Part 4: The Payments System and the Market for Interbank Funds

The two papers in a session on systemic issues in the federal funds market examined the interbank payments system and the market for borrowing funds used to settle interbank payments. Both analyses were based on data on payments made through the Federal Reserve's Fedwire system.

Fedwire—the nation's primary interbank payments system—is a real-time gross settlement system through which payments are settled individually and with finality. More than 9,500 participants use Fedwire to send and receive time-critical and/or large-value payments on behalf of corporate, financial, and individual clients as well as to settle positions with other participants stemming from payments in other systems, securities settlement, and interbank loans. In the first half of 2006, an average of 525,000 Fedwire payments were made each day, with the daily value of funds transferred averaging \$2.1 trillion.

Complementing the Fedwire payments system is the U.S. federal funds market, where banks borrow funds to settle the payments they make over Fedwire. Fed funds are the bank reserves on deposit at the Federal Reserve used to settle payments between banks. The fed funds market plays a critical role in allocating liquidity in the financial system as well as in supporting banks' ability to finalize the settlement of interbank payment obligations.

The first paper, by Adam Ashcraft of the Federal Reserve Bank of New York and Darrell Duffie of Stanford University, explored whether trading frictions in the fed funds market affect the reallocation of excess reserves to banks requiring funds to complete their payments. Ashcraft and Duffie found that fed funds trading is driven partially by individual banks' precautionary targeting of balances, and that this targeting contributes to systemic stability. The banks' trading of funds mitigates the risk of overconcentration of reserves in some banks, and contributes to the liquidity of the fed funds market and keeps price volatility relatively low. The second paper, by Morten Bech of the New York Fed, Walter Beyeler and Robert Glass of Sandia National Laboratories, and Kimmo Soramäki of the European Central Bank (who did not present), analyzed the network structure of interbank payments and developed a model for a payments system within such a network. The paper concluded that a liquidity market allows a payments system to achieve a specified level of performance with much less liquidity than would otherwise be required—a finding that sheds light on the trade-offs between liquidity within the payments system and friction within the liquidity market. The combination of network topology and bank behavior within a model, the study also found, is critical for analysis of systemic risk in payments systems. According to the authors, it is not sufficient just to understand the topology of the network; knowledge of the processes operating in that topology, such as bank behavior, is equally important.

Systemic Dynamics in the Federal Funds Market

Ashcraft and Duffie analyzed how allocational frictions affect trading in the federal funds market. They also considered

The views expressed in this summary do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

whether these frictions could lead to systemic risk in the form of "gridlock," in which individual financial institutions fail to transfer balances quickly to counterparties as they wait for the counterparties to transfer balances to them. Gridlock creates a self-fulfilling slowdown in the efficient reallocation of excess balances.

Like every over-the-counter market, the fed funds market is subject to allocational frictions because trading is executed through bilateral negotiation. These frictions can be any sources of transaction costs or delays in identifying suitable counterparties, negotiating trades, or executing trades, and

Like every over-the-counter market, the fed funds market is subject to allocational frictions because trading is executed through bilateral negotiation.

they can impact market efficiency. Existing theories of trading dynamics in over-the-counter markets have focused on "search" frictions, whereby traders locate each other with delays, to some extent by trial and error, and negotiate prices that depend in part on the difficulty of finding suitable alternative counterparties. Prices also reflect the relative benefits of making a trade immediately rather than later (and with a newly found counterparty) for each of the two counterparties. As frictions increase and the matching of suitable pairs of counterparties becomes more difficult, a trader with an urgent need to transact has relatively less "leverage" during a bilateral negotiation, and this condition is reflected in the contracted price. The efficiency of allocation and the effect of search frictions on pricing are among the main concerns of the various theoretical studies. Yet there is little empirical work on these aspects of the microstructure of over-the-counter markets.

Ashcraft and Duffie broaden this relatively small body of empirical research as well as add a unique dimension to it. Their analysis of allocational frictions uses transaction-level data from Fedwire, and the majority of their work focuses on the top 100 institutions by payment, or "send," volume for business days in 2005.² This data set permits the construction of real-time balances for each institution and allows for the tracking of the sender and receiver of payments and loans for

every minute of the day. The authors identified a particular send as a loan, as opposed to another form of payment, by analyzing the terms of payments in the reverse direction between the same two counterparties on the next business day. (Fed funds loans are for overnight repayment.)

Ashcraft and Duffie documented evidence of federal funds trading being driven partially by individual banks' precautionary targeting of balances. Banks are motivated to end each day with non-negative balances in their reserve accounts at the Federal Reserve because overnight overdrafts are not permitted except in special circumstances.³ These institutions are also motivated to end each day with relatively small balances, in part because the Federal Reserve does not pay interest on overnight balances and in part because the institutions have other ways to meet their reserve requirements over their two-week maintenance periods, such as by holding currency in large ATM networks and by sweeping funds in reservable accounts into nonreservable accounts.

By targeting its balances, a bank contributes to systemic stability. When its balances are larger than they typically are at a particular time of the day, a bank has an incentive to trade, and especially to lend, so as to reduce the balances. Ashcraft and Duffie showed that, empirically, banks indeed act consistently in response to this incentive. Likewise, when balances are low, banks trade (in particular, borrow) on average so as to raise them. This self-interested balance targeting at the individual bank level promotes systemic stability. It mitigates the risk of overconcentration of reserves in some banks and underallocation in others. Balance targeting reduces the risk of gridlock, and plays a role in keeping the federal funds market liquid and funds rate volatility relatively low.

Drawing on their discussions with fed funds traders, Ashcraft and Duffie reported that fed funds trading is significantly more sensitive to reserve balances in the last hour of the day. For example, at some large banks, fed funds traders responsible for targeting a small, non-negative end-of-day balance ask other profit centers of their institutions to avoid large, unscheduled payments, such as settlement of currency trades, near the end of the day. Once a fed funds trader has a reasonable estimate of the extent of current and yet-to-be-executed send and receive transactions, he can adjust pricing and trading negotiations with other banks so as to push his bank's balances in the desired direction. Ashcraft and Duffie uncovered empirical evidence of this behavior; furthermore,

¹In the fed funds market, brokers reduce but do not eliminate the allocational frictions.

²The researchers protected the confidentiality of the institutions by removing firm-specific details and conducting all data-related work within the Federal Reserve Bank of New York.

³The Federal Reserve's discount window is available, but at terms that make it preferable to achieve non-negative balances through fed funds trading with other banks before the end of the day.

⁴One may think in terms of the usual "eigenvalue" or "mean-reversion" conditions for dynamic stability of a multivariate dynamic system. In this case, the coordinate processes of the system are the current balances of each bank in the system.

they found that such behavior is more pronounced following increases in intraday rate volatility.

In addition, the authors raised the issue of, but did not resolve, whether precautionary balance targeting by banks in the fed funds market, coupled with a regime in which banks forecast the targeting policies of other banks, could have systemically destabilizing consequences. A potential systemic problem could arise, for example, if several large institutions during a day of extreme misallocation of reserves individually "hoarded" reserves, given the heightened risk of other banks doing the same or the institutions' forecasts that other banks are incapable of releasing excess reserves quickly to the market. For instance, Ashcraft and Duffie reported traders' accounts of rumors that this type of behavior was initially feared on September 11, 2001, with the communications disruptions that day affecting the Bank of New York, a large clearing bank.⁵ Any such gridlock was in the end averted by energetic liquidity provision by the Federal Reserve.

Without significant liquidity provision by a central bank during such an event, "a run on reserves" could stress the ability of the fixed intraday supply of reserves to be sufficiently reallocated quickly to meet requirements. (The total amount of reserves in the system is relatively small compared with the total daily volume of transactions.) Even in an extreme scenario, however, access to the discount window as well as infusions of liquidity by the Federal Reserve and other central banks would mitigate adverse systemic effects, as they did on September 11.6

Network-Based Modeling of Systemic Risk in the Interbank Payments System

The cross-disciplinary team of Bech, Beyeler, and Glass began by discussing two new approaches for characterizing the nonlocal, systemwide interconnections that may lead to systemic risk, and then combined the approaches to analyze the Fedwire interbank payments system. Their research first examined the global structure of interconnections in Fedwire by representing bilateral interbank relationships as a network. This approach allows one to quantify the overall pattern of interaction and interdependencies using well-defined measures applied to other complex networks, as well

⁵McAndrews and Potter (2002) describe how the disruption of the regular timing of incoming payments made a bank's liquidity management more difficult, and for some banks the "increased uncertainty (regarding which payments they might receive later in the day) led them to have higher precautionary demand for liquid balances."

⁶For example, see McAndrews and Potter (2002) and Lacker (2004).

as to observe how those measures change during disruptions. Bech et al. next presented a model with simple agents interacting within a payments system network; the model exhibited a transition from independent to highly interdependent behavior as liquidity was reduced. When the authors applied the model to a liquidity market, they demonstrated that a reduction in market frictions can reduce this interdependence and thus the likelihood and size of congestive liquidity cascades.

Network Topology of Interbank Payment Flows

A payments system can be viewed as a specific type of complex network. In recent years, many fields of science have sought to characterize the structure of networked systems and the relationship between network topologies and stability, resiliency, and efficiency. From a communications or information-technology perspective, Fedwire is a "star" network, in which all participants are linked to a central hub: the Federal Reserve. Because of its ability to wire funds, Fedwire is a complete network, as all nodes (participants) can communicate (send and receive payments) with all others.

From a communications or informationtechnology perspective, Fedwire is a "star" network, in which all participants are linked to a central hub: the Federal Reserve.

However, the actual behavior of participants, the flow of liquidity in the system, and thereby the contagion channel for financial disturbances are not captured by these network representations.

The graph of actual payment flows over the Fedwire system includes more than 6,600 nodes and more than 70,000 links, but a subgraph, or core, of just 66 nodes and 181 links accounts for 75 percent of the value transferred. A prominent feature of Fedwire is that 25 nodes form a densely connected subgraph to which all the remaining nodes connect. By itself, this core is almost a complete graph, and this small number of banks and the links between them process the large majority of all payments sent over the network. Soramäki et al. (2006) show that the network shares many characteristics with other empirical complex networks. These characteristics include

⁷See, for example, Newman (2003).

a scale-free degree distribution, a high clustering coefficient, and the "small-world" phenomenon. Apart from the core, the network, like many other technological networks, is disassortative. That is, large banks tend to connect to small banks and vice versa.

Bech et al. showed that the topology of the network was altered significantly by the attack on the World Trade Center on September 11, 2001. First, the massive damage to property and communications systems in Lower Manhattan made it more difficult—and in some cases impossible—for certain banks to execute payments to one another, as some nodes were removed from the system or had their capacity reduced. Second, the failure of some banks to make payments disrupted the coordination upon which banks rely when they use incoming payments to fund their own transfers to other banks.

A Payments System Model

In a paper by Bech et al. (2006), the authors use an elementary, agent-based model in the spirit of models applied to understanding self-organized criticality. 8 Physicists have used these models to study cascading phenomena in a variety of systems (for instance, Jensen [1998]), where models made of very simple agents, interacting with neighboring agents, can yield surprising insights about system-level behavior. In the Bech et al. model, interbank payments occur only along the links of a scale-free network based on the authors' analysis of Fedwire data; the model thus shows that only a very small fraction of the possible interbank exchanges tend to be active. Bank customers randomly instruct the institutions to make payments, and banks are reflexively cooperative: they submit payments if the balance in their payments system account allows them to; otherwise, they queue the instruction for settlement later.

If a bank receiving a payment has instructions in its queue, the payment enables it to submit a payment in turn. If the bank that receives the payment is also queuing instructions, it can make a payment, and so on. In this way, a single initial payment can cause many payments to be released from the queues of the downstream receiving banks. This is an example of the cascade processes typically studied in other models of self-organized criticality.

In the Bech et al. model, a single parameter—overall liquidity—controls the degree of interdependence of interbank payments. Abundant liquidity allows banks to operate independently; reduced liquidity increases the likelihood that a bank will exhaust its balance and begin queuing payments.

When liquidity is low, a bank's ability to process payments becomes coupled with the ability of other banks to process and send payments. In this instance, the output of the payments system as a whole is no longer determined by the overall input of payment instructions but instead is dominated by the internal dynamics of the system, and the correlation between arriving instructions and submitted payments degrades as liquidity is reduced. The model does not exhibit a phase transition between completely noncongested and completely congested states. The distribution of congestive events, or liquidity cascades, is close to a geometric distribution, and it has an inherent length scale independent of network size. This result differs from the finding obtained in systems that exhibit self-organized criticality (such as in Jensen [1998]); the result is directly attributable to differences in the underlying relaxation process that set the simple payments system apart from systems that display self-organized criticality.

The Market for Interbank Payment Funds

Going beyond the study of the payments system in isolation, Bech et al. combined it with a simple model of a liquidity market. Liquidity market transactions were represented as a diffusive process, where liquidity flows are not confined to the links of the payments network. In the model, each bank is directly connected to a central node representing the market, and this connection is characterized by a conductance

An understanding of how participants interact and react when faced with operational adversity will assist payments system operators and regulators in designing countermeasures, devising policies, and providing emergency assistance, if necessary.

parameter that reflects the cost or friction associated with market transactions. The market creates a separate network of liquidity flows that, having a star topology, operates parallel to the network of payment flows.

The inclusion of a liquidity market weakened the coupling between banks and reduced the size of settlement cascades. Bech et al. identified trade-off functions that relate different levels of system performance, in terms of cascade size and

⁸For example, see Bak, Tang, and Wiesenfeld (1987).

queue size, to the value of initial system funding and market conductance needed to achieve that performance. The liquidity market is very effective in reducing cascade size and queue size. For a given level of performance, the rate of liquidity flow through the market relative to the rate of flow through the payments system was surprisingly small. The performance of the system can be greatly improved by the market even though less than 2 percent of system through-put flows in the market. A liquidity market allows a system to achieve a specified level of performance with much less total liquidity than a system without a market would require. Conversely, the performance of such a system is highly dependent on the operation of the market. Disruptions to the market would greatly increase congestion and cascades unless they were mitigated, for example, through the addition of liquidity.

According to the authors, if the combined payments and liquidity market system is modeled with simple processes, the boundary between noncongested and congested states can be explained in terms of the relative values of three time constants: a liquidity depletion time, which is governed by the initial liquidity in the system; a net position return time, which increases with total deposits in the system; and a liquidity redistribution time through the market, which is associated with the market conductance parameter. An understanding of this boundary has significant practical applications because this understanding allows for direct consideration of the trade-offs between liquidity within the payments system and friction within the liquidity market, both of which are modified by central bank policies. While the Bech et al. model does not yet include the behavioral feedbacks that likely factor into the decision making of banks, those feedbacks can be included in order to consider how the congestion boundary may move. A goal of future studies will be to introduce into models this complexity as well as variability in the size of bank payments.

Wide-scale disruptions may not only present operational challenges for participants in a payments system, but they may also induce participants to change the way they conduct business, with the potential to either mitigate or exacerbate adverse effects (Bech and Garratt 2006). An understanding of how participants interact and react when faced with operational adversity will assist payments system operators and regulators in designing countermeasures, devising policies, and providing emergency assistance, if necessary. Accordingly, as Bech et al. argue, it is not sufficient just to understand the topology of the network; knowledge of the processes operating in that topology is as important.

Discussion

Robustness Issues

Much of the discussion following the presentations centered on questions concerning the robustness of the fed funds market and the interbank payments system. In particular, how much shock can the system tolerate—and would a shock move the system to a new, less desirable equilibrium? Moreover, if there are multiple equilibria, which are stable and which are

Although the models presented in this session suggested the resiliency of the payments system when combined with a liquidity market for interbank payment funds, the models were not complete enough to assess these issues definitively.

unstable? Although the models presented in this session suggested the resiliency of the payments system when combined with a liquidity market for interbank payment funds, the models were not complete enough to assess these issues definitively.

The models do not have multiple equilibria because they do not yet include the anticipatory behavior that could produce feedback effects leading to system gridlock. In the combined payments system and liquidity market model, congestive liquidity cascades within the payments system are mitigated through transfers within the fed funds market. Each bank is trying to mean-revert toward a target level of reserves, and the more quickly they adjust toward the target, or the lower the friction, the faster the system moves to a stable and stationary equilibrium. If expectations and anticipatory behavior are introduced, such behavior could produce a model with explosive or unstable equilibria—the "gridlock scenario" that so alarms payments system operators. While the model has not been extended in that direction, it is conceptually feasible to model the self-sustaining volatility that would prevent banks from reaching a stable equilibrium. Doing so empirically, however, would pose a challenge because the rarity of systemic crises severely limits the historical data on the behavior being modeled.

Another challenge in modeling empirically the fed funds market and the payments system is to account for the central bank's deep involvement in the system as well as its ability to take corrective action to stabilize the system. Because of this stabilizing role, no data exist on the extreme states of the system where instability could be found. Those states never occur, or are extremely rare, because of stabilizing intervention. This consideration also complicates the formulation of the anticipatory behavior discussed above: Participants understand the stabilizing role of the central bank and may come to rely on it when forming their expectations—a point we consider later in this summary.

Large-Scale Simulation Models and Policy Decisions

Some of the discussion focused on what lessons might be learned from simulation models and how the integration of models of different scales might inform policy decisions. The models presented analyzed liquidity in terms of the liquidity

At what level of policy can simulation exercises be feasibly aimed—at a high level to inform policymakers of trade-offs and general principles, or at a more granular level to emphasize specific steps to take in a crisis?

properties of the network, such as the transition between congested and noncongested states, and in terms of the functioning of the marketplace for liquidity redistribution. These models allow for simulation of the network's responsiveness to liquidity injections by the central bank. What is striking about them is their sensitivity to these injections. A useful endeavor, suggested during the discussion, would be to construct simulations that shed light on the amount of liquidity needed by the system in different types of stress scenarios. That is to say, how much liquidity should be provided in response to different physical disruptions to the network, and to which institutions should it be provided?

For instance, while we tend to focus on the largest institutions—perhaps on the assumption that they are the most connected—a medium-size network participant might also be important because of the high degree of its connectedness in the network.

Another issue raised was whether the simulation models' degree of resolution would be comparable to the type of information policymakers may actually have in a crisis. For example, is it practical to build simulation models with a high level of resolution? And at what level of policy can simulation exercises be feasibly aimed—at a high level to inform policymakers of trade-offs and general principles, or at a more granular level to emphasize specific steps to take in a crisis?

With regard to these questions, conference participants would like to see closer integration between simulation models at different scales—between the small-scale behavior of individual decision makers and the large-scale aggregate behavior of the system. Accordingly, a hierarchy of models with different scales is needed. The general challenge here, common to multiscale modeling in most domains, is building realistic links between the small-scale behavior at the level of individual agents and the large-scale aggregate behavior of the system. A particular challenge in this application is likely to be the range of behaviors and expectations at the micro level of individual decision makers that can be feasibly modeled in simulations. Thus, to what degree can anticipatory behavior and expectations of economic agents be realistically represented in simulation models, especially in simulations of systemic crises that are not modeling "business as usual"? While the issue has both conceptual and empirical dimensions, the empirical basis for this behavioral component is apt to be limited because systemic crises occur so infrequently.

Behavioral and Mutable Aspects of Network Connections

John O'Brien of the Haas School of Business at Berkeley commented on the application of the web-and-node model to the events of 1987. As O'Brien explained it, the nodes represent the individual investment banks and brokers, and they were all under pressure to be more conservative. Therefore, each node was trying to eliminate its assets and reduce its leverage. The system overall would be harmed by that activity, however, because as each node protected itself it would put pressure on the others. At the system level, the Federal Reserve was trying to make credit more readily available, but that change could not be pushed down to the node level because the top priority of the brokers in the investment banks was to meet margin

⁹As an analogy, consider the electrical power distribution system: What if the government instantaneously opened a new power line that was not there before whenever a failure occurred in the grid?

calls and reduce their leverage. Since the situation in 1987 seems to fit the web-and-node model, O'Brien observed, what does the model suggest about policy if a similar situation were to develop? Robert Litzenberger of Azimuth Trust explained that the answer depends on knowing whether the shock will

Risk analysis of financial systems is more dependent on human behavior than is risk analysis of engineered systems.

dissipate or be self-reinforcing. Because many events will cause only a mild market decline, we need to learn what types of conditions will exacerbate problems. One relevant insight is that economists sometimes partition the market into informed traders and liquidity traders. If the shock is affecting mostly liquidity traders, policymakers might decide to act, Litzenberger said. However, if some of the selling is attributable to informed traders, policymakers might instead opt to refrain from taking action.

Douglas Gale of New York University added that in 1987, some investors could not get in on the other side of the market because dealers essentially would not answer their phones. This problem illustrates one of the important differences between contagion in a complex system that includes rigid links between nodes (such as the electric grid), where the collapse of a node has to be triggered by an event that hits it in a physical sense, and the banking system, where links are somewhat fluid depending on the perceptions and expectations of the nodes. When the links are choices made by people anticipating contagion, the links might start breaking before the contagion reaches them. Gale's comment points to the central importance of human expectations and decisions in the operation of financial markets. Risk analysis of financial systems is more dependent on human behavior than is risk analysis of engineered systems.

Evolution of the Payments System

The session concluded with a discussion of whether the payments system could evolve into a less stable configuration. Participants considered constraints that could be placed on the evolution of the payments system to maintain its stability in the face of ongoing changes in the financial system.

An important issue raised in this context was moral hazard. Whenever the central bank intervenes to mitigate or reduce the effects of a systemic shock, it can influence the future risk-taking incentives of private sector decision makers by weakening the perceived need to plan for the occurrence of extreme shocks. This moral hazard issue can potentially influence the state of the system and its vulnerability to shocks through the risks taken by investors and financial institutions. For researchers, this issue could be a factor when choosing the behavioral features of their models. Consideration of such behavioral issues, more generally, can add to the richness of the results obtained from simulation models.

One notable issue that might be illuminated by the models is the current tendency of payments in the system to migrate toward the end of the day. This shift in payments timing would appear to raise the likelihood of a congested state, as system participants seem to hold back payments until late in the day. Conference discussants noted that the Federal Reserve is looking at how payments system policies affect use of the liquidity pool. For example, which policies might be inducing participants to shift their payments to later in the day, and which might reverse that practice so participants can use the available liquidity in the system more efficiently? Two changes that might have played a role in the shift in payments timing in recent years are the reduction in reserve balances, as banks make more efficient use of reserves, and a higher demand on the pool of liquidity attributable to increases in the volume of obligations to be settled.

Another issue raised was whether the financial system is moving toward increased homogeneity and greater connectivity. Greater connectivity seems to be self-evident, but in some ways we are actually seeing an increase in heterogeneity. A shift has occurred from a bank—oriented financial system to a securities-market-oriented system in which a more diverse population of financial institutions interact. The diversity of investors in today's financial system such as pension funds, insurance companies, mutual funds, private equity and hedge funds, investment banks, and commercial banks—makes for a more heterogeneous set of financial system participants than in the traditional bankoriented models of the financial system. In this sense, the system has become more robust. That said, conference participants observed that a handful of very large financial firms play central roles in the financial system, and we need to look carefully at ways to increase the robustness of the systems and utilities that tie these firms together.

References

- Bak, P., C. Tang, and K. Wiesenfeld. 1987. "Self-Organized Criticality: An Explanation of 1/f Noise." Physical Review Letters 59, no. 4 (July): 381-4.
- *Bech, M., and R. Garratt.* 2006. "Illiquidity in the Interbank Payment System following Wide-Scale Disruptions." Federal Reserve Bank of New York STAFF REPORTS, no. 239, March.
- 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.
- Jensen, H. J. 1998. Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems. Cambridge: Cambridge University Press.

- Lacker, J. M. 2004. "Payment System Disruptions and the Federal Reserve following September 11, 2001." Journal of Monetary Economics 51, no. 5 (July): 935-65.
- 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.
- Newman, M. E. J. 2003. "The Structure and Function of Complex Networks." SIAM REVIEW 45, no. 2: 167-256.
- 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.

The views expressed in this summary do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.