System[a]

Search Friction

A search friction refers to the efforts that would be necessary for a payer (the party that intends to send a payment to some other party) and potential liquidity providers to make contact with one another and to determine the right amounts of liquidity to transfer to the payer's accounts. If a payer did not have sufficient funds in its account and did not have access to overdrafts provided by the central bank, it would have to borrow the amount of the payment prior to sending it. But from whom should it borrow? The payer would not necessarily know which other party has sufficient funds in its account, and so it must search for such a lender.

Timing Friction

The timing friction refers to the operational difficulty of achieving the precise timing for when funds will be delivered during the day. Even if parties overcome the search friction and agree on a particular amount of funds to be delivered by one participant to another for the purpose of funding some time-critical payments of the borrower, how will the borrower be assured of receiving the funds at the given time? A commercial bank may have operational difficulties or experience delays for other reasons. The borrower would simply have to wait for delivery of the funds, which reduces the benefits of the arrangement.

Incentive Frictions

There are two incentive frictions to confront in adapting to a withdrawal of daylight credit. First, the rewards of providing intraday funding need to appropriately reflect the costs and risks of doing so. This is also true with overnight funding arrangements, but the intraday timing possibly exacerbates these frictions. Lending $1 billion overnight at a 4 percent interest rate yields approximately $111,000 in earnings, but lending it for an hour at the same rate would yield only $4,600. Assuming the processing costs for arranging the delivery and return of funds are fixed and roughly similar for an intraday and an overnight loan, then it may not be profitable for potential lenders to enter the market at low interest rates.

Second, payments system participants have the option to delay sending a payment rather than borrowing, if the cost of borrowing is too high. However, if all participants are inclined to delay, the system may be vulnerable to gridlock.

[a] This material is borrowed from McAndrews (2006).

banks, net debit caps range from zero to 2.25 times the bank's risk-based capital.[7]

Net debit caps involve a great deal of monitoring. The Federal Reserve reviews supervisory information, evaluates banks' self-assessments (if applicable), and then uses this information to assess the appropriateness of an institution's cap category.[8] This monitoring alleviates some problems associated with adverse selection.

The Federal Reserve also monitors a bank's use of its daylight overdrafts against the cap, providing an opportunity for banks to establish reputations with their regional Federal Reserve Bank. In most instances, banks that exceed their cap limit are required to explain the reason to the Fed and then be counseled to prevent it from happening again. The Fed reserves the right to reduce net debit caps unilaterally, impose collateralization or clearing-balance requirements, reject or delay certain payments, or, in extreme circumstances, prohibit the bank from using Fedwire. Thus, maintaining a reputation of staying under the cap can help banks avoid such actions and can alleviate certain moral hazard concerns.

The Federal Reserve's policy regarding daylight overdrafts uses a combination of fees, monitoring, reputation, and collateral.

Although most daylight overdrafts are uncollateralized, the Federal Reserve uses collateral in two situations. First, it requires collateral from problem institutions to cover any incidental overdrafts. Second, banks wishing to increase their net debit caps can pledge collateral to do so subject to the Federal Reserve's approval. The amount and type of collateral pledged are determined through an agreement between the bank and the Federal Reserve.[9] Collateral plays an insurance role for the Fed in the event of a loss due to an overdraft. It also plays an incentive role for the bank to control its overdrafts and avoid risky behavior that could lead to its closure and forfeiture of the assets it pledged as collateral. Thus, the collateral here also overcomes certain moral hazard concerns.

[7] For foreign banking organizations, a net debit cap is a function of no more than 35 percent of their worldwide capital (referred to as their U.S. capital equivalency).

[8] Each bank that uses a relatively large amount of overdrafts must perform a self-assessment of its own creditworthiness, intraday funds management and control, customer credit policies and controls, and operating controls and contingency procedures.

[9] The type and value of collateral pledged are consistent with the Federal Reserve's discount window policy.

The Federal Reserve's policy regarding daylight overdrafts uses a combination of fees, monitoring, reputation, and collateral. Changes in payments and securities settlement systems, and their effect on the need for intraday liquidity, have led to periodic reviews of this policy to determine whether changes to it can improve the safety and efficiency of the payments system. Recently, in order to ease intraday liquidity constraints and reduce operational risk, the Board proposed changes to its Payments System Risk policy to supply intraday balances to healthy banks predominantly through explicitly collateralized daylight overdrafts.[10] Under the proposal, the Board would allow banks to voluntarily pledge collateral to support intraday overdrafts. Collateralized intraday overdrafts would be charged a zero fee, while the fee for uncollateralized overdrafts would increase from 36 to 50 basis points.[11] We now describe how increasing the use of collateral could bring benefits as well as costs to the Federal Reserve, to banks, and to the financial system as a whole.

3.1 The Benefits of Increasing Collateral Use

Greater use of collateral has the potential to benefit the Federal Reserve, banks, and the financial system in several ways.

The *Federal Reserve* could benefit because collateral provides it with some insurance in the event a bank cannot repay its overdraft. It may also benefit if greater use of collateral increases the incentives for banks to repay their overdrafts over and above the incentives already in place because of monitoring, reputation, and the existing use of collateral.

Banks could benefit if greater use of collateral relaxed some credit constraints. As we observed in the construction loan example, providing collateral can often allow a borrower to obtain better terms on a loan. For similar reasons, the Board's policy proposal includes a zero fee on collateralized daylight overdrafts. In such a case, banks' overdraft costs could decrease when they pledge collateral.

The *financial system* may benefit if the increased use of collateralized intraday overdrafts at the zero fee speeds up the flow of payments across financial markets. The lower cost of collateralized intraday overdrafts may lead to more payments being made earlier in the day, as banks would have less need to delay payments until they have sufficient incoming funds. By encouraging more banks to have collateral pledged at the Fed, increased use of collateral could make it easier for the Federal

[10] See Board of Governors of the Federal Reserve System (2008).

[11] There are other proposed changes as well. See Table 1 in Board of Governors of the Federal Reserve System (2008) for a summary of all the proposed changes.

Reserve to inject liquidity both intraday and overnight in times of financial stress. This is true in particular because collateral is required for overnight loans. In addition, increased use of collateral may prepare banks for financial stress by increasing their ability to borrow at the discount window. All of these spillover benefits may accrue to the financial system through greater use of collateral.

3.2 The Potential Costs of Greater Collateral Use

There are also possible costs to increasing the use of collateral.

The *Federal Reserve* could face higher costs associated with monitoring collateral, such as making sure it is available and valuing it properly. The Federal Reserve already pays such costs because it accepts collateral for overnight loans, but these costs could rise if the amount of collateral increases. Moreover, if there is a greater reliance on collateral for intraday overdrafts, banks may ask to manage their collateral more actively at the Federal Reserve, requiring the Fed to invest in enhancements to its collateral management systems.

Banks would have to pay costs associated with acquiring, managing, and tracking their collateral. Additionally, they may face an opportunity cost associated with using collateral to secure overdrafts because that collateral may no longer be used for other purposes. Banks may also reallocate their portfolio of assets to acquire enough collateral for daylight overdraft purposes. Whether this would constrain banks much depends on the type of collateral that the Federal Reserve and other banks are willing to accept.[12]

The financial system as well may be negatively affected by the greater use of collateral. Collateralized overdrafts make the Federal Reserve a higher claimant on assets of a failed bank, which reduces the attachable assets to residual claimants in the event of a bank liquidation, adversely affecting the unsecured creditors of that bank. This is an issue mainly if the policy is not explained well in advance so that some long-term contracts cannot be renegotiated. Another potential cost would occur if too much of the banking system's assets are tied to collateralized daylight overdrafts. In extreme situations, this could lead to credit rationing in the economy should a shortage of collateral occur. Thus, the increased use of collateral could have negative spillover effects on the economy. Again, whether such a cost is likely to be large is an empirical question and depends on the range of collateral that would be acceptable to the Federal Reserve and other banks.

As with any policy proposal, a careful analysis of the overall benefits and costs of a change in the daylight overdraft policy is essential. It should be noted that the costs of a greater use of collateral are higher for banks if collateral were required for all overdrafts. But if banks are given the choice between uncollateralized lending and posting collateral (with a zero fee on collateralized daylight overdrafts), then their costs should be lower because collateral would be pledged by those banks for which it is the less costly option.

4. Conclusion

How to best enable the extension of liquidity by a central bank is an important policy question. As the examples presented here suggest, it may be desirable to use a combination of reputation, monitoring, and collateral. However, the relative role of each method of enforcing credit arrangements should depend on the details of the contractual relationship considered.

In the future, we can expect the risk faced by central banks to change over time, but we can also expect central banks to have access to more effective enforcement technologies. As banks find themselves in situations requiring them to take quick actions, credit risk can emerge unpredictably and without warning. However, the quality of the tools used by central banks to mitigate these risks has increased as well. For example, technological progress has the potential to make monitoring less costly and more effective in the future. Moreover, new technologies could reduce risk in a number of ways:

- The development of liquidity-saving methods for safely transferring balances could reduce the demand for daylight overdrafts.
- The development of improved markets during the day could potentially lead to a decreased demand for intraday overdrafts, as they are replaced by better methods of intraday distribution of liquidity. In other words, the frictions that require the provision of daylight overdrafts today may be reduced by enhanced technology.
- Technological progress will influence the need to rely heavily on reputation, monitoring, or collateral as some of these methods may become relatively more effective than others.

[12] For discount window purposes, the Federal Reserve accepts a large range of assets of varying liquidity and credit risk, by which they are categorized. The collateral value of each asset is a discounted value of an asset's determined price. This applied discount is based on an asset's class. If the discounts accurately reflect a liquidity and risk premium, banks can have some flexibility in pledging collateral at the Federal Reserve and can minimize opportunity costs associated with collateral. The proposed policy would follow discount window practices to determine which assets are acceptable, those assets' categories, and the discount applied to the assets' determined price.

References

Board of Governors of the Federal Reserve System. 2007. “Federal Reserve Policy on Payments System Risk” (as amended effective January 11). Available at <http://www.federalreserve.gov/paymentsystems/psr/default.htm>.

———. 2008. “Policy on Payments System Risk.” Federal Register 73, no. 46 (March 7): 12417-43.

McAndrews, J. 2006. “Alternative Arrangements for the Distribution of Intraday Liquidity.” Federal Reserve Bank of New York Current Issues in Economics and Finance 12, no. 3 (April).

Federal Reserve Bank of New York
Economic Policy Review

Single-copy subscriptions to the *Economic Policy Review* (ISSN 1932-0426) are free. All articles can also be downloaded at no charge from the Research Group's website.

Library of Congress Card Number: 77-646559