# Economic Review

March/April 1993
Volume 78, Number 2



## Federal Reserve Bank of Atlanta

**In This Issue:**

Economic Review

March/April 1993, Volume 78, Number 2

# Economic Review

## Federal Reserve Bank of Atlanta



The *Economic Review* of the Federal Reserve Bank of Atlanta presents analysis of economic and financial topics relevant to Federal Reserve policy. In a format accessible to the nonspecialist, the publication reflects the work of the Research Department. It is edited, designed, produced, and distributed through the Public Affairs Department.

Views expressed in the *Economic Review* are not necessarily those of this Bank or of the Federal Reserve System.



Free subscriptions and limited additional copies are available from the Public Affairs Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713 (404/521-8020). Change-of-address notices and subscription cancellations should be sent directly to the Public Affairs Department. Please include the current mailing label as well as any new information. ISSN 0732-1813

## Editor's Note

A new feature appears in this issue of the *Economic Review*. Essays such as "New Tools for Regulators in a High-Tech World," presenting analyses and insight of Atlanta Fed research staff members concerning policy-related matters, will occasionally replace the review essay that has been a regular feature of this publication.

# Contents

March/April 1993, Volume 78, Number 2

## 32 *FYI*—The Use of Mitigating Factors in Bank Mergers And Acquisitions: A Decade of Antitrust At the Fed

**Christopher L. Holder**

This article, the second in a two-part series, discusses how the Federal Reserve has dealt with bank merger applications that had potentially significant anticompetitive effects, according to the 1982 Department of Justice (DOJ) merger guidelines, over the last decade. The Fed's Board of Governors approved most of the applications it reviewed during those years, citing a number of factors as mitigating implied anticompetitive effects—competition from thrift institutions, the likelihood of new entry into a market, the financial health of the firm being acquired, and others. This discussion reviews these mitigating factors and provides insight into the Fed's case-by-case approach to considering bank merger applications.

## 45 *P*olicy Essay—New Tools For Regulators in a High-Tech World

**Stephen D. Smith and Sheila L. Tschinkel**

Traditionally, government subsidies have supported certain activities of financial institutions and regulations have limited others. However, the authors argue, the technological advances that have improved efficiency in trade have also made it possible to avoid regulatory barriers at little cost, resulting in an expansion of trading that may expand risk. They conclude that the existing quantity-based regulatory approach does not work as effectively in a low-trading-cost environment and that a user fee-based system imposing costs on financial transactions, rather than prohibiting trades outright, may more efficiently accomplish social goals.

# The Rise of Electronic Payments Networks and The Future Role of the Fed With Regard to Payment Finality

**William Roberds**

Most participants in the U.S. economy can identify with Goethe's observation that "one goes through life with more credit than money." There is a universal need to carry out economic transactions without tapping scarce cash reserves, a need that is met by various forms of credit from credit cards to corporate bonds. Even most economists would agree that a well-functioning credit market is essential for a successful market economy.

An equally important, though less discussed, aspect of credit concerns the settlement of debt with minimum delay, inconvenience, and legal uncertainty. A successful system of credit clearly depends on the ability of debtors and creditors to agree on terms under which debts are considered paid. Uncertainty surrounding payment finality could make potential lenders overly cautious in their extension of credit.

The settlement of credit-based transactions usually involves exchanging a temporary form of payment (otherwise known as credit) for another, final form of payment (money). In this sense, the issue of payment finality is inextricably linked to the larger issue of monetary policy. Consequently, the Federal Reserve System has historically taken a leading role in formulating the laws and regulations involving payment finality. This article considers the Fed's role with respect to the finality or "moneyness" of a fairly new form of payment, namely, large-value or wholesale electronic payments networks. In the United States, the two largest and best known of these networks are Fedwire, operated by the Federal Reserve System, and the Clearing House Interbank Payments System (CHIPS), operated by the New York Clearing House Association.[1]

*The author is a research officer and senior economist in the macropolicy section of the Atlanta Fed's research department.* *He thanks the many colleagues, both within and outside the research department, who have made helpful and extensive comments on earlier drafts. The author is solely responsible for the contents.*

This discussion argues that the current payment practices on Fedwire, CHIPS, and similar networks amount to the creation of intraday money, by means of either an explicit or perceived guarantee of payment finality. Though the intraday money created in this fashion is very short-lived, it is being produced in larger and larger amounts, a phenomenon that has been of much concern to researchers and regulators alike. The position taken here is that the recognition of certain forms of intraday "credit" as money is the key to successfully understanding and regulating the large-value payments networks. The discussion includes a policy proposal that would place an overall cap on the amount of intraday money created via electronic payments networks. Because it explicitly recognizes the monetary role of these networks, the proposal would be likely to result in a more efficient electronic payments environment than would alternative policy regimes.

## Money and Credit Defined

A useful first step toward analysis of any monetary system is a precise definition of terms. The definition of money that is used below is that proposed by Peter M. Garber and Steven R. Weisbrod (1992). According to their definition, *money* is "an asset that promises to maintain its value in terms of the unit of account and therefore becomes generally acceptable in market transactions." In the discussion below, *credit* will consist of the transfer of some commodity (possibly including money) from one economic agent to another, conditional on the promised future repayment of money.

## The Historical Development of Money

While there are many operational difficulties associated with monetary systems, one of the most persistent of these has been the lack of any lasting agreement about exactly what is generally acceptable as money.

From ancient times to the early twentieth century, money was commonly defined as a certain amount of a precious commodity, which was often but not always gold. A system of pure commodity money suffers from a number of problems, however, the most common being what happens to local economic activity in areas where the monetary commodity is scarce. Peter Spufford (1988), for example, documents how the economies of medieval Europe, which were largely based on the use of commodity money, repeatedly spent themselves into recession by running trade deficits with the Middle East. To settle their trade accounts, European countries were constantly exporting precious metals. Despite all government efforts to the contrary, this situation led to the reduction of the stocks of precious metals to the point that monetary exchange could not be sustained, barring the discovery of new sources of gold or silver.[2]

The economies of western Europe eventually managed to break out of this destructive pattern, thanks in part to an influx of precious metal from the New World and in part to the development of institutions that could provide credit. Credit enabled a given amount of commodity money to support a larger measure of economic activity. Two of these credit-providing means are especially relevant for the analysis of electronic payments networks: the banknote and the clearinghouse. The banknote represented the promise of the issuing bank to pay, upon presentation, a certain amount of the accepted commodity currency.[3] These notes, originally a form of credit, gradually became accepted as money (final payment) for most transactions. With banknotes, payments often could be effected without incurring the risks and costs of moving large amounts of precious metal. Banks were able to economize further on the movements of precious metal by establishing clearinghouses.[4] The role of clearinghouses was to calculate, on a daily basis, each member bank's net obligation vis-à-vis all other members. Net obligations would then be settled at the end of the day, using gold or some other mutually acceptable form of payment.

In the United States, late-nineteenth-century restrictions on banknote issue accelerated the development of yet another type of money—checkable bank deposits.[5] The widespread use of checks allowed banks even greater economies on the use of precious metals or banknotes. The movement toward check money was assisted by the further development of private clearinghouses, and later by the Fed's campaign to establish nationwide par (face-value) check clearing.[6]

## The Economics of Electronic Payments Networks

The evolution of money is a continuing process. Today, the vast majority of transactions are still carried out via the monetary "inventions" discussed above—that is, by currency or check. Increasing numbers and

amounts of transactions are taking place in pure electronic form, however, particularly on the large-value or wholesale wire transfer networks. For example, the overall transactions volume for large-value wire transfers in the United States (primarily Fedwire and CHIPS) has been conservatively estimated at approximately 100 million transactions for 1990. While this number represents only 0.1 percent of all transactions in the United States for 1990, the total value of these transactions represents about $421 trillion—83.7 percent of the value of noncash transactions for that year and roughly thirty-five times the total value of U.S. gross domestic product.[7] Most of the payments over these systems are associated with either the domestic financial markets or markets for foreign exchange.

The emergence of electronic payments networks poses both great opportunities for market participants and great challenges to present or potential regulators. The current state of computer technology is such that the time between the initiation of a transaction and settlement could, from a purely technical viewpoint, be reduced to a matter of minutes in most instances. However, this ongoing type of settlement (known as gross settlement) has not become the accepted norm for payments networks, either domestically or abroad. Instead, many electronic payments networks have opted for once-a-day net settlement. Under such an agreement, at the end of the business day a bank or another payments network participant pays to (receives from) the network reserve funds equal to their total net debit (credit) position vis-à-vis all other network participants. This "clearinghouse" type of arrangement is often referred to as multilateral netting.[8]

Strong economic incentives operate in favor of such an arrangement. Suppose, for instance, that six banks are organized into an electronic payments network. On a certain day each bank wants to transfer $1 million to each of the other five banks. Under gross settlement, a total of thirty transactions would have to occur, and a total of $30 million would change hands. Under a multilateral netting scheme, no money would actually be exchanged. At the end of the day, each bank's net obligation to the other banks would be zero, and no payments would be necessary.[9] Given this opportunity to economize on transactions balances, it is hardly surprising that most electronic payments networks have not tried to further reduce the interval between payment initiation and settlement.

In an era of electronic payments systems, a payments network with multilateral netting serves one of the same functions that banknotes, clearinghouses, and checks served in the pre-electronic era—that is, economizing on costly transactions balances. In the uncertainty of the real world such multilateral netting arrangements also help to reduce the credit risk associated with the payments network. Risk is reduced because, other things being equal, the amount of funds that each participant must "front" to settle is typically smaller, thereby making it less likely that a participant would not have access to sufficient funds to settle.

A potential disadvantage of multilateral netting is that it requires a high degree of mutual trust and cooperation among participating institutions. In practice, this problem is not insurmountable, however, because a major purpose of electronic payments networks is to facilitate payments among firms that are accustomed to doing business with one another. And electronic payments networks of any type require a high degree of cooperation on matters such as computer formats, security procedures, provisions for backups, and so forth. The cooperation necessary for a multilateral netting agreement seems only a natural extension of that already required for the existence of a given payments network.

Both of the major large-value payments networks in the United States—Fedwire and CHIPS—carry out their operations under rules that, to some extent, embody the multilateral netting principle discussed above. However, the details are quite different across the two systems. Fedwire is nominally a gross-settlement system: under Fedwire rules, payments made through the Fedwire network are in almost all instances final and irrevocable. Finality of payment is guaranteed by the Fed, and reserve funds are made available immediately to the receiving institution. De facto multilateral netting can still take place by means of "daylight overdrafts."[10] A daylight overdraft occurs when an institution sends an amount of funds over the network that exceeds its operating reserve balance plus the sum of any incoming transfers. Fed regulations require that any such overdrafts must be repaid by the end of the business day, either by additional incoming transfers or deposits of additional reserve funds. Fedwire overdrafts are also subject to other restrictions, which are discussed below.[11]

In contrast to Fedwire, CHIPS operates on an explicit net-settlement basis. Payment messages sent during the day are not final until end-of-the-day settlement occurs. Normally this settlement occurs around 6:00 P.M. via special Fedwire accounts at the Federal Reserve Bank of New York. Because CHIPS is not a bank and has no bank accounts, there are no daylight overdrafts per se. The equivalent of daylight overdrafts on CHIPS is the credit that participants are willing to

extend to one another. As with Fedwire, the allocation of such credit is subject to certain rules and limits.[12]

## *Is* It Really Money?

Are daylight overdrafts over Fedwire and intraday "credit" over networks such as CHIPS really money or simply a convenient form of credit? By sending payments orders over EFT networks, banks and their customers can de facto expand banks' intraday balance sheets in a process that closely resembles the creation of traditional, overnight bank deposits. The details of this process are spelled out in Appendix 1.

Not every expansion of banks' balance sheets constitutes money creation, however. To qualify as money, a bank liability should pass a standard test of its "moneyness." That is, how close an approximation is this liability to final, irreversible payment? In the case of Fedwire, the Fed's explicit guarantee of payment finality clearly qualifies daylight overdrafts as money. For CHIPS and other private networks with multilateral netting, the money/credit question is more subtle. The fact that finality of payment is not guaranteed over CHIPS by the Fed or any other governmental entity raises the issue of systemic risk.[13] Over a payments network, systemic risk refers to the risk that some network participants may not be able to settle their net obligations at the end of the business day, thereby forcing other participants to come up with funds to cover incoming transfers expected from the failing participant. If this requirement, in turn, caused other participants to fail to meet their obligations, the integrity of the network or of other parts of the payments system could be compromised.

Although such a crisis has never occurred over CHIPS, the CHIPS system has taken a number of steps to limit its participants' exposure to systemic risk. These include limiting the net credit and debit positions of each participant (bilateral credit limits and overall net debit caps, respectively) vis-à-vis other participants. More recently, CHIPS has adopted a loss-sharing arrangement, which is backed by collateral requirements and is designed to ensure that the losses from one participant's failure are borne by more than one other participant. The total amount of collateral required, however, is small (estimated at $3 billion to $4 billion) relative to the average amount of intraday credit extended via CHIPS (approximately $20 billion to $30 billion). Studies performed by CHIPS indicate that these measures probably would cover losses induced by one member's failure to settle but that the simultaneous failure of two or more members could easily exhaust these provisions. As a last resort, CHIPS rules allow for its governing committee to take any measures necessary to complete the settlement process. However, the committee is limited in its ability to impose additional loss-sharing obligations on other CHIPS members.[14]

Despite the fact that the Fed does not guarantee payment finality over CHIPS, it can be argued that for purposes of economic analysis payments over CHIPS and similar networks should be considered "approximately" final and that CHIPS intraday credit should be considered money. The first reason is that the "settling" members of CHIPS and similar networks are U.S. banks, which would in most cases have access to the Fed discount window. A second reason to consider payments over many private networks de facto final has been advanced by David L. Mengle (1990), among others. Even though there is no guarantee of finality over these networks by the Fed, or by any other regulatory agency, it can be in the best interests of participants in these networks to act as if such a guarantee were in place. By pursuing this course of action, Mengle explains, network participants are in effect betting that the network will be "bailed out" by some sort of governmental intervention should a crisis develop that could cause a settlement failure. The odds may favor this bet if regulators have strong incentives to prevent settlement failures and their negative consequences.

The viewpoint that CHIPS is too big to fail appears to be widespread. Marcia Stigum quotes one bank officer who is responsible for his bank's CHIPS operation as saying, "If CHIPS fails to settle, I jump out of my window. CHIPS cannot not settle because, if it were to fail to do so, it would destroy confidence in the money market internationally" (1990, 903). The view that the Fed "stands behind" private payments systems is apparently shared by a large number of private-sector observers of the payments system, including officials of the American Banking Association (Philip S. Corwin and Ian W. Macoy 1990, 11) and quite a few academics (for example, Robert Eisenbeis 1987, 48; Andrew F. Brimmer 1989, 15; Ben S. Bernanke 1990, 150; Hal S. Scott 1990, 187).

As long as this perception prevails in the private sector, it makes little difference whether or not the Fed explicitly guarantees payment finality. In other words, the market has its own view of what policies the Fed would pursue in the event of an impending systemic crisis, and such views are not under Fed control. The bottom line is that payments over many private

networks in effect constitute money because they are perceived as such.

## The Policy Problem and Some Real-World Complications

The above discussion shows that the process of netting payments over electronic networks represents a rational private-sector response to the problem of how to economize on costly transactions balances. By exchanging payments messages during the business day and settling net positions at the end of the day, payments network participants economize on their need for costly reserve balances. In the jargon of monetary economics, these savings result from the substitution of "inside" money (money created by the private sector, in this case the payments messages or daylight overdrafts) for "outside" money (in this case, electronic reserve funds held with the Fed).[15]

The ability to create this new form of inside money has certainly been of benefit to the institutions that have been able to do so, and its creation has likely been of net benefit to society as a whole. At the same time, there is evidently a limit to a good thing. The willingness of the Fed to absorb systemic risk associated with electronic payments systems is large but surely not infinite. As is the case with traditional, overnight bank deposits, the economy achieved by the private sector's substitution of inside money for outside money should be offset by a calculation of the costs of the Fed's guarantee of liquidity in the event of a systemic crisis. It is worth noting that the Fed's total liability in the event of such a crisis would not be bounded by the net amounts due to settle but by the gross amount of payments messages entered into the various networks (see Appendix 2). Maintaining a credible (though perhaps implicit) guarantee against systemic risk in electronic payments networks cannot be consistent with unlimited growth in the number of such networks or indefinite growth in the volume of potential liabilities created via these networks.

Thus, the essential policy problem associated with electronic payments networks is how to contain the systemic risk associated with the creation of intraday money via these networks without imposing undue costs on the private sector. It needs emphasizing that the scope of this particular problem goes beyond the confines of domestic bank regulation. On the international front, several multicurrency payments systems are currently in the planning stages. The advent of cross-border payments networks poses some noteworthy complications for policy concerning electronic payments networks.[16]

The first potential complication lies in the sheer size of the international currency markets, whose daily volume is now close to $900 billion.[17] Because the gains from netting arrangements are proportional to the volume of payments over a given network, the incentives for netting cross-border payments are strong. Also, for a given dollar volume of payments, cross-border payments networks can offer stronger incentives for netting than domestic networks, especially if a substantial number of these obligations have to be converted to another currency before settlement. In other words, it is highly likely that a large volume of inside money will be created over these networks and that much of this money will be dollar-denominated.

A second potential problem is the temporal separation of markets in various currencies. This complication should be of particular concern to U.S. policymakers, given that Western Hemisphere financial markets close (and settle) after Asian and European markets have closed for the day. As a result, for certain foreign exchange transactions there is a risk associated with the fact that payment in the foreign currency will have been finalized before the offsetting payment in dollars becomes final.[18] According to Bruce J. Summers (1991, 85), an average daily volume of as much as $400 billion in foreign exchange transactions is settled (on the dollar end of these transactions) by CHIPS at the end of the U.S. East Coast business day. The time delay between initiation and settlement for some of these transactions can be as long as fourteen hours.

On the domestic front, there are incentives for private, nonbank firms to organize themselves into payments networks, including those that allow for bilateral or multilateral netting of obligations. Such networks, known as delivery-versus-payments systems, already exist for U.S. government securities, mortgage-backed securities, and commercial paper. In principle, there is no reason why such arrangements would not be extended to any heavily traded commodity.[19]

The operation of domestic, nonbank payments networks raises policy concerns similar to those listed above, and particularly the issue of ultimate responsibility for the integrity of the network. The settlement of nonbank obligations will, in all probability, continue to be effected via the banking system (that is, through Fedwire). If there is a market perception of a de facto Fed guarantee against systemic risk in these networks, it would be difficult not to recognize the intraday credit extended over the nonbank networks as intraday money.

## Possible Policy Responses

The discussion to this point has attempted to show that strong incentives exist for the extension of intraday credit, which in many cases is regarded as intraday money via electronic payments networks. Given these incentives and continued technological improvements, the volume of such credit can be expected to grow over time.

The intraday credit extended over these networks, particularly over Fedwire and CHIPS, has hardly escaped the attention of U.S. policymakers. However, the development of the legal and regulatory framework for electronic payments networks has proceeded at a relatively slow pace. The measured pace of regulation in this area reflects a fundamental dilemma of regulating payments systems. Because the critical characteristic of a free-market economy is voluntary, mutually beneficial exchange, policymakers are reluctant to burden payments networks with restrictions that would unnecessarily hinder such exchanges.

At the same time, there is a widely recognized need to provide safeguards for payments network participants against systemic risk. Some such protection is afforded by the Fed discount window. However, there are certain disadvantages associated with a reliance on the discount window as a means of protection against systemic crises. One potential drawback is that the size of discount window loans necessary to avert a systemic crisis could be quite large, potentially conflicting with monetary policy objectives. Another drawback is the set of restrictions imposed on the use of the discount window by the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA).[20] Section 142 of FDICIA limits discount window lending to undercapitalized banks to sixty days within any 120-day period unless the Fed or the undercapitalized bank's primary federal bank regulator certifies that the bank is viable. In the case of critically undercapitalized banks, FDICIA instructs the Fed to demand repayment of discount window loans within five days. If the 5-day limit is violated, the Fed must share with the FDIC in any resulting increase in costs, and Congress must be notified of any payments to the FDIC under this provision.

Larry D. Wall (1993) reports that a major objective of FDICIA was to limit the extent of the too-big-to-fail doctrine. Wall notes that although FDICIA allows for "systemic risk" exceptions to its restrictions, invoking this exception requires approval of the FDIC, the Fed, and the U.S. Treasury (Section 142). Thus, while recognizing the importance of a lender of last resort in averting systemic crises, Sections 141 and 142 of FDICIA mandate a clear set of incentives that discourage excessive reliance on the discount window as protection against systemic risk.

Against this background, the Board of Governors of the Federal Reserve System acted in 1990 to strengthen restrictions on daylight overdrafts incurred over Fedwire. A major objective of these restrictions is to limit the amount of systemic risk borne by the Fed in its operation of Fedwire. The most substantive restrictions cap the intraday credit granted to any Fedwire participant, limiting the amount of this credit to a fixed percentage of the participant's risk-adjusted capital.[21] Beginning in 1994 the Fed will also start phasing in interest charges for overdrafts that exceed a deductible, which is also a fixed percentage of risk-adjusted capital. These charges will gradually rise to a level of 25 basis points at an annual rate.[22] These restrictions, together with the new regulations adopted by CHIPS (discussed above), have no doubt helped to restrict the potential for systemic crises in these two large-value payments networks.

Given the increasingly diverse use of electronic payments networks, however, it is questionable whether existing regulation of intraday netting over Fedwire, or even domestic interbank payments networks more generally, will be sufficient to eliminate the possibility of systemic risk. Commenting on the 1990 changes in the rules regarding Fedwire overdrafts, Corwin and Macoy note that "[i]ronically, the result of . . . Federal Reserve policies seeking to limit the growth and totals of daylight overdrafts on Fedwire is to shift them to private wire systems" (1990, 10).[23] In view of the various reforms recently adopted on CHIPS, Corwin and Macoy's statements could probably be applied to some degree to that system as well. To be effective over the longer term, any scheme for minimizing systemic risk over electronic payments networks will have to address the presence of this type of risk on all networks that make use of intraday netting of payments.

At one end of the policy spectrum, a suggested remedy to this situation would be the elimination of intraday netting in favor of real-time gross settlement, that is, gross settlement without daylight overdrafts. While this policy ignores the potential gains from netting arrangements, it has modern technology on its side. That is, if technological improvements make it possible for gross settlement to proceed on a virtually real-time basis, the cost of a gross-settlement system could be reduced vis-à-vis netting arrangements with

daily settlement. A real-world approximation to such a system is the Swiss Interbank Clearing (SIC) system. Christian Vital and Mengle describe SIC as "a centralized gross settlement system created to process interbank payment transactions with no daylight overdrafts and therefore no systemic risk" (1988, 23). However, even on this system, the time from initiation of a transaction to settlement often exceeds the technologically feasible minimum of thirty seconds. Vital and Mengle note that as of November 1988, 55 percent of transactions were executed within two hours of initiation, and 85 percent, within five hours.

Could such a system work in the United States? Certainly the introduction of gross settlement to U.S. payments networks would pose a larger, though hardly insurmountable, technical challenge. In 1989, SIC had 163 members versus 139 for CHIPS and 11,435 for Fedwire. (The total number of participants for Fedwire is somewhat deceiving because only about 2,000 high-volume participants maintain direct computer-to-computer links to the Federal Reserve Banks.) The 1989 average volume of transactions over SIC was comparable to that of Fedwire: daily averages were 223,000 for SIC versus 238,000 for Fedwire and 146,000 for CHIPS. However, the average daily value of the transactions over SIC was much less than for the U.S. networks: $73 billion for SIC versus $730 billion for Fedwire and $761 billion for CHIPS.[24] The higher value for the U.S. networks means that sustaining similar volumes under a gross-settlement system would raise the probability of a situation known as "payments gridlock," whereby numerous network participants would each be waiting for other participants to make the first payment. The elimination of net settlement could also contribute to payments gridlock by encouraging network participants to wait until the last possible moment to enter payments messages into the network so as to economize on intraday reserve balances. Prevention of payments gridlock would require installing additional hardware to handle the last-minute volume or the introduction of peak-hour pricing of settlement services.

A constraint even more limiting than any operational difficulty associated with gross-settlement systems would be the reluctance of current users of intraday netting to move to gross settlement. In fact, history favors continued development of intraday credit as a form of money. Once a form of credit—for example, banknotes or checks—has become accepted as a form of money, attempts to regulate that form out of existence have ultimately been unsuccessful. And in at least one case in which stringent regulation was successfully introduced—a tax on banknotes by a congressional act of 1865—the ultimate effect of this regulation was the accelerated development of bank account money, an alternative form of inside money.

## A Monetary Alternative

There have been numerous studies and proposed policy responses to the problem of "daylight overdrafts" or "intraday credit."[25] Generally speaking, these studies are of high quality, and most of the suggested policy responses represent sensible approaches to this issue. However, a common failing in this literature is a general reluctance to admit that the extension of intraday credit via electronic payments networks is equivalent to the creation of money.[26] The fact that "electronic intraday money" comes in a form different from currency or bank accounts does not affect the validity of this generalization. Paper currency and check money both developed as claims to a different, more widely accepted form of money. Electronic payments, which began as a form of claim on check money, are coming to be more widely used as money. A reasonable first premise of an effective policy on payments networks would be that once something is used as money, it should be viewed as such for purposes of policy.

A distinguishing feature of electronic intraday money, as it currently exists, is that it is all inside money. By contrast, more traditional forms of money consist of a combination of inside money (transactions accounts at depository institutions) and outside money (currency plus bank reserves with the Fed). Traditionally, the amount of inside money held by a depository institution is limited by reserve requirements to be no greater than a fixed multiple of its holdings of outside money. With electronic intraday money, no such requirements exist. Abstracting from such restrictions as bilateral or multilateral "caps," the sole restriction on the creation of this kind of money is the requirement that it disappear by the end of the trading day. The second premise of an effective payments system policy, in the framework of this proposal, would be the institution of some mechanism analogous to a reserve requirement for all forms of electronic intraday money.

In calling for the institution of a reserve-like requirement for intraday money, it should be pointed out that the institution of such a requirement would not be a panacea for all of the regulatory issues associated with the operation of payments networks that allow for netting of intraday payments. In particular, the institution of reserve requirements is not seen as a

substitute for risk-limiting measures such as capital and/or collateral requirements, real-time monitoring of net debit positions, and so forth. Rather, the establishment of a reserve-like mechanism for intraday money creation would serve some of the same purposes as imposing reserve requirements on ordinary, overnight deposits in transactions accounts: delimiting the Fed's liabilities as lender of last resort and supplying a means of pricing the protection provided by the Fed against systemic crises (which may currently be seen as an implicit, rather than explicit, guarantee).

The third and final premise of an effective payments system policy would be recognition of the principle that the successful operation of a payments system, particularly one with multilateral netting arrangements, requires the existence of an institution analogous to a central bank. In the case of the U.S. banking system, the function of a central bank is carried out by the Federal Reserve System; similar institutions exist in most countries today. Although in the late nineteenth and early twentieth century no such institution existed in the United States, many of the present-day functions of the Fed were carried out by private clearinghouse arrangements. Concerning the role of these private clearinghouses, Gary Gorton has noted that "by the early twentieth century clearinghouses looked much like central banks. They admitted, expelled, and fined members; they imposed price ceilings, capital requirements, and reserve requirements; they audited members and required the regular submission of balance sheet reports. . . . [T]hey issued money and provided a form of insurance during panics" (1985, 283).

The fact that such institutions were created voluntarily suggests that some analogous regulatory organization, be it public or private, will inevitably be associated with any electronic payments network.

## *R*eservable Electronic Intraday Money: Some Details

How could reserve requirements be imposed on the intraday money created by electronic payments networks? It seems that currently available payments technology could be used to create a sort of tradable electronic certificate, as follows.

The certificates would be called something like "electronic intraday cash creation rights" (EICCR) and could only be created by the Fed. An EICCR would confer on its owner the right to create intraday money via an electronic funds transfer system with a netting arrangement, up to some prespecified limit. A network participant placing a payments order over a certain network would deposit the required EICCR "collateral" with the relevant network until settlement. Because EICCR would be created in a limited amount, it would have positive value. After its creation, EICCR would be available for purchase by any depository institution with its reserve funds.[27] EICCR not held as collateral by a payments network could be resold to other depository institutions. Nonbank firms could buy EICCR from a depository institution at which they maintained a transactions account.

The term collateral as used here does not mean that EICCR constitutes collateral in the usual sense of "an item of sufficient value such that its liquidation would provide funds necessary to cover any default on a particular debt." In practice, it would be highly unlikely that EICCR liquidation would cover more than a fraction of the funds necessary to cover the obligations of a failed network participant. EICCR might be better described as proof of payment on an insurance policy, under which the proof of payment could be traded among different network participants provided that it was not currently in use.

There are several details of the EICCR plan not described above that would need to be specified in practice. The first such detail would be a way to decide on an initial allocation of EICCR—the amount of EICCR to be created and who would be entitled to it. The question of how much EICCR to create would pose difficult but not insurmountable problems of the same nature as those faced in determining the optimal rate of growth for the Fed's open market portfolio. From a standpoint of economic efficiency, the initial allocation of EICCRs would be essentially irrelevant. One possible candidate for an initial allocation would be to "grandfather in" participants in existing networks by providing them with EICCRs equal to, say, their average maximum intraday exposure over some specified time period.

Another detail that would have to be worked out would be a mechanism for intraday transfer of EICCR. An obvious candidate for such a mechanism would be real-time delivery of EICCR against payment over Fedwire. To be effective, however, such a mechanism would require a "daylight overdrafts" policy of sufficient stringency so as to prevent the widespread substitution of Fedwire overdrafts for intraday money creation over private networks.

Proposals such as the one described above have not been given serious policy consideration in part because

of the perception that a mechanism such as EICCR would pose insurmountable technical difficulties. For example, Edward C. Ettin states that "the sheer mechanics of calculating deposit and reserve balances second by second make this approach impossible" (1988, 290). It is not necessary to delve into the technical details to find such claims difficult to support, given the current technological capabilities in the banking industry. With the availability of debit cards and point-of-sale (POS) terminals, consumers are able to purchase electronically such items as gas and groceries on a real-time basis.[28] If everyday retail items can be bought in such a fashion, then it seems reasonable to assume that currently available technology could allow for the real-time purchase of the right to create intraday money.

As a method of managing the aggregate amount of systemic risk associated with the payments system, an EICCR-based limit on the creation of electronic intraday money would have several advantages over more direct regulation of electronic payments networks. A requirement that intraday money creation be "collateralized" by EICCR would place an effective aggregate limit on the amount of intraday electronic money outstanding at any given time. An advantage that an EICCR-based constraint would have over simple quantitative caps is that it would encourage the development of a market for intraday funds, particularly intraday EICCR. With such a market, an EICCR-based constraint would be more efficient than quantitative caps in the sense that electronic intraday money would be created in the largest amounts in the networks associated with the highest demands for funds. An obvious and beneficial side effect of the EICCR market would be the availability of a market price for intraday "credit"—that is, a price reflecting the true value of the Fed's safeguards against systemic risk in these markets. An operational difficulty associated with policies that advocate the ad hoc pricing of daylight overdrafts is that such policies provide no direct measure of the appropriate price of intraday money. Incorrect pricing, in turn, would amount to levying an unintended tax on, or providing an unintended subsidy to, the creation of intraday money.

The EICCR proposal outlined above bears a strong resemblance to the idea of marketable emission permits in the area of environmental policy. Under such a policy, firms emitting a harmful pollutant into the environment must purchase a permit to do so. The permit allows for emission of the pollutant up to a specified amount. By limiting the number of permits for a given pollutant, the government can control its total emissions. The permits can be freely traded between polluters at market prices. Although the available evidence on the overall efficacy of such permits is mixed, there is a good deal of evidence to suggest that the use of permits represents a more cost-effective approach to pollution control than does direct regulation (explicit quantitative caps on emissions by each producer).[29]

Under the EICCR proposal, the analog of "pollution" would be systemic risk. That is, in the course of producing a desirable commodity—intraday money—participants in electronic payments networks would create an undesirable by-product, systemic risk, most of which would be borne by the Fed as a lender of last resort. As in the case of pollution permits, it is possible to limit the amount of this risk outstanding by requiring permits for creation of such risk (via the netting of payments) and limiting the total amount issued.

The pollution analogy is also useful for illustrating the key difference between the EICCR proposal and a system of explicit, fixed charges for daylight overdrafts or intraday credit. The equivalent of overdraft charges in terms of environmental economics would be a per-unit pollution tax. Overdraft charges attempt to limit the extent of something undesirable—systemic risk—by fixing the price of the risk and letting the market determine the quantity of the risk that would be incurred, a policy that economists term *price rationing*. In contrast, the EICCR proposal would limit the quantity of risk and let the market determine the price, an approach known as *quantity rationing*. In the case of intraday money, quantity rationing (and the systemic risk associated with creating intraday money) would possess an important advantage over price rationing because the closest substitute for intraday money—overnight reserves or federal funds—is already effectively price-rationed by the Fed by means of daily interventions in this market. To fix prices successfully in the markets for two close substitutes (overnight and intraday money), one would to have to possess precise information on the relative price of the two forms of money. Otherwise, one form of money would be overpriced relative to the other, causing market imbalances as market participants try to convert their funds from the more expensive to the cheaper form of money. Under a quantity-rationing scheme for intraday money, the informational demands on the Fed would be less stringent. The Fed would simply set a cap on the maximum amount of intraday money that could be created on a given day. The market would then decide the appropriate price for this money, a price that would be consistent with the going rate of federal funds.

A proposal for reserve requirements on intraday money, similar in some respects to the EICCR proposal outlined above, has been advanced by E. Gerald Corrigan (1987) and put forth in more detail by Kausar Hamdani and John A. Wenninger (1988). In the case of Fedwire daylight overdrafts, Corrigan and Hamdani and Wenninger propose that overdrafting banks hold supplemental overnight reserves at a level that is the average of these overdrafts. Supplemental balances would earn interest at a rate lower than the overnight fed funds rate. While the basic idea of the Corrigan and Hamdani-Wenninger approach—making intraday money reservable—is the same as that behind the present approach, there are some noteworthy differences. The Hamdani-Wenninger proposal is limited to overdrafts on Fedwire, whereas this proposal suggests that an EICCR-based "reserve requirement" be applied to all electronic payments networks that allow for intraday netting of obligations (Corrigan also suggests that reserve requirements be widely applied). For the reasons outlined above, imposition of a reserve ratio on Fedwire, without imposing similar requirements on private networks, would have the undesirable side effect of encouraging the creation of additional intraday money and thereby additional systemic risk on the private networks.[30] Both the Hamdani-Wenninger and the Corrigan approaches also propose reserve requirements (clearing balances) based on average levels of overdrafts, with the averaging taking place over some specified period. On the other hand, an EICCR-based reserve requirement would effectively place a continuously administered reserve requirement on the creation of intraday money. Because intraday money is created on a continuous basis, a true reserve-based market for such funds would have to reflect the continuous changes in the availability of reserves to be held against such funds. A continuously applied reserve requirement would therefore seem preferable, abstracting from technical difficulties, to reserve requirements based on a time average.

By serving as a uniform international standard for electronic payments networks, an EICCR-reserve-based limitation on the creation of intraday money could serve to lessen the risk associated with cross-border payments networks. Such limits could also help to reduce Herstatt risk (see note 18), for example, by means of a uniform international requirement that cross-border payments network participants granting credits to be settled in foreign currencies hold EICCR reserves in the settling currency. Any such requirements would necessitate that EICCR for each currency be made available on an around-the-clock basis.

## Potential Problems

Aside from technical difficulties, one could foresee other potential difficulties with the introduction of intraday reserve requirements, based on either EICCR or more traditional reserve accounts. One of the most troubling from a regulatory viewpoint would be the blurring of the distinction between banking and commerce. Any statutory recognition of intraday electronic exchanges between nonbank firms as a reservable form of payment would come close to conferring on these firms the legal authority to create money, a right currently reserved for depository institutions that are regulated, examined, and insured by governmental agencies. If nonbank firms are to be involved in the production of intraday money, their involvement raises the question of what kind and degree of regulation would be appropriate for these firms.

The inherent difficulty of such public policy issues does not mean that an EICCR requirement or a similar limitation should not be placed on electronic intraday money. The unfortunate legacy of historical policy toward intraday money has been the encouragement of an unsustainable "bielectronic" monetary standard. That is, current policies allow for a dual standard whereby one form of money (transactions funds held overnight at depository institutions) is reservable, yet another form (electronic intraday payments) is not, and the result is a situation reminiscent of the late nineteenth-century U.S. experiments with bimetallism. Theoretically, bimetallism was supposed to allow for the simultaneous maintenance of gold and silver commodity standards, plus the maintenance of a strict mint ratio of convertibility between the two metals. The fate of such schemes is well known: in practice, because the mint ratio between the two metals rarely equals their relative market prices, the cheaper of the two metals circulates while the more expensive metal is converted to a nonmonetary commodity.[31]

In the present-day situation, the existence of reserve requirements on overnight money means that inside money created in the form of traditional checking accounts is more costly to produce than inside money created via intraday credit over electronic payments networks because the latter is currently nonreservable. If there were a private market for the exchange of the two forms of money, intraday money would probably trade above par with overnight funds.[32] Yet par convertibility of electronic intraday money is maintained by the daily settlement of electronic accounts via the exchange of reserve funds—that is, via Fedwire. Par

settlement of electronic intraday money in effect causes this money to be underpriced relative to the usual, overnight bank funds.[33] The imposition of an EICCR collateral requirement on intraday money could rectify this situation while maintaining par valuation of intraday money in terms of overnight money.

The foregoing discussion assumes the continued existence of reserve requirements on transactions accounts at depository institutions. However, many economists question the need for reserve requirements, and any policy designed to regulate the creation of intraday money must contend with the possibility that these requirements could be abolished.[34] In the case of EICCR, it can be argued that the efficacy of an EICCR collateral scheme need not depend on the existence of explicit legal reserve requirements for overnight money.

For example, suppose that explicit reserve requirements for overnight funds were abolished. Even in this case, reserve accounts would still be of value for the liquidity they provide. Most banks would almost certainly continue to hold positive reserve balances for clearing purposes. If these "clearing balances" paid no interest or paid interest at a rate below the market rate on overnight funds,[35] in the absence of EICCR or a similar collateral requirement there would still exist a positive cost differential between the cost of overnight and intraday money, a differential that would reflect the cost of holding liquid, but low-yielding, reserve funds. Hence, the right to create intraday money, which would be conferred by posting of EICCR collateral, would still be of positive value. The price of EICCR would reflect the value of intraday money in reducing the need for costly reserve balances.

A more fundamental objection to placing reserve requirements on electronic intraday money is implied in, for example, the views expressed by Lawrence H. White (1984, 1989). White argues that governmental regulation of the money supply has been at best inefficient and that free market forces would, over time, deliver better monetary institutions. Those sharing White's views might object to imposing reserve requirements on electronic intraday money as precluding the development of efficient private-sector mechanisms for regulating the supply of intraday money.

This criticism is difficult to answer, given how little the economics profession really knows about the workings of money. Because there is currently no widely accepted theoretical framework for money, it is difficult to rank systematically the "efficiency" of various monetary arrangements. Furthermore, the historical evidence concerning private banking systems is decidedly mixed. Richard H. Timberlake, Jr. (1984), Gorton (1985), and Gorton and Donald J. Mullineaux (1987) have argued that the central-bank-like features of nineteenth-century bank clearinghouses represented a spontaneous private-sector response to the various risks typically associated with banking, including the sort of systemic payment risk inherent in multilateral clearing arrangements.[36] Thus, one could argue that something akin to reserve or collateral requirements would be imposed on intraday money by the private sector.

There are several arguments that could be made in favor of increased public-sector regulation of the electronic intraday money, however. The first is the argument made above concerning the CHIPS network—that if participants in a given network expect the Fed (and other central banks) to bear the systemic risk associated with intraday netting schemes, these participants may act as if the risk were covered, irrespective of the Fed's actual policy stance on the issue. If these expectations are backed by similar ones on the part of the general public, there would be greater difficulty in dissociating Fed policy from any implied guarantee against systemic crises. In other words, simple disclaimers of responsibility for systemic risk are not likely to be credible if such disclaimers run against the grain of public expectations. Any credible transfer of the Fed's responsibility as lender of last resort to the private sector would likely require an unambiguous statutory backing and widespread political support.

Another argument in favor of governmental, as opposed to private, regulation of these payments networks concerns the allocation of regulatory responsibility. Under a system of private regulation, would one payments network bear any responsibility for a systemic crisis on another network? In the historical example of the pre-Fed clearinghouses, it is well known that these private clearinghouses provided effective safety nets (from systemic liquidity crises) for their member banks. During the 1907 panic, however, these carefully planned private arrangements were unable to prevent a systemic crisis from developing within the closely allied but relatively unregulated trust companies.[37] If a similar "contagion" of systemic crises were to develop over more than one private payments network, questions would certainly arise about how to allocate the responsibility for managing such a crisis. It is conceivable that these allocations could be governed by private contracts, but the scope and extent of such arrangements would be unprecedented. In addition, this approach would require a credible commitment on the part of the U.S. government to limit the scope of the Fed's control over the intraday money supply.

A related problem with pre-Fed clearinghouse arrangements had to do with the distributional effects of systemic crises. During pre-Fed money panics, clearinghouse member banks would often protect themselves from bank runs by resorting to suspensions of payment (of specie or its equivalent). These suspensions constituted a reasonable (though, strictly speaking, illegal) mechanism for protecting the banking system against liquidity crises. However, the suspensions also imposed significant costs on bank depositors who were unable to convert their bank deposits into hard cash, effectively removing political support for this mechanism. Similar considerations would afflict any private arrangement for dealing with systemic crises over electronic payments networks. If it imposed large costs on a considerable segment of the population, even an economically efficient mechanism could prove politically untenable.[38]

## Conclusion

The history of money is one of ever-increasing sophistication in the technology for economizing on transactions balances. The latest milestone in this historical trend is the advent of electronic intraday money. This money is in effect created by the extension of intraday "credit" over electronic payments networks, networks whose integrity is often seen as either de jure or de facto guaranteed by the Fed or by foreign central banks.

The development of electronic intraday money is a natural outgrowth of the development of computer and communications technology necessary for electronic payment. These payments technologies offer the potential to economize on the use of scarce bank reserves and to increase payments system efficiency. Unfortunately, the legal and regulatory framework of the payments system has not kept pace with the rate of technological innovation. In particular, current technologies allow for virtually continuous payments while the laws, regulations, and conventions regarding settlement are still largely oriented toward settlement on a once-a-day basis.

An unforeseen consequence of this disparity has been the encouragement of a "bielectronic" monetary standard. That is, money in overnight transactions accounts is reservable while intraday money is not reservable but is convertible to overnight money at par. As a result, the existing regulatory framework has favored the creation of the less costly electronic intraday money in ever larger amounts. The creation of intraday money carries with it a certain amount of systemic risk. A number of policy measures have been aimed to limit the amount of such risk borne by the Fed, and it is likely that the legal and regulatory framework in this area will see additional changes. This discussion has argued that a major step toward containing the systemic risk associated with electronic payments networks would be to eliminate the artificial cost advantage associated with intraday money by creating tradable "intraday money creation rights." By encouraging the development of an intraday money market, this step would lead to more efficient allocations of intraday funds than would the continuing imposition of ad hoc limits on the creation of such funds.

The literature on electronic payments networks has tended to focus on the very complex institutional and technological characteristics of such systems. Without minimizing the complexity of such issues, it is this author's opinion that such analyses have failed to emphasize the fundamental truth that electronic payments are increasingly being used as an efficient form of money. Conceptually, electronic payments are no more innovative than were other devices that allowed for the creation of inside money: banknotes in seventeenth-century Sweden, clearinghouses in eighteenth-century London, or checks in the nineteenth-century United States. The introduction of marketable rights to the creation of electronic intraday money would be a useful first step toward moving electronic payments systems away from the realm of technocracy and into the mainstream of the marketplace.

## Appendix 1
## The Mechanics of Money Creation: Some Simple Examples

The process by which banks can create money is covered in most college courses on money and banking. However, the notion that banks can create money remains a foreign one to many people. The first example describes how a theoretical banking system can create traditional, overnight money.

Suppose that a small, closed economy has only two banks. In this simple economy, banks hold liabilities in the form of demand deposits and assets in the form of loans to customers and reserves in an account held with a central bank. There is a 10 percent reserve requirement, meaning that each bank must end the banking day with a ratio of reserves versus deposits of at least 10 percent. No interest is paid on reserves.

Suppose further that the two banks' initial balance sheets are the following:

**Bank A's Initial Balance Sheet**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $900 | Deposits | $900 |
| Reserves | $110 | Net Worth | $110 |

**Bank B's Initial Balance Sheet**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $900 | Deposits | $900 |
| Reserves | $90 | Net Worth | $90 |

Note that the aggregate M1 money supply—the total demand deposits in the economy—equals $1,800. Also note that Bank B is "loaned up," that is, Bank B's reserve holdings equal exactly 10 percent of its deposits. Bank A, however, holds an additional $20 in reserves beyond its legal requirement, or $20 in *excess reserves*. Because these reserves earn no interest, Bank A decides to loan the $20 to a creditworthy customer. When the loan is made, Bank A's assets and liabilities both expand. That is, its assets are increased by the addition of a $20 loan, and its liabilities are increased by the amount of additional funds made available to the borrowers' account, which is also $20. This means that Bank A's balance sheet can now be written as

**Bank A**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $920 | Deposits | $920 |
| Reserves | $110 | Net Worth | $110 |

Note that the money supply has expanded to $1,820. Suppose now that the Bank A customer who received the loan writes a check to a customer of Bank B. The Bank B customer deposits the check in Bank B. Initially, assume that the check clears instantaneously through a clearing system run by the central bank. In the clearing process, Bank A's balance sheet is contracted by $20, and Bank B's balance sheet is augmented by a corresponding amount. The new balance sheets for both banks are as follows:

**Bank A**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $920 | Deposits | $900 |
| Reserves | $90 | Net Worth | $110 |

**Bank B**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $900 | Deposits | $920 |
| Reserves | $110 | Net Worth | $90 |

Note that after clearing the check, the money supply has remained constant at $1,820, and that both banks continue to meet their legal reserve requirements.

While the example is highly stylized, it is relevant because it demonstrates some salient features of banking under a fractional (less than 100 percent) reserve requirement. First, it shows that the banking system can create money in the form of demand deposits. Second, it can be used to exemplify how reserve requirements place a constraint on the creation of such deposits. To illustrate, suppose that Bank B wishes to lend out its excess reserves to a creditworthy customer. However, in contrast to the earlier case with Bank A, there are only $18 of excess reserves available for Bank B to lend. If Bank B decides to lend out the $18, then the M1 money supply will expand by $18. If the Bank B customer writes a check for $18 to a Bank A customer, Bank B will lose its excess reserves to Bank A after the check clears. After clearing, Bank A can loan out its excess reserves, in the amount of 90 percent of $18, or $16.20. This process can be repeated over and over again, but the total amount of money in the economy is constrained by the reserve requirement and by the amount of available reserves to be less than or equal to $2,000, which is the amount of reserves available, or $200 divided by the reserve requirement of 10 percent or one-tenth.

Now consider a second example that is exactly the same as the first, except that the two banks decide to eliminate checks and to exchange all payments via a private electronic payments network. The rules of the payments network are as follows.

At the beginning of each business day, the network assigns each member a net debit cap, which is a limit on the amount of intraday net indebtedness to the other bank. For purposes of comparison with the first example, suppose that this limit is given by each bank's initial excess reserves. At the end of the day, the two banks settle by having the bank in the net debit position remit reserve funds via a wire system run by the central bank. Further suppose that it is also the custom of both banks to credit their customers' accounts with good funds as soon as a payment message to the customer's account is received via the EFT network.

At the beginning of a certain business day, suppose that each bank's initial balance sheet is as it was in the previous example. That is, both banks have $900 in loans and deposits, Bank A has $100 in reserves, and Bank B has $90 in reserves. According to the network rules, Bank A's net debit cap is $20, and Bank B's net debit cap is zero.

Suppose that when the network opens for business, each bank has twelve extremely creditworthy customers, each of whom would like to send $20 to a customer of the other bank. Suppose that the customers are named C1, C2, C3, . . . , C24 and that odd-numbered customers use the payment services of Bank A and even-numbered customers use Bank B. Each customer $i$ wishes to send customer $i + 1$ the sum of $20 (C24 wishes to send funds to C1). In this very predictable economy, the same set of payments are made every day. However, the order of payments is random.

Suppose that on a particular day, C1 is first in line at Bank A and wishes to wire $20 to C2's account at Bank B. Customer C1 has no deposits at Bank A but has a $20 line of credit against some good collateral. Further, C1 expects a $20 payment sometime during the day from C24, who banks at Bank B. He taps his line of credit and instructs Bank A to send a $20 payment message to C2, after which the banks' balance sheets look like

**Bank A**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $920 | Deposits | $900 |
| Reserves | $110 | Due to B2 | $20 |
| | | Net Worth | $110 |

**Bank B**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $900 | Deposits | $920 |
| Reserves | $90 | Net Worth | $90 |
| Due from B1 | $20 | | |

Note that the M1 money supply has expanded by $20. Bank B's net debit position is now –$20. Because it is below its net debit cap of zero, Bank B starts processing its customers' payment orders. The first customer in line is C14, who wants to send $20 to C15, who banks with Bank A. As was the case with C1 and Bank A, C14 has no deposits with Bank B but has a $20 line of credit with Bank B that she taps and uses to send a payment message to C15, knowing that the loan will be repaid sometime later in the day by C13. After C14's payment, the balance sheets of the two banks and the money supply have again expanded by $20, as shown below.

**Bank A**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $920 | Deposits | $920 |
| Reserves | $110 | Due to B2 | $20 |
| Due from B2 | $20 | Net Worth | $110 |

**Bank B**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $920 | Deposits | $920 |
| Reserves | $90 | Due to B1 | $20 |
| Due from B1 | $20 | Net Worth | $90 |

Suppose that on this particular day, the remaining customers in line at Bank A are customers C3, C5, . . . , C23 and that the remaining customers in line at Bank B are C16, C18, . . . , C24, C2, C4, . . . , C12. None of the customers had any funds on deposit with either bank at the beginning of the day, but each had a line of credit for $20. Now consider the balance sheets of the two banks after C1, C3, . . . , C11 have sent payment messages to C2, . . . , C12; and C14, C16, . . . , C22 have sent messages to C15, . . . , C23. These are as follows:

**Bank A**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $1,020 | Deposits | $1,000 |
| Reserves | $110 | Due to B2 | $120 |
| Due from B2 | $100 | Net Worth | $110 |

**Bank B**

| Assets | | Liabilities | |
|---|---|---|---|
| Loans | $1,000 | Deposits | $1,020 |
| Reserves | $90 | Due to B1 | $100 |
| Due from B1 | $120 | Net Worth | $90 |

At this point in the business day, the money supply has expanded to $2,020, an amount exceeding the upper limit of the first example by $20. As more customers

send their payment messages, the intraday loans will start to be paid off. (The example abstracts from the fees and interest that would be charged for such loans in real life.) At the close of the business day, both banks end up with a net debit position of zero and their balance sheets look exactly as they did at the start of the day.

While this second example also is highly stylized, it is again a relevant one because it illustrates two salient features of a private electronic funds transfer system. First, it shows that such an EFT system with a netting arrangement allows for an intraday expansion of balance sheets that closely parallels the traditional process of inside money creation. Second, it shows that this expansion of balance sheets need not be constrained by traditional reserve requirements.

## Some Real-World Complications

The simple examples above abstract from several important features of real-world payments systems. While some of these features complicate the logic of the argument that intraday payments amount to money creation unconstrained by reserve requirements, none of them renders the basic point invalid. Some of these complications are considered below.

The first complication is the existence of "check float." Check float occurs when the funds associated with a check simultaneously appear as deposits on two banks' balance sheets. Float occurs as a result of the time lag associated with the check-clearing process. In certain instances, this delay can result in a customer at a payee bank having access to funds before the check for those funds has been presented to the payor bank. In such instances an expansion of bank balance sheets and transactions deposits can occur, in a fashion analogous to that described for the hypothetical wire transfer system in the second example above. Check float is not counted in the official monetary aggregates, however.

The existence of check float means that it is possible for a banking system under a fractional reserve requirement to exceed the theoretical upper bound on inside money creation (provided one is willing to include float in the definition of "money"), at least on a short-term basis. This last fact, in turn, could be interpreted to mean that the creation of intraday money via electronic payments networks is really no more problematic than the creation of check float.

There are at least two key distinctions, however, between the creation of money via check float and the creation of money via the netting of electronic payments. One distinction is a technological one. That is, it is physically much easier to run up a large net debit position on a wire transfer network than it is systematically to write checks that will result in equally large amounts of check float. This first distinction is a direct consequence of a second, more fundamental distinction between intraday EFT netting and the purposeful management of check float—the differing policy stance of the Fed and other regulatory bodies toward these formally similar phenomena.

Simplistic attempts by individuals to exploit lags in the check-clearing process are more popularly known as "check kiting," and such activities carry a criminal penalty in most instances. More subtle and systematic attempts by larger organizations to exploit the lag in check clearing are usually designated by a less ominous term, "remote disbursement." Though widely practiced, remote disbursement is officially discouraged by the Federal Reserve, and some of the more flagrant practices associated with remote disbursement have been effectively prohibited. By contrast, the netting of intraday payments has been generally tolerated (subject to certain risk controls) as a means of introducing greater efficiency into the payments system. Thus, it seems that the expansion of bank balance sheets, and the consequent creation of inside money, enjoys both technological and regulatory advantages when this activity occurs over EFT networks as compared with its occurrence by means of check float.

A second real-world complication is that reserves need not stay constant. As pointed out in Marvin Goodfriend and Monica Hargraves (1983), traditionally the Fed has chosen to accommodate short-run fluctuations in banks' demand for reserves rather than allow these demand fluctuations to affect short-term interest rates. Hence, reserve requirements have not historically posed a barrier to banks' expansion of the money supply, at least in the short run. For this reason, the first example above exaggerates the constraining effect of reserve requirements on money creation. In real life, reserves can be borrowed overnight in the Fed funds market at a rate that changes little from day to day.

It should be noted, however, that the Fed's accommodation of short-run fluctuations in reserve demand have not reduced the marginal cost of adding further reserves (and hence of creating additional reservable deposits) to zero. Further, the real marginal cost of adding reserves—the real Fed funds rate—fluctuates over the course of the business cycle and is typically highest near the business cycle peak. In the case of intraday payments over EFT networks, the marginal cost of additional payments is often negligible, given that preset caps on such payments and their associated net debit positions have not been breached. It seems unlikely, therefore, that such quantitative caps would be as effective as the existence of reserve requirements in constraining the growth of banks' balance sheets.

A third complication is that bank customers may not have free access to lines of credit, as the second example

above supposes. However, the same aggregate expansion of banks' balance sheets can be obtained by introducing additional banks to the payments network. As long as each participating bank is willing to credit its customers immediately with funds sent over the network, the example goes through.

A final complication that could affect the validity of these examples is that the electronic payments in the second example could presumably be sent over Fedwire, which would change the accounting somewhat. Because the finality of payment is guaranteed over Fedwire, any overdrafts of banks' reserve accounts would become "due to the Fed," and the "due from's" would be actual increases in each bank's reserve account. These entries would be exactly offset by new entries on the Fed's balance sheet, which would consist of the payor bank's "due from's" on the asset side and increases in the payee bank's reserve account on the liability side, as shown more formally in Appendix 2.

## Appendix 2
## The Algebra of Netting Schemes

The properties of market clearing have been studied extensively in economic theory (see, for example, Gerard Debreu 1959). However, implementation of market clearing has not been as thoroughly studied. One fairly recent, systematic study is Alfred Lorn Norman (1987). The following discussion uses a framework similar to Norman's to analyze netting schemes.

### Example 1: A Domestic Payments Network

The first abstract setting to be considered is a domestic payments network. There are $N + 1$ participants in this network, $N$ participants ("banks") plus a network manager. In the jargon of equilibrium theory, there are $N + 1$ "commodities"—that is, things that will be traded. Each commodity $i$, $i = 1, \ldots, N$ consists of (reserve account) funds to be held by bank $i$. The commodity $N + 1$ consists of funds in a settlement account with the network.

Assume that at the beginning of a business day, each bank $i$ knows the total amount of funds that all of its customers need to wire to all other banks $j \neq i$. Denote the amount of this "excess demand" of bank $i$ for commodity $j$ as $Z_{ij}$. Note that $Z_{ij} \geq 0$ and $Z_{ii} = 0$. It is assumed that transactions costs are negligible and that all banks' funds are valued at par, which in this abstract setting means that all commodities have a price equal to one. Each bank has a large endowment of commodity $i$ that is more than sufficient to meet all demands for commodity $i$ but has no endowment of any other commodity. Three types of clearing schemes can now be considered: gross settlement, settlement with bilateral netting, and settlement with multilateral netting.

**Case 0**: gross settlement, meaning no netting. In this case, the network manager does nothing other than to verify that the transfers take place. Each bank $i$ will transfer a total of $\sum_{j \neq i} Z_{ij}$ to other banks. The total funds required to clear the market will be

$$F_0 = \sum_i \sum_{j \neq i} Z_{ij}.$$

Under gross settlement, the maximum number of transfers necessary to settle is $N(N-1)$.

**Case 1**: settlement with bilateral netting. In this case, the network manager still does not play an active role. Instead, each bank $i$ first determines its net debit (credit) position vis-à-vis bank $j$, $Z_{ij} - Z_{ji}$, where $1 \leq i, j \leq N$. All banks in a net debit position then settle by paying $|Z_{ij} - Z_{ji}|$ to the bank in the net credit position. Under the assumptions of this analysis, half of the transactions that occurred under gross settlement need not occur under bilateral netting. Settlement thus requires a total amount of funds equal to

$$F_B = \sum_i \sum_{j<i} \left| Z_{ij} - Z_{ji} \right|$$

and involves at most $N(N-1)/2$ transfers.

**Case 3**: settlement with multilateral netting. In this case, the network manager presents each bank $i$ with its net credit or debit position vis-à-vis all banks in the network, which will be $\sum_{j \neq i} (Z_{ij} - Z_{ji})$. Banks with a net debit position will pay this amount to the network manager, who will use these funds to pay the amounts due net creditors. The total amount of funds needed to clear the market will be

$$F_M = (1/2) \sum_i \left| \sum_{j \neq i} \left( Z_{ij} - Z_{ji} \right) \right|,$$

and settlement will involve, at most, $N$ transactions—that is, settling the network accounts of $N$ banks.

The "economy" of netting schemes derives from the following results.

*Lemma 1*. The number of transactions involved in multilateral netting ($\leq N$) is no greater than the number of transactions for bilateral netting [$\leq N(N-1)/2$], which is no greater than the number of transactions for gross settlement [$N(N-1)$].

*Lemma 2*. $F_M \leq F_B \leq F_0$. (This equation represents just the triangle inequality.)

In words, *Lemma 1* means that fewer transactions are needed as a result of netting, and *Lemma 2* means that less cash is needed. In this simple example, if clearing is instantaneous at the beginning of the business day and settlement is made at the end of the day, the amount of daylight credit extended is $F_0$ for either type of netting scheme.

In the example above, the market is static, meaning that money has no time value, and the excess demands $Z_{ij}$ are considered fixed quantities. In real-world networks, money will have time value, and the granting of costless or underpriced daylight credit will result in an increase in the demands for funds at other banks—the $Z_{ij}$'s. In a network with bilateral or multilateral netting, a cap on bilateral credit requires that two banks $i$ and $j$ switch to gross settlement when $|Z_{ij} - Z_{ji}|$ exceeds the cap value. In a network with multilateral netting, a cap on overall credit requires that bank $i$ must switch to gross settlement with all other banks $j \neq i$ when bank $i$'s overall net debit position, $\sum_{j \neq i}(Z_{ij} - Z_{ji})$, exceeds the specified limit.

The example above also assumes that each bank $i$ has a sufficient endowment of (reserve account) funds to clear exchanges under all settlement schemes. Conceivably, these endowments could be so small as to make gross settlement impossible with only bilateral exchanges, even though settlement could occur under a netting scheme. This situation corresponds to what is known as network "gridlock."

*Arithmetic examples. Example 1*: A payments network has eleven member banks. On a particular day, each bank owes the other ten banks the gross amount of $10 million each. In this example, $F_0$ = ($10 million • 10 recipients • 11 banks) = $1.1 billion, whereas $F_B = F_M$ = $0. Under gross settlement, there are 110 settling transactions, but no transactions are required to settle under either netting scheme. *Example 2*: Each bank $i$ owes bank $i+1$ a sum of $10 million and owes bank $i-1$ a sum of $5 million. In this case $F_0$ = ($5 million + $10 million) • 11 banks = $165 million, $F_B$ = ($10 million – $5 billion) • 11 banks = $55 million, and $F_m = 0$. Under gross settlement, 22 transactions are required to settle; there are 11 transactions under bilateral netting; under multilateral netting, no funds are required to settle.

## Example 2: A Payments Network with Different Currencies

Now consider the case of an interbank payments network in which payment may be made in any of $K$ different currencies, where the $K$th currency is the "dollar." Let $P_k$ represent the market-clearing dollar price of currency $k$, where $P_K = 1$. In this network, there are $K(N+1)$ commodities—that is, accounts in $K$ different currencies at $N$ different banks plus $K$ settlement accounts with the network. Let $Z_{ij}^k$ denote bank $i$'s excess demand for the typical commodity—money in currency $k$ in an account of bank $j$. As in the previous example, banks do not need to transfer money to themselves, so $Z_{ii}^k = 0$ for all $i$ and $k$.

Associated with each bank $i$ is a "home currency" $\kappa(i)$. Initially, consider the especially unrealistic case in which each bank is well supplied with funds in every currency. One way to view the operation of the $K$-currency payments network would be to model this network as the simultaneous operation of $K$ parallel domestic payments networks. A maximum of $KN(N-1)$ transactions would occur under gross settlement, one-half that number under bilateral netting, and, at most, $KN$ transactions would occur under simultaneous multilateral settlement in all currencies.

While this example is easy to analyze, the assumption that each network participant has access to and wishes to hold a "large" stock of funds in every currency seems particularly unrealistic. In a real-world situation, each bank $i$ would prefer to hold a specialized portfolio of funds in a few currencies. In an extreme example of multilateral netting, each network participant $i$ would only settle their net credit or debit position

$$\left| \sum_k \sum_{j \neq i} P_k \left( Z_{ji}^k - Z_{ij}^k \right) \right| \{\text{expressed in dollars}\}$$

by payment in their own domestic currency $\kappa(i)$. (Recall that the prices $P_k$ by assumption clear the currency market so that no one gets stuck with an unwanted currency.) The total number of transactions involved in settlement would be $N$ while the total dollar value of funds needed to settle would be

$$F_M^I = (1/2) \sum_i \left| \sum_k \left[ \sum_{j \neq i} P_k \left( Z_{ji}^k - Z_{ij}^k \right) \right] \right|.$$

The total dollar value of funds needed to clear under gross settlement would be

$$F_0^I = \sum_i \sum_k \sum_{j \neq i} P_k Z_{ij}^k.$$

In such a situation, the overall reduction in the number of transactions required for clearing, by going to multilateral netting from gross settlement, would be from a maximum of $KN(N-1)$ to $N$.

*Arithmetic example*. Suppose, as in the previous example, that there are eleven banks in a payments network, only this time it is supposed that each bank is based in a different currency and that each bank wishes to hold domestic money only at the beginning and end of

each trading day. As before, each bank $i$ owes bank $i+1$ an amount equivalent to $10 million, and bank $i$ also owes bank $i-1$ an amount equivalent to $5 million. To settle a debt with another bank, a given bank has to make two transactions. First, it has to convert the required payment to foreign currency; second, it discharges the debt in the required currency. Under a gross settlement regime, settlement results in forty-four transactions having a total dollar value of $F_0^I = 2 \cdot (\$5 \text{ million} + \$10 \text{ million}) \cdot 11 \text{ banks} = \$330 \text{ million}$. Under bilateral netting, settlement requires twenty-two transactions having a dollar value of $F_B^I = 2 \cdot (\$10 \text{ million} - \$5 \text{ million}) \cdot 11 \text{ banks} = \$110 \text{ million}$. Under multilateral netting, no payments are needed to settle because $F_M^I = 0$.

### Example 3: A Generalized Payments Network for *K* Commodities

The framework for Example 2 can be adapted to payments for $K$ different types of goods (for example, government bonds or mortgage-backed securities). The only difference is that network participants may have endowments of any commodity, but that settlement is always in the $K$th commodity—that is, in dollars.

### Observational Equivalence of Multilateral Netting (with Guaranteed Settlement) and Gross Settlement with Daylight Overdrafts

Consider a domestic payments network, as in Example 1. Bank $i$ wishes to send payment $Z_{ij}$ to bank $j$, and the payments will be settled by multilateral netting.

Before settlement bank $i$'s balance sheet looks like the following (showing only intraday changes):

**Bank *i***

| Assets | Liabilities |
|---|---|
| $+ \max\{\sum_j Z_{ji} - \sum_j Z_{ij}, 0\}$ | $+ \max\{\sum_j Z_{ij} - \sum_j Z_{ji}, 0\}$ |
| Due from network | Due to network |
| | $- \sum_j Z_{ij}$ |
| | Debits to sending customers' accounts |
| | $+ \sum_j Z_{ji}$ |
| | Credits to receiving customers' accounts |

After settlement, bank $i$'s balance sheet will change as follows:

**Bank *i***

| Assets | Liabilities |
|---|---|
| $- \max\{\sum_j Z_{ji} - \sum_j Z_{ij}, 0\}$ | $- \max\{\sum_j Z_{ij} - \sum_j Z_{ji}, 0\}$ |
| Net due from erased | Net due to erased |
| $\sum_j (Z_{ij} - Z_{ji})$ | |
| Change in reserves | |

Under a Fedwire-type system, bank $i$'s payments may be covered either by bank $i$'s reserve balance or by overdrafting this balance. Before settlement on such a system, bank $i$'s balance sheet would be

**Bank *i***

| Assets | Liabilities |
|---|---|
| $+ \max\{-R_i, \sum_j Z_{ji} - Z_{ij}\}$ | $+ \max\{\sum_j Z_{ij} - \sum_j Z_{ji} - R_i, 0\}$ |
| Change in reserve account | Amount of daylight overdraft, if any |
| | $- \sum_j Z_{ij}$ |
| | Debit senders' accounts |
| | $+ \sum_j Z_{ji}$ |
| | Credit receivers' accounts |

After settlement, the balance sheet of bank $i$ would be

**Bank *i***

| Assets | Liabilities |
|---|---|
| $- \max\{\sum_j Z_{ij} - \sum_j Z_{ji} - R_i, 0\}$ | $- \max\{\sum_j Z_{ij} - \sum_j Z_{ji} - R_i, 0\}$ |
| Cover daylight overdraft, if necessary | Overdraft obligation erased |

There are two principal differences between the "multilateral netting" and "Fedwire" T-accounts. The first is the presence of the initial reserve balance, $R_i$, in the Fedwire accounts. However, the historical incentives of the Fedwire system have been such that banks would try to minimize $R_i$. If it is assumed that $R_i$ is "fairly small," then the remaining difference between the two

systems is that under Fedwire, net creditors would be in possession of the reserve funds "due to" them from the network, *before settlement*. By contrast, under an explicit multilateral net-settlement mechanism, net creditors would not receive funds until *after settlement*. If settlement is guaranteed, however, this difference would be inconsequential for the behavior of the network participants.

## Notes

1. According to Juncker, Summers, and Young (1991), the Fed provides settlement for more than 160 private, small-value payments netting arrangements involving checks, Automated Clearing House (ACH) transactions, and so forth. To date, the relatively small amount of intraday credit extended via these networks has not been a major policy concern. Besides CHIPS, other domestic large-value payments networks include those operated by Participants Trust Company (PTC) and by Depository Trust Company (SFDS). See Juncker, Summers, and Young (1991) for more details on the PTC and SFDS networks.
2. Of course, monetary exchange could have been carried out with fiat money (legal tender), which was unknown in Europe at the time.
3. Clough and Cole (1941, 276-77) attribute the first widespread issue of banknotes to the Bank of Stockholm in 1661. The popularity of this note issue is at least partly explained by the fact that Sweden was on a copper standard at the time.
4. Clough and Cole (1941, 493) and Braudel (1984, 606-607) date the founding of the first clearinghouse for banks, the London Clearing House, to 1773. As noted by Braudel, nonbank clearing organizations were in existence centuries before this date.
5. See for example, Timberlake (1978, 87).
6. Timberlake (1984), Gorton (1985), and Gorton and Mullineaux (1987) each provide descriptive accounts of the workings of nineteenth-century clearinghouses in the United States. Duprey and Nelson (1986) describe the Fed's efforts to introduce par checking.
7. Figures are from Bank for International Settlements (1991, 47).
8. Two excellent glossaries of terms commonly used in the literature on electronic payments are provided in Bank for International Settlements (1989) and Gilbert (1992).
9. Similar examples can be found for domestic clearinghouses in Juncker, Summers, and Young (1991) and for cross-border clearing arrangements in Gilbert (1992). Readers interested in the mathematical details are referred to Appendix 2.
10. "De facto multilateral netting" means that the behavior of Fedwire participants is essentially the same as if Fedwire were a multilateral net settlement system, which it is not. Garber and Weisbrod (1992, 300-302) discuss the behavioral equivalence of "daylight overdrafts" and multilateral netting. A more formal discussion of this equivalence is given in Appendix 2.
11. The vast majority of daylight overdrafts (on a value basis) are incurred by large banks in the business of clearing financial-markets transactions. One recent estimate attributed 60 percent of daylight overdrafts to only three money-center banks.
12. A useful summary of basic information on Fedwire and CHIPS is provided in Bank for International Settlements (1990a).
13. There are types of risk other than systemic risk associated with private payments systems such as CHIPS. From a public policy point of view, however, systemic risk is the most important for at least two reasons. First, a truly widespread or "systemic" crisis would be the sort of risk that the private sector is least able to either control or insure against. Second, existing rules covering EFT netting arrangements typically do not provide a complete set of contingent rules in the event of a systemic crisis. See, for example, Federal Reserve Bank of New York (1991) for a more complete description of the CHIPS risk management procedures or Stehm's (1992) description of risk management on the Participants Trust Company network.
14. See Federal Reserve Bank of New York (1991).
15. The classification of Fedwire overdrafts as inside or outside money is somewhat problematic. The accepted definition of outside (inside) money is money that does (does not) represent a *net* claim of the private sector against another party outside the private sector (see, for example, Sargent 1987, 103, or Gurley and Shaw, 1960, 73). Any Fedwire payment that is funded by a daylight overdraft causes the instantaneous creation of a claim by the payee bank against the public sector (the Fed). In this sense, Fedwire overdrafts resemble more traditional forms of outside money such as overnight reserves. Under normal circumstances, however, the claim against the Fed caused by a daylight overdraft is exactly offset by a claim of the Fed on the payor, a claim that must be paid at par by the end of the business day. In this "expectational" sense, no net liability has been created, and daylight overdrafts more closely resemble inside money. Another "inside" feature of daylight overdrafts is that they are automatically created (up to the amount of any quantitative cap) at the behest of banks and their customers, at a negligible marginal cost. By contrast, the more traditional Fed liabilities commonly equated to "outside money," such

as overnight bank reserves, are under explicit control of the Fed. The Fed may choose to accommodate fluctuations in the demand for overnight reserves, but it also exerts a high degree of control over both the amount of the accommodation and the (typically nonnegligible) price charged. On balance, Fedwire daylight overdrafts seem more "inside" than "outside."

16. More detailed treatments of the issues surrounding cross-border networks can be found in Bank for International Settlements (1989, 1990b) and Gilbert (1992).
17. Estimate by Bank for International Settlements (1993).
18. This risk is referred to as Herstatt risk, after the 1974 failure of a German firm, Bankhaus Herstatt. A detailed study of Herstatt risk in the foreign exchange markets is presented in Kamata (1990).
19. McAndrews (1992) presents a more detailed treatment of delivery-versus-payments systems.
20. The discussion of FDICIA below draws heavily on that of Wall (1993).
21. For a summary of the restrictions enacted in 1990, see Board of Governors of the Federal Reserve System (1991, 79-80).
22. For a summary of the pricing scheme for Fedwire daylight overdrafts, see *American Banker*, October 1-2, 1992.
23. A very rough estimate of the magnitude of such a shift, provided in an unpublished Federal Reserve System study, is calculated to be no more than one-third of the value of all Fedwire transfers. More precise estimates must await full implementation of pricing of daylight overdrafts.
24. Figures are from Bank for International Settlements (1990a).
25. Some representative studies from this rather large body of literature include Board of Governors of the Federal Reserve System (1988), Faulhaber, Phillips, and Santomero (1990), and Flannery (1988).
26. The "Angell" report notes that "systems for the binding netting of . . . financial obligations provide a service that is a very close substitute for the function of money as a medium of exchange" (Bank for International Settlements 1990, 7). The recognition of electronic payments as money is at least implicit in the discussions of electronic payments systems by Corrigan (1987) and Flannery (1988). In a footnote, Ettin (1988) characterizes electronic payments as money and attributes this characterization to Jeffrey Marquardt.
27. Note that EICCRs as described here are not "reserves" in the traditional sense, nor do they confer a right to borrow at the discount window. Rather, EICCRs confer on their owner the right to create intraday money, which is a close substitute for reserves.
28. Several readers of early drafts of the paper have correctly pointed out that debit card and POS transactions are currently not settled on a real-time basis. It seems likely, however, that if the float associated with the settlement on these transactions were appropriately priced, then real-time settlement would become the norm.
29. The idea of marketable emission permits was suggested by Dale (1968). Baumol and Oates (1988, chap. 12) provide a discussion of these permits and a survey of the literature that analyzes them.
30. Note that this proposal does not claim that the imposition of compensating balances on Fedwire would increase the aggregate level of systemic risk in the payments system. The claim is that imposition of reserve requirements (via compensating balances) on Fedwire would shift such risk from a network in which finality is explicitly guaranteed by the Fed to other networks. As argued above, participants in such systems may well see themselves as protected from systemic risk by an implicit Fed guarantee.
31. See, for example, chapters 10 and 11 of Timberlake (1978) for an account of the U.S. experiments with bimetallism. In a nutshell, these experiments consisted of repeated, unsuccessful attempts by the U.S. government to circulate silver coinage at a mint value above its market value. For a history of earlier experiments with bimetallism by various European countries, see Kindleberger (1984, chap. 4).
32. The validity of this statement does not require that the (market) relative prices of the two types of money equal the ratio of their production costs. As long as the market price of each type of money increases as its production cost increases, intraday money would trade above par with overnight money.
33. A notable difference between the current situation and nineteenth-century bimetallism is the direction of the mispricing of the newer form of money. The introduction of silver money was a flop because the mint, or official, value of the new money was above its market price in terms of gold. Presently, electronic intraday money has succeeded at least partly because of its official valuation at par with ordinary overnight bank money. This value is below its market price, which would be above par, as discussed above.
34. For example, Goodfriend and Hargraves (1983) offer an in-depth assessment and critique of the performance of reserve requirements as a component of monetary policy. For another treatment of the costs and benefits of reserve requirements, see chapter 13 of Garber and Weisbrod (1992).
35. Currently the Fed is required by law to pay interest on any such balances on a quarterly basis, at a rate corresponding to the average rate of return on the Fed's open market portfolio for the previous quarter. A reduction in this rate would require statutory authorization.
36. It should be mentioned that the member banks of the nineteenth-century clearinghouses were hardly unregulated institutions. In other words, one does not have to endorse a system of pure laissez-faire banking to believe that private mechanisms could deal with some of the risk associated with electronic payments networks.
37. See Tallman and Moen (1990) for an account of the trusts' role in the 1907 panic.
38. For example, Donaldson (1992, 78) contends that a common pre-Fed mechanism for dealing with bank panics, that is, suspension of payments and issue of clearinghouse certificates, often resulted in abnormally large profits for the members of the clearinghouses.

## References


Bank for International Settlements. *Report on Netting Schemes*. Basle, 1989.

_____. *Large-Value Funds Transfer Systems in the Group of Ten Countries*. Basle, 1990a.

_____. *Report of the Committee on InterBank Netting Schemes of the Central Banks of the Group of Ten Countries*. Basle, 1990b.

_____. *Statistics on Payments Systems in Eleven Developed Countries*. Basle, 1991.

Baumol, William J., and Wallace E. Oates. *The Theory of Environmental Policy*. 2d ed. Cambridge: Cambridge University Press, 1988.

Bernanke, Ben S. "Clearing and Settlement during the Crash." *Review of Financial Studies* 3 (1990): 133-51.

Board of Governors of the Federal Reserve System. *Controlling Risk in the Payments System*. Washington, D.C., 1988.

_____. *Annual Report for 1990*. Washington, D.C., 1991.

Braudel, Fernand. *The Perspective of the World*. Translated by Sîan Reynolds. New York: Harper and Row, 1984. Originally published as *Les Temps du Monde* (Paris: Librarie Armand Colin, 1979).

Brimmer, Andrew F. "Central Banking and Systemic Risks in Capital Markets." *Journal of Economic Perspectives* 3 (Spring 1989): 3-16.

Clough, Shepard Bancroft, and Charles Woolsey Cole. *Economic History of Europe*. Boston: D.C. Heath and Company, 1941.

Corrigan, E. Gerald. *Financial Market Structure: A Longer View*. Federal Reserve Bank of New York, 1987.

Corwin, Philip S., and Ian W. Macoy. "A Comprehensive Look at Electronic Payments System Risk." *Banking Expansion Reporter* 9 (February 5, 1990): 1, 8-17.

Debreu, Gerard. *Theory of Value*. New Haven: Yale University Press, 1959.

Donaldson, R. Glen. "Costly Liquidation, Interbank Trade, Bank Runs and Panics." *Journal of Financial Intermediation* 2 (1992): 59-82.

Duprey, James N., and Clarence W. Nelson. "A Visible Hand: The Fed's Involvement in the Check Payments System." Federal Reserve Bank of Minneapolis *Quarterly Review* 10 (Spring 1986): 18-29.

Eisenbeis, Robert. "Eroding Market Imperfections: Implications for Financial Intermediaries, the Payments System, and Regulatory Reform." In *Restructuring the Financial System*, 19-54. Kansas City: Federal Reserve Bank of Kansas City, 1987.

Ettin, Edward C. "Commentary on Flannery." In *Restructuring Banking and Financial Services in America*, edited by William S. Haraf, 288-95. Washington D.C.: American Enterprise Institute, 1988.

Faulhaber, Gerald R., Almarin Phillips, and Anthony M. Santomero. "Payment Risk, Network Risk, and the Role of the Fed." In *The U.S. Payments System: Efficiency, Risk and the Role of the Federal Reserve*, edited by David B. Humphrey, 197-213. Boston: Kluwer Academic Publishers, 1990.

Federal Reserve Bank of New York. *The Clearing House Interbank Payments System*. New York, 1991.

Flannery, Mark J. "Payments System Risk and Public Policy." In *Restructuring Banking and Financial Services in America*, edited by William S. Haraf and Rose Marie Kushmeider, 261-87. Washington D.C.: American Enterprise Institute, 1988.

Garber, Peter M., and Steven R. Weisbrod. *The Economics of Money, Banking, and Liquidity*. Lexington, Mass.: D.C. Heath, 1992.

Gilbert, R. Anton. "Implications of Netting Arrangements for Bank Risk in Foreign Exchange Transactions." Federal Reserve Bank of St. Louis *Review* 74 (January/February 1992): 3-16.

Goodfriend, Marvin, and Monica Hargraves. "A Historical Assessment of the Rationales and Functions of Reserve Requirements." Federal Reserve Bank of Richmond *Economic Review* (March/April 1983): 3-21.

Gorton, Gary. "Clearinghouses and the Origin of Central Banking in the United States." *Journal of Economic History* 45 (March 1985): 277-83.

Gorton, Gary, and Donald J. Mullineaux. "The Joint Production of Confidence: Endogenous Regulation and Nineteenth Century Commercial-Bank Clearinghouses." *Journal of Money, Credit and Banking* 19 (November 1987): 457-68.

Gurley, John G., and Edward S. Shaw. *Money in a Theory of Finance*. Washington, D.C.: The Brookings Institution, 1960.

Hamdani, Kausar, and John A. Wenninger. "The Macroeconomics of Supplemental Balances." Appendix to *Controlling Risk in the Payments System*. Washington, D.C.: Board of Governors of the Federal Reserve System, 1988.

Juncker, George R., Bruce J. Summers, and Florence M. Young. "A Primer on the Settlement of Payments in the United States." *Federal Reserve Bulletin* 77 (November 1991): 847-58.

Kindleberger, Charles P. *A Financial History of Western Europe*. London: Allen and Unwin, 1984.

McAndrews, James J. "Where Has All the Paper Gone? Book-Entry Delivery-against-Payments Systems." Federal Reserve Bank of Philadelphia *Business Review* (November/December 1992): 19-30.

Mengle, David L. "Legal and Regulatory Reform in Electronic Payments: An Evaluation of Payment Finality Rules." In *The U.S. Payments System: Efficiency, Risk and the Role of the Federal Reserve*, edited by David B. Humphrey, 145-80. Boston: Kluwer Academic Publishers, 1990.

Norman, Alfred Lorn. "Theory of Monetary Exchange." *Review of Economic Studies* 54 (1987): 499-517.

Sargent, Thomas J. *Macroeconomic Theory*. 2d ed. Orlando, Fla.: Academic Press, 1987.

Scott, Hal S. "Commentary on Mengle." In *The U.S. Payments System: Efficiency, Risk, and the Role of the Federal Reserve*, edited by David B. Humphrey, 181-95. Boston: Kluwer Academic Publishers, 1990.

Spufford, Peter. *Money and Its Use in Medieval Europe*. Cambridge University Press, 1988.

Stehm, Jeff. "Clearance and Settlement of Mortgage-Backed Securities through the Participants Trust Company." Federal Reserve Board Finance and Economics Discussion Series #214, November 1992.

Stigum, Marcia. *The Money Market*. 3d ed. Homewood, Illinois: Dow-Jones Irwin, 1990.

Summers, Bruce J. "Clearing and Payments Systems: The Role of the Central Bank." *Federal Reserve Bulletin* 77 (February 1991): 81-91.

Tallman, Ellis W., and Jon R. Moen. "Lessons from the Panic of 1907." Federal Reserve Bank of Atlanta *Economic Review* 75 (May/June 1990): 2-13.

Timberlake, Richard H., Jr. *The Origins of Central Banking in the United States*. Cambridge, Mass.: Harvard University Press, 1978.

_____. "The Central Banking Role of Clearinghouse Associations." *Journal of Money, Credit and Banking* 16 (February 1984): 1-15.

Vital, Christian, and David L. Mengle. "SIC: Switzerland's New Electronic Interbank Payments system." Federal Reserve Bank of Richmond *Economic Review* 74 (November/December 1988): 12-27.

Wall, Larry D. "Too-Big-to-Fail after FDICIA." Federal Reserve Bank of Atlanta *Economic Review* 78 (January/February 1993): 1-14.

White, Lawrence H. "Competitive Payments Systems and the Unit of Account." *American Economic Review* 74 (September 1984): 699-712.

_____. "What Kinds of Monetary Institutions Would a Free Market Deliver?" *The Cato Journal* 9 (Fall 1989): 367-91.

# Beyond Duration: Measuring Interest Rate Exposure

**Hugh Cohen**

While many factors contributed to the savings and loan industry's extensive losses in the 1980s, the biggest losses, those that brought on the savings and loan crisis, resulted primarily from interest rate fluctuations during the late 1970s and early 1980s (see George J. Benston and George G. Kaufman 1990). Those losses demonstrated the importance of calculating and avoiding interest rate risk for financial practitioners who fund and manage all sizes of portfolios. They also focused the attention of financial regulators, the public, and, ultimately, Congress on potential losses from interest rate risk. In the aftermath, hedging instruments and techniques have been applied more broadly.[1] Congress, in the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA), has also instructed federal bank regulators to account for interest rate risk in their risk-based capital requirements.

A simple and potentially inadequate approximation of interest rate risk exposure results from the use of a technique called "modified duration." This technique is used to gauge the changes in the value of an asset or portfolio of assets that occur in response to a parallel shift in interest rates. It thus measures the portfolio's sensitivity to interest rate fluctuations. Modified duration gauges interest sensitivity by making equal interest rate shifts at all maturities of the current term structure and revaluing a portfolio under the new (parallel) term structure.

Acceptance of modified duration as a measure of interest rate exposure can be seen in federal bank regulators' recent proposal of the method for the purpose of integrating interest rate risk exposure into risk-based capital

*The author is a senior economist in the financial section of the Atlanta Fed's research department.* *He thanks Peter Abken, David Heath, Frank King, Stephen Smith, and Larry Wall for helpful comments.*

guidelines and in a modification of that proposal discussed by the Federal Reserve Board on March 31, 1993. The joint proposal seeks to approximate an institution's exposure to interest rate changes by measuring changes in its net economic value that would result from 100 basis point parallel shifts in interest rates over a three-month period. The change in net economic value would be measured as the change in the present value of its assets minus the change in the present value of its liabilities and off-balance-sheet positions. The more recent Federal Reserve proposal adds 200 basis point shifts and a nonparallel shift based on interest rate changes over the past five years to the proposed exposure measures.[2]

In addition, modified duration's simplicity has made it a common topic in textbooks. As useful as the method may be for teaching purposes, however, it is an insufficient measure for hedging interest rate exposure in the real world. This article identifies two major problems with using modified duration for this purpose. The discussion first presents the theory underlying modified duration and illustrates its benefits as a hedging model. For the analysis a simple mock portfolio was constructed and revalued using simulated term structures. The analysis points out some of the faults of modified duration, which failed to capture major elements of interest rate exposure, and suggests more accurate measures.

## *U*nderstanding Duration

Duration is a term that is usually applied to bonds but can be used in reference to any cash-flow stream. The duration of a portfolio's cash flow may be thought of as the weighted average maturity of its securities' cash flows, where the weights are the proportion of the cash flows' present value in the current period over the total present value of the portfolio's future cash flows.

For example, consider the prices in Table 1 for $100.00 default-free securities to be paid off at a specified time in the future. The price of a zero-coupon, $100.00 face-value bond maturing in two years would be $89.96. The duration of the bond would be (2 • 89.96)/89.96 = 2.

Next, consider a $100.00 face-value bond that pays 5 percent coupons semiannually. (The bond pays $5.00 [or .05 • $100.00] in six months, one year, and one and one-half years. Additionally, the bond pays $105.00 in two years, reflecting both interest and face-value payments.) Assuming that the price of the bond is the sum of its individual payments, the price of this bond would be

$$\begin{aligned} Price &= (.05 \cdot 97.89) + (.05 \cdot 95.56) + (.05 \cdot 92.77) \\ &\quad + (1.05 \cdot 89.96) \\ &= 4.8945 + 4.778 + 4.6385 + 94.458 \\ &= 108.769. \end{aligned}$$

The duration of the bond would be

$$\begin{aligned} Duration &= [(0.5 \cdot 4.8945) + (1.0 \cdot 4.778) + (1.5 \cdot 4.6385) \\ &\quad + (2.0 \cdot 94.4586)]/108.769 \\ &= 1.87. \end{aligned}$$

The duration of a two-year zero-coupon bond is two years, and the duration of a two-year 10 percent coupon bond (5 percent semiannually) is 1.87 years, illustrating that the duration of a zero-coupon bond is the maturity of the bond and that duration decreases as the coupon rate increases. (Duration declines as the proportion of the total income stream paid early increases.) The box on page 30 shows that a cash flow's duration is the sensitivity of its present value to a parallel shift in interest rates. The implication, therefore, is that the price of the zero-coupon bond is more sensitive to parallel shifts in interest rates than the price of the 10 percent bond is. This concept is important in hedging interest rate exposure. Moreover, if the cash flow's duration is zero, the cash flow will not change value in response to a small parallel change in interest rates. In other words, when the duration of a cash flow is zero, the present value of the cash flow is hedged against small parallel movements in the term structure. (See the box for a more complete discussion of duration.)

**Table 1**
**Prices of $100.00 Default-Free Securities**

| Years until Maturity | Price of $100.00 Bond |
|---|---|
| 0.5 | $97.59 |
| 1.0 | $95.56 |
| 1.5 | $92.77 |
| 2.0 | $89.96 |

## *Hedging with Duration*

A simple example will illustrate the process of hedging with duration. Consider a portfolio on July 1, 1992, that consisted of receiving $100.00 on July 1 of each year from 1993 through 1996 (face value: $400.00). According to the term structure constructed from the July 1, 1992, *Wall Street Journal*, this portfolio would have the price and duration depicted in Table 2.

The first column shows the date of payment, and the second column lists its present value. The sum of the second column is the portfolio's price. The third column is the time (years) remaining until the payment date. The fourth column weights the time into the future, multiplying it by the payment's price and dividing that figure by the total portfolio price. The sum of the weighted times is the duration of the portfolio. The fifth column is the new price of the payments if the term structure were shifted up by 1 basis point.[3]

Given that the portfolio's price changed with the shift in interest rates, is it possible to find a single cash flow that would hedge the portfolio's present value to this shift? One hedging instrument would be a single cash flow with a duration of 2.42 years (for simplicity approximated as 2.5 years) and a face value of $350.51. Because the duration of a single cash flow is the maturity of the cash flow, this security would be one that would mature on January 1, 1994. According to the term structure on July 1, 1992, a $398.81 face-value security maturing January 1, 1994, would be priced at $350.51. If the term structure were shifted 1 basis point higher, the new price of the cash flow would decrease to $350.43. The two asset prices change by the same amount with the shift in interest rates. Thus, the present value of the cash flow of the four-year portfolio can be hedged for small parallel movements of the term structure by shorting, or selling, the single cash-flow security that would mature in two and one-half years. Table 3 illustrates the benefits of using duration as a hedging tool.

A second example of hedging with duration involves a portfolio with a greater duration. In the interest of simplicity the example analyzes only default-free, fixed-income securities. A security is constructed to

**Table 2**
**The Cash Flow Portfolio of a Four-Year Security***

| Date of Payment | Price | Years until Payment | Weighted Time | Adjusted Price (+1 basis point) |
|---|---|---|---|---|
| July 1993 | $96.03 | 1.0 | 0.27 | $96.02 |
| July 1994 | $90.87 | 2.0 | 0.52 | $90.85 |
| July 1995 | $84.94 | 3.0 | 0.73 | $84.92 |
| July 1996 | $78.67 | 4.0 | 0.90 | $78.64 |
| Total | $350.51 | | 2.42 | $350.43 |

**The portfolio receives $100.00 on each July 1 from 1993 through 1996. The term structure is constructed from the July 1, 1992,* Wall Street Journal.

**Table 3**
**A Portfolio Hedged with a Single Cash Flow**

| Asset | Current Price | Adjusted Price | Difference |
|---|---|---|---|
| Long 4-Year Security | $350.51 | $350.43 | +$0.08 |
| Short 2.5-Year Security | −$350.51 | −$350.43 | −$0.08 |
| Combined Portfolio | 0.0 | 0.0 | 0.0 |

resemble a thirty-year mortgage. However, again for simplicity, the prepayment option and default risk are not included and only biannual payments are considered. Specifically, at time July 1, 1992 (the beginning of the third quarter), a cash flow is considered that consists of $100.00 payments on January 1 and on July 1 in the years from 1993 through 2022 (a face value of $6,000.00). Using a term structure of interest rates constructed from the prices of stripped Treasury bonds as reported in the *Wall Street Journal* on July 1, 1992, this security had a market price of $2,316.38 and a duration of 9.54 years.

To hedge the price of this security, a bond with a single payment on January 1, 2002 (duration 9.5 years), was selected. Using the same term structure, a face value of $4,714.52 maturing on January 1, 2002, was calculated as having a market price of $2,316.38. Thus, this security was chosen as the hedging instrument. Imagine a portfolio that is long the thirty-year security and short the nine-and-one-half-year security.[4] Such a portfolio would have a face value of zero and a duration of approximately zero. To demonstrate the usefulness of matching duration, a 1 basis point parallel shift increase to the entire term structure was implemented, and the securities were repriced. After the shift, the thirty-year security has a market price of $2,314.17 and the nine-and-one-half-year security has a market price of $2,314.17. Thus, even though the securities' prices have changed by $2.21 (.1 percent), the price of the portfolio is unchanged. Table 4 illustrates how matching the duration of a portfolio can hedge the portfolio to small parallel shifts of the term structure.

## Testing Parallel Shift Simulations

Users of duration-based models realize that the models are useful only for small movements in the term structure. However, interest rates in the United States may become very volatile in relatively short periods of time. To capture a more realistic measure of parallel movement interest rate exposure over three months, many practitioners simulate larger parallel shift movements. This study continues the previous example of a portfolio that is long the thirty-year security and short the nine-and-one-half-year security, altering the July 1 term structure by plus and minus 100 basis points throughout the curve (as in the interagency proposal cited earlier), and revaluing the securities with this new term structure. For a 100 basis point parallel shift increase in interest rates the portfolio price was +$4.99. For a 100 basis point decrease in rates the portfolio price was +$8.29. This analysis indicates that the portfolio faces little interest rate exposure. In fact, for any significant parallel shift in the term structure, the price of the portfolio increases. Thus, modified duration indicates that there should be no concern about losses from interest rate fluctuation.

As a test of this measure's accuracy, the portfolio price was recalculated using the actual term structure constructed from the stripped Treasury bond prices reported three months later, on October 1, 1992—and the difference in the price of the portfolio was −$54.75 (see Table 5). Modified duration would have grossly underestimated the actual interest rate exposure of the simplest portfolio during the third quarter of 1992. There are two important possible sources of such results: mismatched convexity and nonparallel term structure movements.

**Adjusting for Convexity**. While duration is the amount the price of a portfolio will change for small parallel movements in the term structure, convexity is how much duration will change for small parallel shifts in the term structure.[5] Thus, if durations are matched and convexities are not, the portfolio prices are hedged only to small changes in the term structure. After a small shift the durations would no longer be

**Table 4**
**A Portfolio Hedged with Matching Durations**

| Asset | Current Price | Adjusted Price | Difference |
|---|---|---|---|
| Long 30-Year Security | $2,316.38 | $2,314.17 | +$2.21 |
| Short 9.5-Year Security | −$2,316.38 | −$2,314.17 | −$2.21 |
| Combined Portfolio | 0.0 | 0.0 | 0.0 |

matched, and in the event of a larger parallel shift the portfolio prices would no longer be hedged.

The examples discussed demonstrate the results of unmatched convexity. Recall that the portfolios were perfectly hedged for a 1 basis point increase in the term structure but that their prices differed for a 100 basis point shift. Unmatched convexity is clearly evident in Table 6, in which the portfolio is priced for a 200 basis point shift. Compared with the price changes for a 100 basis point shift (+$4.99 to +$8.27), the price changes for a 200 basis point shift (+$19.21 to +$35.09) seem to indicate a nonlinear increase in the magnitude of the differences with the size of the parallel movement increases.

Eliminating convexity errors would be the first suggested improvement in simulating 100 basis point parallel shifts. This step is taken in the Federal Reserve's revised proposal, where simulations of 200 basis point shifts are included. Such shifts approximate two standard deviations of historical volatility. Because convexity errors can be large, at least two standard deviations should be simulated.[6]

Incorporating convexity clearly improves the accuracy of duration-based models. However, in the example above convexity was not a problem. Movements exceeding 100 basis points would have shown profits in the portfolio. Recall that the portfolio had a large positive price difference for both a 200 basis point increase and a 200 basis point decrease.

**Nonparallel Shifts in the Term Structure**. The biggest problem with using modified duration and parallel shift simulations is that term structure movements historically have rarely been parallel. Unfortunately, portfolios hedged for parallel movements of the term structure may have considerable exposure to nonparallel movements. A statistical technique called principal component analysis is a useful tool for illustrating this point. Principal component analysis breaks down a sequence of random motions into its most dominant independent components, with the first principal component being the most dominant, or most often occurring, component in the random sequence. The second principal component is the next dominant component after removing the first one. Chart 1 shows the two largest principal components of historical forward interest rate volatility.[7] In the chart the first principal component of forward interest rate fluctuation is similar to a parallel shift in that the entire curve moves in the same direction. Observe, however, that short-term rates are more volatile than long-term rates (a point missed by parallel shift simulation). This characteristic is similar to the nonparallel

**Table 5**
**Simulating a Portfolio under a 100 Basis Point Shift**

| Simulation | Price of the 30-Year Security | Price of the 9.5-Year Security | Difference |
|---|---|---|---|
| +100 Basis Points | $2,111.24 | $2,106.25 | +$4.99 |
| −100 Basis Points | $2,555.74 | $2,547.47 | +$8.27 |
| Actual Outcome | $2,475.81 | $2,530.56 | −$54.75 |

**Table 6**
**Simulating a Portfolio under a 200 Basis Point Shift**

| Simulation | Price of the 30-Year Security | Price of the 9.5-Year Security | Difference |
|---|---|---|---|
| +200 Basis Points | $1,934.39 | $1,915.18 | +$19.21 |
| −200 Basis Points | $2,836.71 | $2,801.62 | +$35.09 |

shift that the revised proposal discussed by the Federal Reserve Board uses for monitoring interest rate risk. The second principal component of historical forward rate movement is fundamentally different from parallel shifts. It involves "twists" of the curve, or short-term and long-term rates moving in different directions. Combined, these two principal components account for more than 98 percent of the historical interest rate fluctuation (see Robert Litterman and Jose Scheinkman 1991).

Given that historically the most likely changes in the term structure are the independent movements of its principal components, a useful measure of interest rate exposure would be the change in the portfolio price in relation to the movements resulting from possible combinations of historical principal components of term structure fluctuation. Table 7 recalculates the market price of the portfolio for simulated term structures. The term structures are the result of 0, 1, 2, and 3 standard deviation movements of the historical principal components. The row number is the number of standard deviations of the first principal component. (For example, +1 in the row means that the term structure was raised by one standard deviation of the first principal component, and −2 means that the term structure was lowered by two standard deviations of the first principal component.) The column number is the number of standard deviations of the second principal component (so that +1 in the column means that the term structure was steepened by one standard deviation of the second principal component and −2 means that it was flattened by two standard deviations of the second principal component, assuming the curve was initially steep). All standard deviations are for a three-month period. For every simulated term structure the profit/loss of the portfolio is calculated.

The greatest portfolio loss arising from the combinations of the first two historical principal components is −$59.18. This simulated loss is close to the actual loss of −$54.75 (see Table 5). Simulating more than parallel shifts indicates that the actual loss should not have been unexpected. Use of only parallel shift simulations was misleading as to the size of, and even the existence of, possible losses.[8] It is important to note that using historical interest rate fluctuations does not require any reporting information about the securities beyond what is required for modified duration; it simply requires the user to simulate more than parallel shift scenarios. Thus, better information is available at no additional cost.

In order to compare the different portfolios' exposure, the simulated portfolio values can be combined

**Chart 1**
**Principal Components of Historical Volatility**

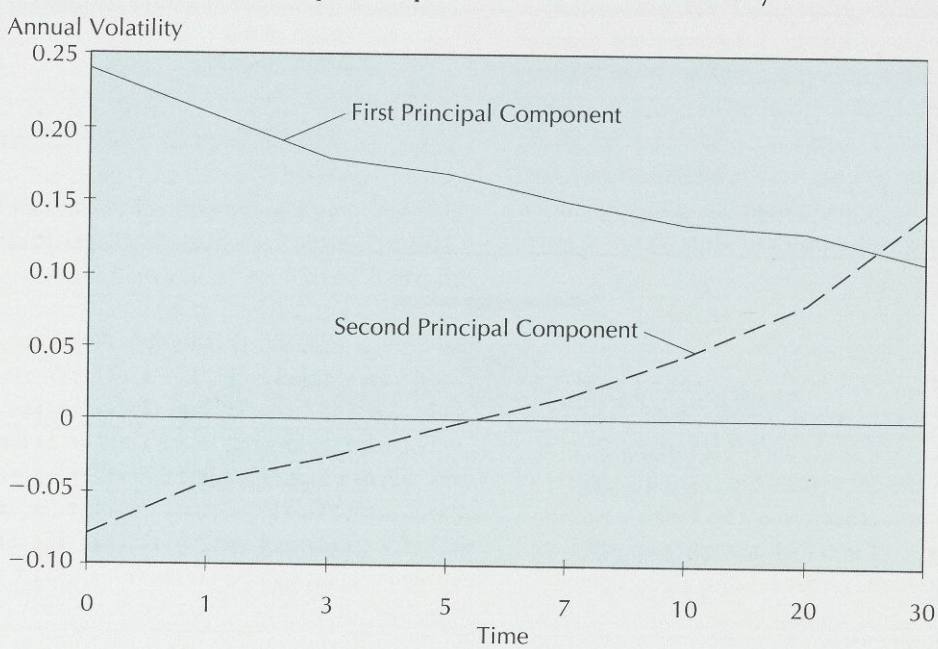

**Table 7**
**Simulated Portfolio Values***

| Standard Deviations Of First Principal Component | Standard Deviations of Second Principal Component | | | | | | |
|---|---|---|---|---|---|---|---|
| | –3 | –2 | –1 | 0 | 1 | 2 | 3 |
| –3 | +53.99 | +35.30 | +16.79 | –1.54 | –19.70 | –37.69 | –55.53 |
| –2 | +46.73 | +28.69 | +10.82 | –6.90 | –24.47 | –41.90 | **–59.18** |
| –1 | +45.59 | +28.26 | +11.07 | –5.97 | –22.89 | –39.68 | –56.35 |
| 0 | +49.34 | +32.77 | +16.33 | 0.00 | –16.22 | –32.33 | –48.35 |
| 1 | +56.98 | +41.20 | +25.53 | +9.95 | –5.54 | –20.94 | –36.27 |
| 2 | +67.63 | +52.67 | +37.79 | +22.98 | +8.25 | –6.42 | –21.03 |
| 3 | +80.58 | +66.44 | +52.37 | +38.35 | +24.39 | +10.47 | –3.41 |

* *In dollars.*

into different test statistics. For example, the loss of –$59.18—the worst-case scenario—would be a useful statistic for determining margin (or capital) for the portfolios. However, this method may still yield errors. First, although primary principal components capture more than 98 percent of the historical movements, term structures do not move exactly as historical patterns predict. Thus, there is additional "noise" that does not get simulated. Second, it is possible (although unlikely) for interest rates to move more than three standard deviations during the three-month period. For this reason, some may argue that caution calls for more than three standard deviations to be included in the simulation. Third, any number of historical principal components can be used in the simulation. Clearly, including more components reduces the amount of unmonitored interest rate risk. Performing simulations with these dimensions in mind permits a more realistic assessment of the portfolio's actual interest rate exposure and results in a statistic with a greater degree of accuracy than modified duration.[9]

In the example discussed, one may question why the zero and one standard deviation movements were included in the simulations when the big gains and losses occurred in the two and three standard deviation movements. The smaller movements were included because, when options are part of a set of securities, portfolios may exist that make money for all large movements of the term structure but lose money when the term structure is relatively stable. It is, therefore, necessary to simulate more than just the extreme outcomes. For instance, consider a portfolio consisting of long positions in a far, out-of-the-money call and put options on Treasury bond futures contracts. If interest rates fluctuate by only small amounts, all options in this portfolio would expire out-of-the-money and the original cost of the options would be lost. However, if interest rates fluctuate by a large amount in either direction, the portfolio has options that will finish in-the-money.

## Conclusion

Both Hugh Cohen (1991) and James H. Gilkeson and Stephen D. Smith (1992) show that the nature of cash flows is important in evaluating prices and risks. This article shows that the evolution of interest rate movements is also important in these evaluations. Modified duration and parallel shift simulations give useful rough approximations of interest rate exposure. However, because of the very simplicity that makes them attractive, these models have restrictions that affect their accuracy, especially over long or volatile periods of time.

This article illustrates that at the beginning of the third quarter of 1992, parallel shift simulations failed to detect the possibility of any losses to a simple portfolio, which in actuality sustained significant losses over the quarter. However, simulations based on historical term structure fluctuations, requiring no additional reporting information, would have warned the user that losses of the magnitude actually sustained were possible.

## Using Duration to Hedge Interest Rate Exposure

### Hedging with Constant Interest Rates

Consider at time 0 a default-free bond that pays $1.00 at time $T$ in the future. Assuming a constant interest rate and continuous compounding, the result is the relationship

$$b(T) = \exp(-RT), \tag{1}$$

where $R$ is the constant interest rate per unit of time, $T$ is the time in the future when the bond matures, and $b(T)$ is the price of the bond. For a coupon-paying bond,

$$Price\ of\ the\ bond = \sum_{i=1}^{i=n} CF_i \exp(-RT_i), \tag{2}$$

where $n$ is the total number of cash flows contained in the bond and $CF_i$ is the $i$th cash flow at time $T_i$. Duration, a well-known function of a bond, is defined as

$$Duration = \frac{\sum_{i=1}^{i=n} T_i CF_i \exp(-RT_i)}{Price\ of\ the\ bond}. \tag{3}$$

In words, duration is the weighted average maturity of the cash flow of the bond. Differentiating the price of a bond with respect to $R$ finds that

$$\frac{d(Price\ of\ the\ bond)}{dR} = \sum_{i=1}^{i=n} -T_i CF_i \exp(-RT_i), \tag{4}$$

which leads to the well-known relationship

$$\frac{\frac{d(Price\ of\ the\ bond)}{dR}}{Price\ of\ the\ bond} = -Duration. \tag{5}$$

In words, the percent change in a bond's price in response to an infinitesimal positive change in the constant interest rate is minus the duration. Thus, under the assumption of a flat term structure, the duration of a bond is a single number that indicates the sensitivity of the bond price to a small change in interest rates. This result can be extended for more than constant interest rates.

### Hedging with a Term Structure

Replace the assumption of a constant interest rate, $R$, with a forward interest rate curve denoted by $f(T)$. The forward interest rate is the interest rate agreed upon now at time 0 for an instantaneous default-free loan at time $T$. For example, if $f(30) = 8\%$, it is implied that the annualized interest rate on a default-free loan agreed upon today that will mature thirty years in the future and will be instantaneously repaid is 8 percent. The forward interest rate curve is the forward rate, $f(T)$, for all $T \geq 0$. The forward interest rate curve can be used to price default-free cash flows. Again, let $b(T)$ be the time 0 price of a default-free bond that pays $1.00 at time $T$, and then

$$b(T) = \exp\left[-\int_0^T f(t)dt\right]. \tag{6}$$

For a coupon-paying bond,

$$Price\ of\ the\ bond = \sum_{i=1}^{i=n} CF_i \exp\left[-\int_0^{T_i} f(t)\,dt\right]. \tag{7}$$

Duration is similarly defined as the weighted average maturity of the cash flows:

$$Duration = \frac{\sum_{i=1}^{i=n} T_i CF_i \exp\left[-\int_0^{T_i} f(t)\,dt\right]}{Price\ of\ the\ bond}. \tag{8}$$

If the price of the bond is differentiated with respect to a parallel shift in the forward rate curve [substitute $f(t) + R$ for $f(t)$ in equation 7 and differentiate with respect to R], the result as $R$ approaches 0 is

$$\frac{d(Price\ of\ the\ bond)}{dR} = \sum_{i=1}^{i=n} -T_i CF_i \exp\left[-\int_0^{T_i} f(t)dt\right]. \tag{9}$$

Substituting,

$$\frac{\frac{d(Price\ of\ the\ bond)}{dR}}{Price\ of\ the\ bond} = -Duration. \tag{10}$$

This equation demonstrates the advantages of using duration as a measure of interest rate exposure. For any forward interest rate curve, the duration of a cash flow is the sensitivity of that cash flow to a small parallel shift in the term structure. The examples in the text illustrate the benefits and limitations of hedging with duration. For small parallel fluctuations in the term structure, the portfolios are well hedged. However, for larger parallel movements or nonparallel movements, the portfolios may sustain severe losses.

The fact that 100 and 200 basis point parallel shifts failed to detect that the mock portfolio could sustain any loss owing to interest rate exposure, or that a single-factor model detected only the possibility of small losses, should be alarming for those who depend solely upon these measures to determine their interest rate exposure. Furthermore, the mock portfolio constructed is the most straightforward sort of portfolio possible, consisting of only deterministic default-free cash flows. In contrast, the set of securities available to investors in interest rate contingent claims contains extremely complex securities. Even a "simple" fixed-rate mortgage contains a complicated prepayment option. In addition, caps, floors, swaps, futures, options on futures, and countless embedded options add to the complexity of the problem. The failure to capture the true interest rate exposure of this relatively simple mock portfolio illustrates that a large amount of interest rate exposure is undetected by these measures.

The findings reported here should serve as a warning to both investors and regulators interested in determining interest rate exposure. It is important to know that oversimplified approaches to measuring interest rate exposure can be misleading, even for simple securities. Given the complex nature of securities that are common within interest rate contingent claims, the results of parallel shift and single-factor simulations should not, by themselves, be viewed as accurately reflecting interest rate exposure.

## Notes

1. One indication of this development has been the increase in the open interest of the Treasury bond futures contract. (Open interest is the number of futures contracts in existence.) Over the period from March 31, 1981, to March 31, 1993, the open interest of the nearest June futures contract increased from 51,847 to 317,804.
2. See Docket R-0764, an interagency proposal of the Federal Deposit Insurance Corporation, the Office of the Comptroller of the Currency, and the Board of Governors of the Federal Reserve System. The modified proposal presented to the Federal Reserve Board was reported in the *American Banker*, April 1, 1993, 1. It was not available in the *Federal Register* at the time of publication.
3. A basis point is 1/100 of 1 percent. If interest rates were 3 percent, a 1 basis point increase would raise them to 3.01 percent.
4. Selling a security short is equivalent to borrowing the security and selling it at its current market price with the intention of repurchasing the security at a future date and returning it to its original owner. A short seller profits when the price of the underlying security declines. Longing a security is equivalent to purchasing the security.
5. If duration is considered the first derivative of the portfolio price with respect to parallel interest rate movements, convexity would be the second derivative. For a discussion of the "convexity trap" in pricing mortgage portfolios see Gilkeson and Smith (1992).
6. The actual deviation of interest rates would lie within one standard deviation approximately 65 percent of the time. It would lie within two standard deviations approximately 95 percent of the time.
7. These components were supplied by a large financial institution in 1991.
8. Note that a one-factor historical model similar to the regulators' nonparallel shift would not have worked much better. The 0 column in Table 7 simulates only the first historical factor shifts, and the worst loss is −$6.90. Thus, two factors are the minimum number necessary for an adequate measure of this portfolio over this period.
9. If options were included in the portfolio, one would also want to simulate the effects of changes in the market's implied volatility of interest rates to the term structure simulation.

## References


Benston, George J., and George G. Kaufman. "Understanding the Savings and Loan Debacle." *Public Interest* 99 (Spring 1990): 79-95.

Cohen, Hugh. "Evaluating Embedded Options." Federal Reserve Bank of Atlanta *Economic Review* 76 (November/December 1991): 9-16.

FDIC Improvement Act of December 19, 1991, Pub. Law 102-242, 105 Stat. 2236.

Gilkeson, James D., and Stephen D. Smith. "The Convexity Trap." Federal Reserve Bank of Atlanta *Economic Review* 77 (November/December 1992): 14-27.

Interagency proposal on revising risk-based capital standards as prescribed by Section 305 of FDICIA (Docket R-0764), Press release, *Federal Register*, July 31, 1992.

Litterman, Robert, and Jose Scheinkman. "Common Factors Affecting Bond Returns." *Journal of Fixed Income* (June 1991): 54-61.

FYI

# The Use of Mitigating Factors in Bank Mergers And Acquisitions: A Decade of Antitrust At the Fed

**Christopher L. Holder**

An important aspect of the industry consolidation experienced over the past decade by the U.S. banking system is the increased pace of bank mergers and acquisitions.[1] From an average of 170 mergers per year from 1960 to 1979, the yearly average grew to 498 during the period from 1980 to 1989 (see Stephen A. Rhoades 1985a and John P. LaWare 1991). The increased number and size of bank mergers in recent years, as well as the relatively large number of bank failures, have renewed interest in how antitrust enforcement is pursued by the federal banking agencies. The federal authorities having primary responsibility for the aspects of bank mergers related to competitiveness are the Federal Reserve, the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC), and the Department of Justice (DOJ).[2]

The purpose of antitrust regulation in mergers is to prevent an acquirer from being able to exercise market power, thereby earning abnormal profits at the expense of customers within the market where the merger occurred. From a policy standpoint, a proposed merger may be denied if it carries with it the possibility of significant anticompetitive effects on prices and consumer and business welfare. The Fed's guidelines help anticipate a bank merger's effects on competition. However, a mechanical application of these guidelines, because they provide only approximations, can be misleading, and it may be appropriate to consider additional factors.

*The author is an analyst in the financial section of the Atlanta Fed's research department.* *He gratefully acknowledges comments from Jim Burke, Frank King, Bobbie McCrackin, Steve Rhoades, Aruna Srinivasan, Sheila Tschinkel, and Larry Wall.*

There are substantive, positive reasons for regulators to refrain from interfering in the market for corporate control. For example, mergers may eliminate inefficiencies or poor management. They also provide diversification and reduce excess capacity in local markets. There is a strong argument that industry consolidation is a healthy and even necessary development for U.S. banks to become stronger and remain globally competitive. With these things in mind, the Fed's approach is generally to approve mergers unless competitive effects are significantly adverse. In merger applications that apparently pose problems with regard to competitiveness the Fed looks for factors that might mitigate the anticompetitiveness implicit in a breach of its guidelines.

The essential elements in antitrust analysis of bank mergers are specification of the correct geographic and product markets, determination of all the direct and potential competitors, and the analysis of the merger's effects on the structure of individual markets. The Federal Reserve reviews these factors in a two-stage process, determining first whether a competitive problem potentially exists and, if so, whether the merger could in fact significantly affect competition adversely. The Fed's approach to identifying potential competitive problems is discussed in detail in an article in the January/February 1993 issue of this *Review*. That article examines the Fed's initial screening of proposed transactions for those that could have a significantly adverse effect on competition.

This article, the second in a two-part series detailing how the Fed deals with antitrust issues, deals with the other stage of the Fed's competitive analysis. If a proposed merger's effects exceed the Fed's structural benchmarks and the application goes to the Board of Governors of the Federal Reserve System, the Fed then seeks to determine to what extent the merger might be anticompetitive.[3] During the last decade, the Board has approved most bank merger applications it has reviewed, citing a number of mitigating factors such as competition from thrift institutions, the likelihood of new entry given the market's attractiveness, and the financial health of the firm being acquired. These and other mitigating factors cited by the Board will be the focus of this article.

## The Data

Bank merger applications dating from November 19, 1982, through December 1992 were examined in order to identify those that presented possible antitrust concerns and went to the Board of Governors for review.[4] The applications considered were filed by bank holding companies or state member banks to acquire another bank or bank holding company. (Applications from institutions that had a primary regulator other than the Fed and applications involving acquisitions of thrift institutions were not examined.) Acquisitions were judged to pose potential antitrust problems on the basis of the Board's rules regarding delegation of authority to the Reserve Banks that were applicable at the time the merger application was filed.

The Department of Justice guidelines issued in June 1982 are the foundation for the Fed's initial screening of applications.[5] The guidelines discussed market concentration in terms of the Herfindahl-Hirschman Index (HHI) and established three postmerger HHI concentration ranges for considering the likelihood that a particular acquisition would have significant anticompetitive effects. A postmerger HHI below 1,000 is considered unconcentrated; between 1,000 and 1,800, moderately concentrated; and higher than 1,800, highly concentrated. (For a discussion on the calculation and use of the HHI see Christopher L. Holder 1993, 28-30.) The Department of Justice stated that it was more likely than not to challenge transactions with a change in the HHI greater than 100 points in a moderately concentrated market or in a highly concentrated market. A change between 50 points and 100 points in a highly concentrated market might be challenged, depending on the postmerger market concentration, the size of the resulting increase in concentration, and the presence or absence of several other market-specific factors.

For the purposes of this article, these Department of Justice criteria were applied as stipulated in the Fed's Delegation of Authority guidelines for three distinct subperiods over the decade studied: (1) November 19, 1982, to December 1985, (2) January 1986 to June 1987, and (3) July 1987 to December 1992.[6]

A total of 155 merger applications were identified as posing potential competitive problems. Of these, sixteen involved issues of "prior common control" not relevant in most applications and were dropped from the data set. Of the remaining 139 applications, involving 297 local banking markets, applicants in 86 of these markets proposed totally divesting all of either their own or the target's branches, ensuring that the postmerger market share of the applicant was not higher than either its or the target's premerger market share.[7] These 86 markets were dropped from the data set, leaving a total of 211 local banking markets for which competitive issues potentially remained.

## Mitigating Factors

A majority of the applications involving the remaining 211 markets were approved. In justifying these approvals, the Board cited a number of factors that mitigated the potential anticompetitive effects of these transactions as indicated solely by the structural, or HHI, numbers. The following discussion examines the fifteen mitigating factors cited by the Board in reference to applications reviewed between November 1982 and December 1992. The factors are grouped here into five categories: strong remaining competition, misleading HHI, potential competition, convenience and needs considerations, and procompetitive effects on the market. Individual markets could, and often did, involve multiple mitigating factors as identified in the Board's decision.[8] (Table 1 presents a summary of the mitigating factors cited over the last decade, and Table 2 presents the results of this analysis summarized by year.)

## Strong Remaining Competition

Each of the mitigating factors cited in this category was used to indicate significant competition that was

**Table 1**
**Factors Cited during the Last Decade as Mitigating Potential Anticompetitive Effects of Bank Mergers**

| Mitigating Factor | Number of Markets | Percentage of Markets |
|---|---|---|
| Strong Remaining Competition | | |
| Thrift Competition | 113 | 53.6 |
| Numerous Remaining Competitors | 107 | 50.7 |
| Nonbank and Out-of-Market Competition | 10 | 4.7 |
| **Total** | **230** | |
| Misleading HHI | | |
| Partial Divestiture | 76 | 36.0 |
| Deposit Runoff | 4 | 1.9 |
| Total Deposits Incorrect | 3 | 1.4 |
| Passive Investment | 3 | 1.4 |
| Limited Competition | 3 | 1.4 |
| **Total** | **89** | |
| Potential Competition | | |
| Likelihood of Entry | 34 | 16.1 |
| Expected De Novo Entry | 1 | 0.5 |
| **Total** | **35** | |
| Convenience and Needs Considerations | | |
| Financial Health of Target Firm | 30 | 14.2 |
| No Less Anticompetitive Solution | 3 | 1.4 |
| **Total** | **33** | |
| Procompetitive Effects on Market | | |
| Benefits to Acquiring Bank | 7 | 3.3 |
| Market Share of Dominant Firm(s) | 3 | 1.4 |
| Applicant's Small Size in the Market | 1 | 0.5 |
| **Total** | **11** | |
| Denials | 6* | |

** Five merger applications, involving competition in six banking markets, were denied for competitive reasons.*

**Table 2**
**Factors Mitigating Potential Anticompetitive Effects of Bank Mergers, Summarized by Year**
*(Number of Markets)*

| Mitigating Factor | Dec. 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Strong Remaining Competition | | | | | | | | | | | | |
| Thrift Competition | 1 | 9 | 22 | 29 | 20 | 18 | 1 | 3 | 1 | 7 | 2 | 113 |
| Numerous Remaining Competitors | 0 | 8 | 9 | 17 | 16 | 19 | 5 | 3 | 8 | 8 | 14 | 107 |
| Nonbank and Out-of-Market Competition | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 2 | 4 | 1 | 10 |
| **Total** | **1** | **17** | **31** | **46** | **36** | **40** | **6** | **6** | **11** | **19** | **17** | **230** |
| HHI Misleading | | | | | | | | | | | | |
| Partial Divestiture | 0 | 4 | 3 | 2 | 5 | 0 | 9 | 1 | 3 | 6 | 43 | 76 |
| Deposit Runoff | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 |
| Total Deposits Incorrect | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 3 |
| Passive Investment | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 3 |
| Limited Competition | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 3 |
| **Total** | **0** | **5** | **4** | **4** | **5** | **1** | **10** | **1** | **6** | **6** | **47** | **89** |
| Potential Competition | | | | | | | | | | | | |
| Likelihood of Entry | 0 | 1 | 1 | 3 | 0 | 4 | 2 | 4 | 6 | 2 | 11 | 34 |
| Expected De Novo Entry | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| **Total** | **0** | **1** | **1** | **3** | **0** | **4** | **2** | **4** | **7** | **2** | **11** | **35** |
| Convenience and Needs Considerations | | | | | | | | | | | | |
| Financial Health of Target Firm | 0 | 5 | 6 | 3 | 0 | 6 | 2 | 1 | 1 | 5 | 1 | 30 |
| No Less Anticompetitive Solution | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| **Total** | **0** | **7** | **7** | **3** | **0** | **6** | **2** | **1** | **1** | **5** | **1** | **33** |
| Procompetitive Effects on Market | | | | | | | | | | | | |
| Benefits to Acquiring Bank | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 1 | 7 |
| Market Share of Dominant Firm(s) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 3 |
| Applicant's Small Size in the Market | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Total** | **0** | **1** | **1** | **0** | **0** | **6** | **0** | **0** | **1** | **0** | **2** | **11** |
| Denials | 0 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0[a] | 6[b] |

[a] *The Federal Reserve denied two applications in 1992 in which a bank holding company sought to acquire a thrift institution (see the box on page 41).*

[b] *Five merger applications, involving competition in six banking markets, were denied for competitive reasons.*

not captured by bank deposit market share data and would remain an important aspect of competition in the postmerger banking market.

**Thrift Competition**. The Board cited competition from thrift institutions more frequently than any other mitigating factor (see Holder 1993, 31).[9] The Board considered such measures as the number, size, and share of deposits held by thrifts in a market, as well as how the thrifts ranked in size within a market. The higher these measures, the more likely it was that thrifts were included as a mitigating factor. In addition, the Board also looked for evidence that thrifts were actually competing with banks by offering the full cluster of traditional banking services. Types of business and consumer transaction accounts (for example, NOW accounts), commercial and industrial loan ratios, the existence of a commercial lending department or commercial lending officers, and active advertisements for business customers were all used as evidence that thrifts were actively competing with banks.

From November 1982 through June 1987, thrifts were generally not explicitly included in HHI calculations but were considered a mitigating factor in 98 out of 116 markets, or 87.7 percent of the markets with competitive issues.[10] After June 1987 the Board automatically assigned thrifts a 50 percent weight in calculating HHIs and gave them an even higher weighting in 15 out of 95 markets.[11]

**Numerous Remaining Competitors**. The second most often-cited mitigating factor was the Board's recognition that the number of competitors remaining in a particular market after a merger was a signal about the likelihood of monopoly power developing.[12] The expectation was that remaining competitors would rise to the occasion in the event that an acquirer attempted to exercise market power through prices. (This potential is not adequately captured in the HHI because the index is a static measure of competitive structure.) Although the Board has not specified the number of competitors necessary for their presence to be considered a mitigating factor, the type of market (rural or urban, small or large deposit base) apparently played a role in this determination.[13] It appears, though, that while the existence of numerous remaining competitors was often cited as a mitigating factor, it did not play a major role in decisions regarding the transactions studied.[14] In addition, the Board sometimes noted that large statewide or regional banks having a small market share in a particular market may exert a stronger competitive influence than their small market share indicates because of their significant financial and managerial resources. Implicit in the Board's opinion is the assumption that large banks can price independently of market leaders in a particular local market because they can operate with financial support from the home office.[15]

**Nonbank and Out-of-Market Competition**. Nonbank, nonthrift financial institutions were cited as a source of competition that did not show up in the structural numbers. These financial institutions were viewed as competing with banks in a broad array of financial services, and their presence was considered a mitigating factor if they provided significant competition within a local banking market. In all, the Board referred to this mitigating factor ten times over the period under study: three times in 1987, twice in 1990, four times in 1991, and once in 1992.[16] The appearance of this mitigating factor in decisions in only the latter half of the decade is consistent with, and largely the result of, the increased competition and institutional deregulation generally experienced by the financial services industry during this period.

The most common nonbank, nonthrift competitor mentioned was credit unions—with presences specifically mentioned in six out of the ten markets. Competition from credit unions was assessed by reviewing membership requirements (liberal requirements would attract many more customers), relative and absolute size, loan-to-total-asset ratios, and business accounts offered.

Other nondepository institutions were also cited as providing significant competition for banks—including consumer and commercial finance companies, industrial loan companies, and securities brokerage firms.[17] For one market Mexican financial institutions were cited, and savings and credit union societies (in Puerto Rico) were cited in two markets.[18] Decisions on two applications acknowledged significant competition for financial services from institutions that solicited business from within a market even though they maintained no offices in the market.[19]

## *Misleading* HHI

Mitigating factors in this category were used when a mechanical interpretation of the structural numbers might be misleading. The issues raised relate to the accuracy of using the market share of total bank deposits as the sole indicator of competitive influence in a particular market, a data problem. The factors cited were partial divestiture, deposit runoff (with-

drawal of monies because of recent acquisitions), total deposits incorrect, passive investment, and limited competition.

**Partial Divestiture**. Partial divestiture, reducing an acquirer's new market share by selling some of the deposits and loans of either the applicant or bank involved, was considered a mitigating factor. In such instances, concentration numbers based on the assumption that all the deposits and loans of a target institution would be acquired misrepresented a merger's effects on competition. To compensate, the Board adjusted the HHI for the divestitures by calculating new concentration numbers reflecting the proposed sale of a bank's branches. While the divestiture was sometimes deemed adequate to correct any potential problems, in other cases additional factors played a role.

**Deposit Runoff**. Deposit runoff was a mitigating factor in two applications (four markets). Because branch-level deposit data are collected annually and as a result often do not reflect an institution's current holdings, this factor can be important. In the first market, the Board noted that the applicant had recently acquired a failed bank in the market and projected that significant deposit and loan losses would result from that acquisition, reducing the applicant's market share. The Board agreed with the applicant's contention that its competitive position as measured by deposits was overstated.[20]

A single application accounted for each of the three remaining instances of deposit runoff cited as a mitigating factor. The applicant had acquired all of its offices in the three markets by acquiring failed or failing thrifts from the Resolution Trust Corporation (RTC). Since these acquisitions, the applicant had experienced significant deposit runoff that other competitors in these markets had not been subject to. The Board concluded that the latest branch-level deposit data available overstated the competitive influence of the applicant in these markets.[21]

**Total Deposits Incorrect**. Total deposits may not be a perfect measure of competitive influence. Recognizing that fact the Board has, in two merger applications, stated that the deposits of individuals, partnerships, and corporations (IPC deposits) can be the better measure to use when calculating market concentration.[22] The Board's position is that "IPC deposits may be the proper focus of the competitive analysis in mergers and acquisitions in markets, such as those including state capitals, in which government deposits constitute a relatively large share of total deposits."[23] Because government deposits are often short-term (monies from tax collections) or must be invested in lower-yielding, relatively safe assets, they can inflate total deposit figures and be misleading.

In a third merger decision, the Board found that commercial banks in the relevant market had a substantial portion of their deposits in amounts greater than $100,000 that were predominantly short-term in nature. The applicant was cited as having almost 50 percent of its deposits in such accounts. The Board stated that these types of deposits "do not serve as a base for significant lending by banks in this market, and tend to overstate the competitive influence of banks in the market."[24] As above, the Board's conclusion was that total deposits were not the best measure of competition within this market.

**Passive Investment**. The Board has also cited the fact that in three applications the acquirer was investing passively in a bank and was not seeking control of the institution.[25] Thus, the structural changes as reflected in the HHI overstated the transaction's actual effects on competition. The Board noted that if these proposals had involved acquiring control of the bank, competition most likely would have been substantially diminished in the relevant markets. Relying on commitments that applicants would not seek to influence the bank's independent activities, the Board concluded that control of the bank would not be acquired by the applicant.

The Board pointed out, however, that one company did not need to acquire control of another to reduce competition between them. Partial ownership could dilute independence of action and encourage collusive activities. In approving the applications, therefore, the Board also noted that there would be no director interlocks among applicants and banks and that stock ownership was meant strictly as a passive investment. In two of the applications, the Board also pointed out that the bank was under the firm and active management of a family that collectively owned more than 50 percent of the outstanding stock, the implication being that the applicant's likely influence over the bank's actions would be limited.

**Limited Competition**. In some cases competition between an applicant and a target bank was already limited by their having common principals or ownership. Thus, the amount of competition actually eliminated would have been less than the HHI indicated. The Board noted in one application that the applicant's principals had formed the target bank de novo (as a newly chartered bank) in 1965 and that the applicant's shareholders already owned 77 percent of the bank.[26] In a

second application, a principal of the applicant was also a management official of the target bank.[27] In another application, it was noted that brothers owned both the applicant and the bank. Each owned stock in the other's institution in addition to having numerous other business relationships.[28] In all of these cases, the Board approved the merger.

## *Potential Competition*

In this category the Board cited likelihood of new entry into the relevant market and expected de novo entry as mitigating factors. The expectation was that future competitors could at least partially offset any current anticompetitive effects of a proposed merger.

**Likelihood of Entry**. In line with market theory, the Board has tried to assess the likelihood of new entry and its effect on competition within the market of a proposed merger. The relaxation of legal barriers to entry in a large number of states in recent years has significantly increased the pool of potential entrants into most banking markets. Recent empirical evidence supports the hypothesis that an increase in laws permitting interstate banking and statewide branching has made potential competition a more important factor in banking markets (see Dean F. Amel and J. Nellie Liang 1991). Correspondingly, the Board has cited the likelihood of new entry as a mitigating factor much more frequently in recent years. Twenty-seven of the thirty-four occasions in which potential new entry was cited as a mitigating factor occurred after July 1987, with eleven occurring in 1992.

If a market is considered attractive for entry and has few or no legal barriers (restricting branching or prohibiting entry by out-of-state bank holding companies), then new entry can be expected to lessen the possible anticompetitive effects of a merger. The Board reasons that if a bank (or banks) within a market implements noncompetitive pricing and earns greater-than-normal profits, other firms could be expected to enter the market to capture some of this excess profit, forcing more intramarket competition and a return to competitive pricing. A market is attractive for entry if (1) it can easily support a new bank or banks, (2) there are banks that are likely to expand quickly into the market, and (3) the market has certain characteristics associated with market attractiveness.

The Board delineated several characteristics that add to a market's attractiveness—large market size, urban location, rapid population and deposit growth, a relatively high ratio of population per bank or banking office, and a relatively high ratio of deposits per bank or banking office (with high ratios tending to indicate that the market is underbanked or that the population and deposits are enough to support new entrants). Higher-than-average per capita income indicates that a market is attractive as does recent de novo entry into the market. Rapid growth is especially important, making it easier for entrants to attract an adequate customer base.

In the extreme case of an unattractive, declining market, a case can be made that an institution's exit from the market is a necessary adjustment because the market can no longer support the existing number of independent institutions. This factor is generally cited when the bank is in danger of failing. In addition, declining markets are often unattractive for expansion by out-of-market firms so that an in-market merger may be the only means of preventing a bank's failure.

Empirical evidence supports considering the likelihood of entry as a mitigating factor in merger decisions. Studies have developed a fairly consistent set of variables that are positively related to the entry of firms into banking markets, either by acquisition—the more common means—or de novo. These variables include market size, market concentration, profitability, rate of growth, and the number of customers per bank, all of which have been cited by the Board. In addition, urban markets have been found to be significantly more likely to experience entry than rural markets (see Amel 1989).

A high likelihood of entry because of a market's attractiveness was cited as a mitigating factor in applications involving twenty-one markets over the sample period; twelve markets were determined to be unattractive or declining. Legal issues affecting entry played a role in nine markets, and statewide branching or permissible interstate mergers and acquisitions were cited as mitigating factors in eight of those. In one market there were legal barriers preventing branching or interstate mergers or acquisitions, and this factor weighed against approval of the proposed merger.

**Expected De Novo Entry**. The Board cited the expected de novo entry of a new competitor as a factor that mitigated any potentially anticompetitive effects in one market.[29] No further explanation of the use of this factor was given. It seems obvious, however, that the Board expected the new entrant to provide enough competition to offset at least partially any anticompetitive effects of the merger.

## Convenience and Needs Considerations

Section 3(c) of the Bank Holding Company Act (1956) specifies that in supervising bank mergers and acquisitions federal agencies must consider the convenience and needs of the community to be served. If a merger would result in a favorable impact on the convenience and needs of the community, that consideration may outweigh concerns about anticompetitive effects. In Board decisions, two factors fall into this category: the target firm's financial health and the lack of a less-anticompetitive solution.

**Financial Health of the Target Firm**. The Board cited the financial health of the target firm as a mitigating factor in a total of thirty markets. In eight of these markets, the Board was relatively certain that the bank would fail. The decision indicated that serving the convenience and needs of the community outweighed the anticompetitive effects of allowing the merger. Specific public benefits cited included uninterrupted banking service, continued operation of conveniently located offices, and maintaining employment within the community.

In the remaining twenty-two markets, the Board concluded that the bank had proven to be a weak competitor with a possibility of failing in the future and that this fact lent some weight toward approval. Citing a particular bank as a weak competitor is based on the hypothesis that deposit-share data probably overstate the firm's competitive influence in its market and thus misrepresent the anticipated anticompetitive effects of the merger. The Board cited several factors it considered in reaching its conclusion: regulatory exam results, deteriorating capital levels, past and projected earnings records, declining market share, a low loan-to-deposit ratio, small bank size, and the failure to offer the full range of banking services. Often, the failure of a weak bank to provide a full range of services to its customers is addressed in an acquiring institution's application, with the acquirer promising to improve the range and quality of the services provided to the community.

**No Less-Anticompetitive Solution**. Another issue considered in such cases is whether a failing bank has potential buyers other than the anticompetitive applicant. The presence of bidders promising less-anticompetitive effects who could also satisfy the convenience and needs considerations of the community is likely to weigh against approval of the merger. On the other hand, the lack of other potential acquirers tends to weigh heavily toward approval of the merger.

In two cases involving acquisition of either a failing bank or one that was a very weak competitor unlikely to survive on its own, the Board recognized that the merger would have some negative effects on competition but cited as a mitigating factor the absence of a better solution. In these markets, the target bank was either offered to or had attracted some interest from investors outside the market or institutions other than the applicant. However, in both applications only the applicant actually agreed to purchase the bank.[30] In a third application, the Board cited the FDIC's conclusion that no less-anticompetitive solution was available.[31]

## Procompetitive Effects on a Market

The Board has indicated that factors enhancing competition within a market work in favor of a merger application's approval. Three mitigating factors of this sort have been cited: benefits to the acquiring bank, the market share of dominant firm(s), and an applicant's small size in the market.

**Benefits to the Acquiring Bank**. The benefits expected to accrue to an acquiring bank were cited in two applications as a mitigating factor supporting approval of the application. In the four markets affected, the regional economy encompassing both the acquirer and the target was suffering an economic downturn, reflected in the operating results of the institutions involved. The Board concluded that the cost savings resulting from the merger would better position the applicants to survive this downturn.[32]

A third application citing benefits to the applicant as a mitigating factor involved an opinion by the Board that the applicant's management could gain financial and operating efficiencies through elimination of duplicate boards of directors and through the pooling of capital accounts, thus positioning itself to be a stronger competitor in the future.[33] In a fourth application, the Board concluded that because both the applicant and the target bank were small in absolute size, they might derive some economies of scale from consolidation.[34] In the final application involving this mitigating factor, the Board found that the acquisition would not disturb the competitive balance within the market, noting that after the merger five of the remaining seven institutions would have market shares greater than 10 percent. The Board concluded that the merger would result in a viable, but not dominant, competitor.[35]

**Market Share of Dominant Firm(s)**. If a high HHI for a market was caused exclusively by the large market

share of one or two firms, this factor worked in favor of mergers that involved other institutions in the market and against approval for transactions involving the dominant firm(s).[36] The implication is that the Board is more willing to approve mergers that result in a market of more nearly equal-sized competitors, thereby reducing the market power of the dominant firm(s).[37]

**Applicant's Small Size in the Market.** In one application the Board cited the relatively small size of an applicant as a mitigating factor, noting that the applicant had not increased its market share in recent years despite a significant increase in the market's deposits generally.[38] The Board also pointed out that the merger would result in only a modest increase in market concentration relative to the market's overall competitive structure. (The change in the HHI would be 265 points, which would not greatly exceed the applicable guidelines.) While the Board did not express its reasoning, one possible explanation for its decision is that the acquisition presumably would enable the acquirer to become a more effective competitor in the market. The underlying assumption would be that more evenly sized banks would increase competition within a market.

## Denials

During the past decade, the Board has denied five applications for which competitive issues were a factor in proposed state member bank and bank holding company acquisitions of another bank or bank holding company.[39] Those applications would have involved the structural changes depicted in Table 3.

Several mitigating factors were considered in these five denials, but they were not seen as overcoming the significantly adverse effects of these proposals. (Citing competitive issues, the Board has also recently denied two bank holding company applications in which those institutions were trying to acquire thrifts. For a discussion of these two denials see the box on page 41).

In all five bank acquisition denials the Board considered competition from thrifts. Including thrifts at 100 percent weight produced the structural changes shown in columns 5 and 6 of Table 3. Even after including thrifts at 100 percent weight, each merger exceeded guidelines. In addition, the Board noted in three of the cases—Pikeville National, Saver's Bancorp, and Sunwest—that the facts of the cases did not warrant 100 percent thrift inclusion.

The Board noted in three of the applications (Pennbancorp, Pikeville National, and Saver's Bancorp) that the acquirer proposed to expand the services currently being provided by the target bank. While these improvements in services apparently lent some weight toward approval, they were not enough to outweigh the potential adverse effects on competition.

The Board noted several factors working against approval of the mergers. In one denial, Pikeville National, the Board noted that significant legal barriers to entry in the market made it unlikely that new competition would mitigate the anticompetitive effects of the transaction. In another case, Saver's Bancorp, the number of competitors in the market was limited, and the Board noted that consummation of the proposal would further reduce that number. The Board also considered financial and managerial factors in the Saver's Bancorp application. Although the bank to be acquired had

**Table 3**
**Merger Applications Denied on the Basis of Competitive Issues during the Last Decade**

| Applicant | Date | Postmerger HHI | Change in HHI | Postmerger HHI (Thrifts at 100%) | Change in HHI (Thrifts at 100%) |
|---|---|---|---|---|---|
| Pennbancorp | 1983 | 3,058 | 741 | 2,024 | 435 |
| Dacotah BHC | 1984 | 2,251 | 526 | 2,016 | 461 |
| Pikeville National | 1985 | 2,573 | 526 | 2,405 | 490 |
| Saver's Bancorp | 1985 | 5,338 | 658 | 3,481 | 287 |
| Sunwest (Market #1) | 1987 | 3,738 | 868 | 1,915 | 388 |
| (Market #2) | 1987 | 5,092 | 752 | 3,642* | 513 |

* *Thrift weighting in this market is only 50 percent. The structural numbers with 100 percent thrift inclusion were not given by the Federal Reserve Board of Governors for this market.*

## Thrift Acquisition Denials

In addition to denying five bank mergers for competitive reasons since November 1982, the Board has recently denied two acquisitions of thrifts by bank holding companies. The first, an application from Norwest Corporation, was denied on April 3, 1992.[1] It involved a change in the HHI, with thrifts accorded half weight, of 565 points, to a postmerger level of 2,727. (The deposits of the thrift being acquired are accorded 100 percent weight in the calculation of the postmerger HHI.) The Board noted four decisive factors: (1) market structure, (2) potential competition, (3) financial health of the target firm, and (4) competition from credit unions. In this case, the structure of the market weighed against approval. Norwest controlled more than twice the share of the market's second-largest competitor. In addition, after consummation, Norwest would control twenty of the forty-eight branches in the market, with only one other depository institution controlling more than three branches. The Board also noted that most of the remaining depository institutions were small.

The Board also found that the market was unattractive for entry and that the merger's negative effects on competition were unlikely to be offset by new entry. The market's small size, the fact that it had not experienced a high growth rate, and the fact that no new competitors had entered the market during the previous five years were all noted by the Board in reaching its conclusion. The Board found that, owing to the financial condition of the thrift, there were public benefits to the merger, but these benefits did not clearly outweigh the likely adverse effects on competition. It also noted that the RTC had received qualified bids from prospective purchasers that did not have a significant presence in the market. In addition, although the Board considered Norwest's argument that the measures of market share did not adequately take into account competition from credit unions in the market and overstated the competitive effects of the merger, this point was not addressed in detail. The Board's decision makes it clear that this factor did not overcome the likely anticompetitive effects of the proposal.

The Board also denied an application from SouthTrust Corporation to acquire a thrift institution on July 9, 1992.[2] The proposed acquisition would have produced a change in the HHI of 672 points, to a postmerger level of 2,488, with thrifts given 50 percent weight (again, in the postmerger HHI the target thrift was accorded a 100 percent weight). Several competitive factors were important in the decision: (1) the structure of the market, (2) potential competition, (3) SouthTrust's contention that the structural numbers overstated the anticompetitive effects because the thrift did not compete with SouthTrust in several banking product lines, and (4) convenience and needs considerations.

SouthTrust contended that the large number of competitors remaining mitigated the potential anticompetitive effects of the proposed merger. However, the Board concluded that other structural factors weighed against approval. These included the fact that upon consummation SouthTrust would become the market's largest competitor with a market share more than 50 percent greater than the second-largest competitor. In addition, SouthTrust would control eight of the market's twenty-two depository institution offices with only one other firm controlling more than two offices. Most of the remaining eleven institutions would be small ones, with seven of them having market shares of less than 5 percent.

SouthTrust also suggested that recent entry made the market attractive to potential competitors. However, the Board disagreed, noting that the market was rural, small, and poor by Florida norms and had experienced slow population growth and deposit growth below the state average for rural counties. In addition, population and deposits per bank and banking office were below comparable rural markets in Florida. The Board also stated that while there had been several indirect acquisitions of branch offices in the market, there had been no de novo entry since before 1987.

SouthTrust contended that the thrift was not a competitor in several product lines, including commercial lending. SouthTrust's approach differed, however, from the traditional concept of a cluster of banking products, and the Board reaffirmed its position that the cluster concept introduced by the Supreme Court in the Philadelphia National Bank case is still the appropriate framework for analyzing the competitive effects of bank mergers.[3]

The Board also noted that potential convenience and needs benefits to the community to be served did not outweigh the expected anticompetitive effects of the proposed acquisition. The decision pointed out that the thrift was in satisfactory financial condition and was an important provider of services in the market, having, for example, an important role as a lender in the market for one-to-four-unit residential mortgages .

### Notes

1. "Letter to Norwest Corporation, April 3, 1992," *Federal Reserve Bulletin* 78 (1992): 452.
2. "SouthTrust Corporation," *Federal Reserve Bulletin* 78 (1992): 710.
3. U.S. v. Philadelphia National Bank, 374 U.S. 321 (1963).

previously suffered losses, it had improved markedly over the last few years, and the Board concluded that its prospects were favorable and that it had demonstrated its ability to remain an effective competitor.

In the first market in the Sunwest application, the Board acknowledged Sunwest's claim that various nonbank financial institutions existed in the market but concluded that the record did not clarify the extent to which other institutions competed with banks in the market. The Board noted that it would be "willing to consider any additional facts or information that Applicant may be able to submit regarding this issue."[40] In the second market involved in this application, the Board disagreed with Sunwest that the market was declining and therefore that its decline mitigated the anticompetitive effects within the market.

Each of the above denials involved some question about the correct definition of the relevant geographic market affected by the transaction, and the applicants disagreed with their Reserve Bank's market definitions. The Board noted in detail the points considered in deriving the market definitions used by the Fed in each of the above applications.

## Conclusion

The Federal Reserve analyzes the competitive effects of bank mergers in a two-stage process. First, the Fed conducts an initial screening, based largely on the Department of Justice's 1982 merger guidelines, to identify the proposed mergers that may threaten competition. Then, if a proposed merger seems to involve potential competitive issues, the Board and the Reserve Banks conduct an in-depth analysis to determine what the merger's actual effects on competition would be. The Fed's analysis over the last decade cites several factors that can mitigate a merger's potentially harmful effects on competition as indicated by the HHI.

The deregulation and innovations of nonbank financial institutions—especially thrifts—in recent years have allowed many firms to compete more directly with banks in providing financial services. The Fed now generally gives thrifts an automatic weighting of 50 percent when considering potential competitive effects from bank mergers. In addition, the removal of many legal restrictions on statewide branching and out-of-state acquisitions has decreased the anticompetitive effects of mergers in many markets by substantially increasing the likelihood of new entry. The current financial health and competitiveness of the target firm, partial divestitures, and any procompetitive effects on the market were also considered by the Board as important factors mitigating the potential anticompetitive effects of some mergers.

Most bank merger applications that fail the Fed's initial screening for potential anticompetitive effects are eventually approved by the Board. While the Fed has denied only five applications for reasons related to competition over the last decade (plus two denials of thrift institution acquisitions in 1992), antitrust considerations still play an important role in the industry's approach to consolidation. The Fed's consistent use of its guidelines in antitrust enforcement has lead to self-screening on the part of potential acquirers who can proceed with relative certainty about the Fed's reaction to a specific merger proposal. Many proposals are initially structured to include divestiture that addresses likely antitrust concerns, and an unknown number of banks are deterred from even attempting certain acquisitions. The Fed has shown that it examines transactions on a case-by-case basis and is willing to give consideration to mitigating factors unique to specific markets.

### Notes

1. Throughout this article the terms merger and acquisition are used synonymously.
2. The Federal Reserve has primary jurisdiction over mergers of state member banks and mergers and acquisitions by bank holding companies. The OCC has primary responsibility for national banks, and the FDIC oversees insured state nonmember banks. In addition, Section 18(c) of the Federal Deposit Insurance Act provides that "before acting on any application for approval of a merger transaction, the responsible agency . . . shall request reports on the competitive factors involved from the Attorney General and the other two banking agencies." Once a merger or acquisition has been approved by the appropriate federal banking agency, the DOJ, by law, has thirty days in which to file suit if it feels the transaction would violate antitrust statutes. If the DOJ does file suit, the merger is automatically stopped pending resolution of legal action.
3. Applications for mergers that seem to involve no issues of competitiveness are "delegated" to the appropriate Federal Reserve Banks for handling. If a particular transaction has potentially significant issues (competitive, legal, financial, and so forth) it is subject to extensive Board review. Au-

thority to deny an application rests solely with the Board. The criteria used to determine whether an application is delegated (processed by the Reserve Banks) or nondelegated (processed by the Board) is given in the Fed's "Rules Regarding Delegation of Authority." See Holder (1993).

4. November 19, 1982, is the date the Board first referred to the 1982 DOJ merger guidelines and the Herfindahl-Hirschman Index (HHI). See "First Bancorp of New Hampshire, Inc.," *Federal Reserve Bulletin* 78 (1982): 769. The Board's actions on applications discussed in this article that potentially posed significant competitive issues are available in the *Federal Reserve Bulletin.*
5. U.S. Department of Justice Merger Guidelines, June 14, 1982.
6. Bank-specific antitrust guidelines differed in these three subperiods. See Holder (1993, 31, 33).
7. Divestiture is considered by the federal agencies as an acceptable means of reducing potential anticompetitive effects of a proposed merger (see Holder 1993).
8. For a previous treatment of the use of mitigating factors by the Board, see Loeys (1985).
9. Academic research generally supports the inclusion of thrifts as competitors of commercial banks. See Burke, Rhoades, and Wolken (1987) and Watro (1983).
10. In eleven of these markets the Board gave thrifts 100 percent weight; in seventy-two markets, 50 percent weight; in two markets, 25 percent weight; and in one market, 15 percent weight. A thrift weighting was not specified in the remaining twelve markets.
11. In thirteen of these markets the Board gave thrifts 100 percent weight, and in two markets, 75 percent weight.
12. The Board usually cited this factor in a simple statement saying that despite elimination of a competitor, numerous banking alternatives would remain in the market.
13. In applications involving divestitures to an out-of-market competitor, the Board cited as a mitigating factor the fact that the number of independent competitors within the market would remain the same after the merger. Because the number of competitors within a market is already reflected in weighting in the calculation of the market HHI, it is not clear why the Board has considered this factor as mitigating the anticompetitive effects indicated by the market's structural numbers.
14. In 103 of the 107 applications in which the presence of numerous remaining competitors was used as a mitigating factor, other mitigating factors were also cited in the Board's decision. Only four applications that exceeded guidelines were approved with numerous remaining competitors cited as the sole mitigating factor. In each of these cases the postmerger HHI and the change in HHI did not greatly exceed applicable guidelines:

| Year | Change in HHI | Postmerger HHI |
|---|---|---|
| 1983 | 149 | 1,138 |
| 1985 | 101 | 1,474 |
| 1987 | 212 | 2,220 |
| 1987 | 269 | 1,930 |

See "1st Source Bank," *Federal Reserve Bulletin* 69 (1983): 311; "The Marine Corporation," *Federal Reserve Bulletin* 71 (1985): 262; "Houghton Financial, Inc.," *Federal Reserve Bulletin* 73 (1987): 870; and "U.S. Bancorp," *Federal Reserve Bulletin* 73 (1987): 941.

15. This mitigating factor is similar to one cited by the OCC in a November 1984 merger decision. In this transaction, the OCC argued that market shares understate the competitive influence of firms that are in the market but have most of their resources elsewhere, and "these market shares would not reflect the capacity of such firms to divert resources from the external market in response to an attempt to exercise market power in the relevant market." "Decision of the Comptroller of the Currency on the Application to merger Farmers Community Bank, State College, Pennsylvania, into Peoples National Bank of Central Pennsylvania, State College, Pennsylvania," November 5, 1984, Press Release. This argument is sometimes referred to as the "deep pockets" hypothesis.
16. "AmSouth Bancorporation," *Federal Reserve Bulletin* 73 (1987): 351; "Sunwest Financial Services, Inc.," *Federal Reserve Bulletin* 73 (1987): 463; "Hartford National Corporation," *Federal Reserve Bulletin* 73 (1987): 720; "First Union Corporation," *Federal Reserve Bulletin* 76 (1990): 83; "WM Bancorp," *Federal Reserve Bulletin* 76 (1990): 788; "BanPonce Corporation," *Federal Reserve Bulletin* 77 (1991): 43; "First Hawaiian, Inc.," *Federal Reserve Bulletin* 77 (1991): 52; and "Laredo National Bancshares, Inc.," *Federal Reserve Bulletin* 78 (1992): 139.
17. "AmSouth Bancorporation," *Federal Reserve Bulletin* 73 (1987): 351; "Sunwest Financial Services, Inc.," *Federal Reserve Bulletin* 73 (1987) 463; "Hartford National Corporation," *Federal Reserve Bulletin* 73 (1987): 720; "First Union Corporation," *Federal Reserve Bulletin* 76 (1990): 83; "WM Bancorp," *Federal Reserve Bulletin* 76 (1990): 788; "First Hawaiian, Inc.," *Federal Reserve Bulletin* 77 (1991): 52.
18. "Laredo National Bancshares, Inc.," *Federal Reserve Bulletin* 78 (1992): 139; and "BanPonce Corporation," *Federal Reserve Bulletin* 77 (1991): 43, respectively.
19. "Hartford National Corporation," *Federal Reserve Bulletin* 73 (1987): 720; and "Laredo National Bancshares, Inc.," *Federal Reserve Bulletin* 78 (1992): 139.
20. "First Tennessee National Corporation," *Federal Reserve Bulletin* 69 (1983): 298.
21. "BankAmerica Corporation," *Federal Reserve Bulletin* 78 (1992): 338.
22. "Norstar Bancorp, Inc.," *Federal Reserve Bulletin* 70 (1984): 164; and "Valley Bank of Nevada," *Federal Reserve Bulletin* 74 (1988): 67.
23. "United Bank Corporation of New York," *Federal Reserve Bulletin* 66 (1980): 61.
24. "Laredo National Bancshares, Inc.," *Federal Reserve Bulletin* 78 (1992): 139.
25. "Sun Banks, Inc.," *Federal Reserve Bulletin* 71 (1985): 243; "First State Corporation," *Federal Reserve Bulletin* 76 (1990): 376; and "SunTrust Banks, Inc.," *Federal Reserve Bulletin* 76 (1990): 542. The percentage ownerships involved in the three applications were 15 percent, 24.9 percent, and 24.99 percent, respectively.
26. "Central Wisconsin Bankshares, Inc.," *Federal Reserve Bulletin* 71 (1985): 895.

27. "Fairfax Bancshares, Inc.," *Federal Reserve Bulletin* 73 (1987): 923.
28. "Lisco State Company," *Federal Reserve Bulletin* 76 (1990): 31.
29. "Centura Banks, Inc.," *Federal Reserve Bulletin* 76 (1990): 869.
30. "Van Buren Bancorporation," *Federal Reserve Bulletin* 69 (1983): 811; and "First National Bankshares of Sheridan," *Federal Reserve Bulletin* 70 (1984): 832.
31. "Indiana Bancorp," *Federal Reserve Bulletin* 69 (1983): 913.
32. "RepublicBank Corporation," *Federal Reserve Bulletin* 73 (1987): 510; and "Alaska Mutual Bancorporation," *Federal Reserve Bulletin* 73 (1987): 921. In addition to a proposal by the applicant to raise additional capital in the Alaska Mutual Bancorporation transaction, the FDIC agreed to make a significant capital contribution to the applicant.
33. "F.S.B., Inc.," *Federal Reserve Bulletin* 78 (1992): 550.
34. "Fairfax Bancshares, Inc.," *Federal Reserve Bulletin* 73 (1987): 923. In drawing this conclusion, the Board relied on a body of empirical work indicating that there are economies of scale in banking. Academic research suggests that banks have a U-shaped cost curve that implies some scale economies. However, the scale-efficient bank size is disputed (see Bauer, Berger, and Humphrey 1992; Evanoff and Israilevich 1991; Humphrey 1990; Hunter, Timme, and Yang 1990; and Ferrier and Lovell 1990). In addition, the efficiency gains are usually small (see Berger and Humphrey 1991). Similar results have been found for thrifts (see Mester 1987).
35. "Old Kent Financial Corporation," *Federal Reserve Bulletin* 69 (1983): 102.
36. "Community Bancshares, Inc.," *Federal Reserve Bulletin* 70 (1984): 770; "Norwest Corporation," *Federal Reserve Bulletin* 76 (1990): 873; and "CB&T Financial Corporation," *Federal Reserve Bulletin* 78 (1992): 704.
37. For an empirical analysis of the results of increasing the size of fringe firms in a market see Rhoades (1985b).
38. "AmSouth Bancorporation," *Federal Reserve Bulletin* 66 (1987): 351.
39. "Pennbancorp," *Federal Reserve Bulletin* 69 (1983): 548; "Dacotah Bank Holding Company," *Federal Reserve Bulletin* 70 (1984): 347; "Pikeville National Corporation," *Federal Reserve Bulletin* 71 (1985): 240; "Saver's Bancorp, Inc.," *Federal Reserve Bulletin* 71 (1985): 579; and "Sunwest Financial Services, Inc.," *Federal Reserve Bulletin* 73 (1987): 463.
40. "Sunwest Financial Services, Inc.," *Federal Reserve Bulletin* 73 (1987): 463.

## References


Amel, Dean F. "An Empirical Investigation of Potential Competition: Evidence from the Banking Industry." In *Bank Mergers: Current Issues and Perspectives*, edited by Benton E. Gup, 29-68. Norwell: Kluwer Academic Publishers, 1989.

Amel, Dean F., and J. Nellie Liang. "The Relationship between Entry into Banking Markets and Changes in Legal Restrictions on Entry." Board of Governors of the Federal Reserve System unpublished paper, 1991.

Bauer, Paul W., Allen N. Berger, and David B. Humphrey. "Efficiency and Productivity Growth in U.S. Banking." In *The Measurement of Productive Efficiency: Techniques and Applications*, edited by H.O. Fried, C.A.K. Lovell, and S.S. Schmidt. Oxford: Oxford University Press, 1992.

Berger, Allen N., and David B. Humphrey. "The Dominance of Inefficiencies over Scale and Product Mix Economies in Banking." *Journal of Monetary Economics* 28 (August 1991): 117-48.

Burke, Jim, Stephen A. Rhoades, and John Wolken. "Thrift Institutions and Their New Powers." *Journal of Commercial Bank Lending* 69 (June 1987): 43-54.

Evanoff, Douglas D., and Philip R. Israilevich. "Scale Elasticity and Efficiency for U.S. Banks." Federal Reserve Bank of Chicago Working Paper 91-15, 1991.

Ferrier, Gary D., and C.A. Knox Lovell. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence." *Journal of Econometrics* 46 (October/November 1990): 229-45.

Holder, Christopher L. "Competitive Considerations in Bank Mergers and Acquisitions: Economic Theory, Legal Foundations, and the Fed." Federal Reserve Bank of Atlanta *Economic Review* 78 (January/February 1993): 23-36.

Humphrey, David B. "Why Do Estimates of Bank Scale Economies Differ?" Federal Reserve Bank of Richmond *Economic Review* 76 (September/October 1990): 38-50.

Hunter, William C., Stephen G. Timme, and Won Kuen Yang. "An Examination of Cost Subadditivity and Multiproduct Production in Large U.S. Banks." *Journal of Money, Credit and Banking* 22 (November 1990): 504-25.

LaWare, John P. Testimony before the Committee on Banking, Finance and Urban Affairs of the U.S. House of Representatives. September 24, 1991. Reprinted in *Federal Reserve Bulletin* 77 (November 1991): 932-48.

Loeys, Jan G. "The Fed's Use of Mitigating Competitive Factors in Bank Merger Cases." Federal Reserve Bank of Philadelphia Working Paper 85-8, August 1985.

Mester, Loretta J. "Efficient Production of Financial Services: Scale and Scope Economies." Federal Reserve Bank of Philadelphia *Business Review* (January/February 1987): 15-25.

Rhoades, Stephen A. "Mergers and Acquisitions by Commercial Banks, 1960-83." Board of Governors of the Federal Reserve System, Staff Study 142, January 1985a.

_____. "Market Performance and the Nature of a Competitive Fringe." *Journal of Economics and Business* 37 (February 1985b): 141-57.

Watro, Paul R. "Thrifts and the Competitive Analysis of Bank Mergers." Federal Reserve Bank of Cleveland *Economic Review* (Winter 1983): 13-22.

# Policy Essay

# New Tools for Regulators In a High-Tech World

**Stephen D. Smith and Sheila L. Tschinkel**

Many people believe that the deregulation of financial institutions in the 1980s has caused substantial losses to the public. At the same time, there are feelings in some quarters that the continued subsidization of particular financial institutions, along with appropriate restrictions, would enable these firms to remain viable and to continue providing low-cost credit to certain sectors of the economy. After all, when a system that seemed to work for many years was dismantled, the public faced disasters like the losses at savings and loans (which are still being paid for). While the correlation of events—the lifting of regulations and the losses that followed—appears to suggest cause and effect, it is important to look behind this relationship if we want to improve the system that delivers financial services.

This article argues that much of what has seemingly resulted from deregulation is actually caused by technological change that is outside the control of a traditional regulatory system. The resulting ability to avoid regulatory barriers at little cost generates an expansion of trading that may also expand risk. In this case, price regulation of transactions may work better than quantity regulation in achieving public policy objectives.

Traditionally, the government has designed systems that subsidize certain activities of financial institutions or has developed regulations that limit others. These policy instruments may be viewed as opposite sides of the same coin. Subsidization may be seen as establishing a very low, or zero, price for a good or service while regulation increases the price, perhaps at times to infinity.

Technology's ability to break down this type of regulatory system has been demonstrated in the savings and loan industry's massive losses in the

*Smith holds the H. Talmage Dobbs, Jr., Chair of Finance in the College of Business at Georgia State University and is a visiting scholar at the Federal Reserve Bank of Atlanta. Tschinkel is Senior Vice President and Director of Research at the Atlanta Fed.*
*The authors thank Curt Hunter, Frank King, Bobbie McCrackin, Joycelyn Woolfolk, and especially Will Roberds and Larry Wall for comments that have substantially improved the essay.*

early 1980s. At that time, the value of mortgage portfolios at thrift institutions plunged as interest rates soared to double-digit levels. Within the traditional framework, thrift portfolios were restricted primarily to mortgages (to encourage housing). The reasoning was that any losses on these assets could be offset by gains from paying below-market rates of interest on deposits. However, depositors soon learned about an alternative to holding low-interest bank and thrift accounts. Developments in computer and communications technology had allowed money market mutual funds to be created and to expand rapidly, and the money flooded out of regulated depositories into unregulated money funds. The government was left with a choice: (1) it could do nothing and thereby preserve some low-cost deposits at troubled thrifts but cripple the ability of many depositories to make new loans, or (2) it could deregulate deposit interest rates, a move that would allow viable depositories to compete in the market.[1] The eventual decision was to decontrol deposit interest rates. However, in a very real sense this deregulation did not result from ideological beliefs in the benefits of free markets. Rather, it was generated by financial markets that exploited low-cost technologies. In fact, technological advances developed in the marketplace can and almost always will be able to circumvent so-called direct regulation, which relies on regulators' ability to know the feasible set of alternative strategies available.

In the discussion that follows, we begin by providing a general discussion of transactions costs as they relate to existing regulations. To illustrate, we focus on two situations created by advances in technology that undercut regulatory attempts to control the production of a socially desirable good. The first example demonstrates the role of organized financial exchanges in "producing" the prices of financial instruments. For our purposes, an exchange is defined as an organized trading system requiring the payment of fixed costs. This definition includes physical entities (such as the New York Stock Exchange) but also incorporates over-the-counter transactions. The National Association of Securities Dealers (through the NASDAQ system) is, rather obviously, in the business of producing prices, as is the very large over-the-counter market in government securities. We argue that the franchise value of this socially useful function has been eroded by low-cost technology that allows traders to use the product (prices) without paying for it.

Our second example covers deposit insurance guarantees, which are a form of subsidization that prevents bank runs. Historically, the high cost of technology helped regulation contain the amount of the subsidy by limiting the movement of deposits.

Our goal is not to debate the merits of the myriad government subsidy programs. Rather, we argue that existing regulations no longer work in a low-trading-cost environment. Simply stated, private benefits that arise from circumventing these restrictions depend directly on the cost of technology. As the technological cost of trading drops to near zero, the number of privately beneficial trades soars, and traditional regulation can no longer limit potential public losses. We suggest, therefore, that consideration should be given to replacing the subsidy/prohibition regulatory paradigm with a user fee–based system that may more efficiently accomplish social goals. Such a system would replace direct intervention with indirect intervention. A fee is suggested as a preferable alternative to regulation seeking to limit quantities because a fee would likely cause less loss of efficiency. In international trade, for example, tariffs can in many instances be shown to be more efficient than quotas. Moreover, other examples, such as seignorage in monetary theory, suggest that a user fee may lead to fewer distortions. Imposing a user fee would not shut down markets, as direct intervention may, and therefore it would not eliminate privately beneficial sharing arrangements. It is important to note that any fee would need to be levied against particular transactions as opposed to institutions.[2]

## *T*ransactions Costs and Effective Regulation

In most economic settings, reducing transactions costs improves social welfare. Economists generally believe that the volume of trade in a security tends to increase when there exists an easily accessible, low-cost market in which new and existing holdings can be traded. To the extent that the initial holders of such claims have the option to transfer them to someone who values them more highly, this concept makes sense. When transfer costs are high, there is little effort put forth to make a "secondary" market in the asset. Conversely, as technology improves and lowers the cost of transactions, there is an ever greater incentive to create a secondary market. The transfers that result benefit the initial parties and others.

While improved efficiency in trade may be helpful when no social goods are involved, it may undermine regulatory attempts to limit access to scarce social

goods like insurance on deposits. Enhanced efficiency in trade, by definition, makes it easier to (a) avoid prohibitions on activities (that is, it makes "tax avoidance" easier) and (b) expand the growth of valuable subsidies.[3] As a result, unregulated secondary markets develop whereby private benefits are maximized and public well-being minimized in what amounts to a zero-sum game. Direct prohibitions will never be effective in this case if perfect, or nearly perfect, financial substitutes can be derived quickly in response to new regulatory guidelines.

## *P*rivate Use of Prices Obtained from Public Exchanges

The availability of publicly observable prices that aggregate the information of all traders promotes the efficient allocation of resources by agents throughout the economy. This function of prices is a critical one in a free market system, as described by F.A. Hayek (1945), but it is not a costless process. When transactions costs are high, additional trading in secondary markets is not worthwhile and the producers (exchanges) can recover costs. However, low-cost trading technology fosters growth in secondary market transactions, pulling trading away from the primary market and thereby threatening the quality of information generated because exchanges have less and less incentive to create prices.[4]

A purely "private-party" example of technology's role in the cost of such "tax" avoidance relates to the so-called gray market in listed equities. In this market traders use the information provided by organized exchanges to determine a "fair" price for the security but then avoid the exchange's trading tax (higher transactions costs) by trading off-market. The growing availability of programs that allow strategies to be formulated quickly permits traders to act on price information as it becomes available, typically through services or media organizations that broadcast the data in real time. While the exchanges may charge these organizations a fee for accessing the data, they are unable to obtain revenues from all users.

Such "free-riding" is certainly not restricted to private markets for equity securities. Swap markets, for example, have experienced tremendous growth by using costly price information generated in exchanges to create new securities. The prices that participants are willing to pay to swap interest payments, for instance, are based in large part on interest rates that prevail on the original securities. This information, which the exchanges produce, can be used by swap participants at essentially no cost. Moreover, to the extent that swaps are unregulated, the default risk engendered by these contracts is not subject to oversight by any private or public regulatory agency.[5]

The solution to the free-riding problem suggested by Ronald Coase (1960) would rely on side payments between participants to generate a socially optimal level of the activity. The "winners" in some senses would compensate the "losers" so that the latter continue to play the game. However, as Rafael Rob (1989) and others have pointed out, there is no guarantee that this "gains to trade" model assumed by Coase will work in a noncooperative framework.

## *T*rading on Government Guarantees

The government subsidizes a number of activities thought to be socially useful. One common means of subsidization is to provide guarantees to purchasers of selected financial instruments in order to encourage an activity deemed to be socially desirable, such as home ownership. Alternatively, guarantees may be used to discourage undesirable activities like bank panics. Whatever the rationale, an important part of the "guarantee package" involves the use of regulation to limit the government's risk exposure to manageable levels. However, the guarantees have true market values, and, if trading costs are low enough, holders will be willing to sell some of them to capture the value for their shareholders.

A specific example that touches the pocketbooks of all taxpayers involves the recent trend toward the securitization of assets originally held by depository intermediaries.[6] As regulators moved to impose higher capital requirements, institutions had two options available to them. The first involved issuing more capital against existing assets. The potential dilution suffered by shareholders made this alternative unattractive for many institutions. The other option, which in the past would have been prohibitively costly, involved the transfer of certain assets to third parties by creating a new security, whose promised repayments are backed by the original assets. The current low cost of such "financial engineering" makes it worthwhile for banks or thrifts that might otherwise have raised capital to instead sell high-quality assets to insurance companies and other institutions in order to meet capital requirements. Unfortunately, this process reduces the average quality of assets on the institutions' books, and these

are financed by insured deposits. The windfall from these transactions goes, of course, to the institutions' shareholders.

It will always be possible, as long as transactions costs are low enough, to create secondary market trading in a subsidy and avoid a quantity regulation. Securitization has legitimate benefits, such as added liquidity and reduced cost of diversification for banks (see, for example, Gary Gorton and George Pennacchi 1992). However, it seems that much of the bank activity in this area has been undertaken to avoid the capital tax and thus more of this financial engineering takes place than is socially optimal.

It has been argued that risk-based deposit insurance could solve this problem, although it may be difficult to specify the premium, which varies by risk, ex ante. In any case, deposit insurance premia may be viewed as user fees. Both are a form of price, as opposed to quantity, regulation.

## Some Thoughts on a User Fee Proposal

It should by now be clear that regulations designed to restrict the volume (or quantity) of particular assets or subsidies are not effective when such inexpensive trading strategies exist. Unfortunately, blanket prohibition of large classes of securities may cause more harm than good. What, then, are the options available to the public and private exchanges? It seems logical to seek the answer in price-based regulation. Growth could be limited by removing most traditional regulations and instead imposing a positive cost on selected transactions through a user fee.

While it is true, as many authors have argued (see, for example, Paul Kupiec 1992), that transactions fees, or taxes, impede the efficiency of markets, impeding various types of free-riding transactions is desirable from a social welfare point of view. An imposed cost would indeed limit the volume of, for example, off-balance-sheet securitization, thereby making it worthwhile for institutions to raise capital instead. From a taxpayer's point of view, this result is precisely the goal. A positive cost for trading in securities would inhibit the wealth-transferring trades without prohibiting asset transfers outright.[7] As noted earlier, unlike an outright prohibition this approach would not shut down markets completely. Rather, it would result in managed growth, whereby only the most efficient firms in the industry would be able to produce privately beneficial contracts. It is true that market participants could still use low-cost technology to avoid fees by moving transactions to places where fees do not apply, such as outside the United States. However, besides the fact that customer demand is likely to limit such moves, international coordination of a fee system could benefit both the United States and other countries likely to have similar problems. Although such coordination could prove difficult to accomplish—because individual countries may want to attract activity to their domestic markets—in the long run an unbounded expansion of trading could prove costly.

On the other hand, tax evasion is always a possibility. It would be less likely, however, if penalties were attached to the statute enacting the user fee. Under these conditions most transactors would comply in the same way that most people pay their taxes. Furthermore, evasion would be discouraged by the fact that avoiding the taxes would require collusion among at least two participants. Finally, provisions could be implemented that would tie a financial transaction's U.S. legal standing to payment of the fee.

Many prominent economists, including Nobel Prize–winner James Tobin, have long called for a tax on securities transactions. They argue that excessive resources are being used in financial markets that could more productively be employed elsewhere in the economy. While it may not be possible to determine whether there are too many financial transactions in total, it seems clear that limiting the growth in free-riding transactions could enhance public welfare.

To be sure, a user fee would be expected to impede the growth of transactions, including those that are viewed as socially beneficial. However, the costs of this approach may, on balance, be lower than those generated under the current system. In short, it seems that price regulation of the sort accomplished through a form of user fee may work better than traditional quantity regulation in effectively limiting public losses. This argument is an application of the theory that examines whether price regulation is better than quantity regulation. There is general agreement among theorists that, in many situations, price regulation is superior from a social point of view.[8] In the area of financial regulation, it is a topic worth further investigation by both researchers and public policy specialists.

## Conclusion

The idea of imposing a volume-based fee on financial transactions is not to be considered lightly. We are

all too familiar with regulatory strategies that often have unintended and costly effects. At the same time, the observed costs to taxpayers of uninhibited private transactions appear substantial. Indeed, the realized contingent liability of government agencies alone is enough to make the public take notice. A user fee, which would be a type of tax analogous to a toll paid when a vehicle crosses a bridge, would generate revenue that could be passed through to the institutions that are threatened by free-riding. The pass-through may involve private exchanges or a central pool of government funds such as the FDIC insurance fund. To the extent that Congress has authorized a national market for securities, some of these funds could be employed to create such an exchange. Finally, the tremendous volume of transactions suggests that even a relatively insignificant fee (for example, fractions of a basis point) could create a substantial pool of funds. These funds could prove useful in the event of unanticipated financial stress or emergency.

Imposing costs on financial transactions, rather than prohibiting trades outright, is not proposed as a panacea for limiting risks to taxpayers or protecting the rights of private producers of financial information. Thinking about this alternative scheme is, however, a start in the direction of recognizing that modern technology, combined with limited liability, has belied the old adage that "there is no such thing as a free lunch" for traders in many financial markets. Unfortunately, their lunches must be paid for by the rest of us.

## Notes

1. A third choice, placing rate controls on money market mutual funds, was not politically viable. Many small depositors did not see any reason why they should subsidize banks and thrifts by accepting below-market rates on deposits. Some of these depositors, including groups such as the gray panthers, became politically active in efforts to prevent Congress from extending interest controls to money funds.
2. While this proposal has theoretical appeal, considerable study is needed to assess its practical effects. The design of a user fee or how much revenue it could generate is also beyond the scope of this article.
3. Kane (1981) has made a similar point in terms of a regulatory "dialectic." His idea, that institutions find ways to circumvent regulation and that regulators then respond by imposing new constraints, is similar to the argument here. He contends that improvements in technology help institutions in the "cat and mouse" game. The argument here is that while technology has now made the tool kit used by regulators obsolete in the current environment, a substitute one may be possible.
4. This is an example of the so-called Grossman-Stiglitz (1980) paradox. If the returns to producing information are zero (or close to zero), traders will have no incentive to collect information, and, in turn, the informational role of prices is destroyed. Because no transactor has incentive to pay for the socially beneficial component of price information, and if there is no way to prevent free-riding, charges by exchanges will not recover costs.
5. Private exchanges, like government, try to limit, through various devices, the risk exposure of participants. Mulherin, Netter, and Overdahl (1991) provide an excellent discussion of this point.
6. There are many other examples as well. For example, Roberds (in this *Economic Review*) discusses the free-riding of private clearinghouses on the central bank. Smith (forthcoming, 1993) has brought out similar points with regard to the securities firms and the central bank.
7. To be sure, financial innovation might indeed create situations in which a fee could be avoided. While it may be possible, in theory, to generate substitutes for securities, the cost in terms of transactions or liquidity would likely be higher than paying a fee. This argument suggests that because the cost of substitution between securities is low, a fee should be broadly based as opposed to levied against a subset of securities, such as options and futures.
8. See, for example, Bhagwati (1965) for an analysis of this choice in terms of quotas versus tariffs in international trade theory. He shows that under imperfect competition a tariff system may provide superior "welfare" when compared with a quota system. His setting is, however, static, and Rotemberg and Saloner (1989) have argued that the results may be reversed in a dynamic setting. In other situations it might be socially harmful to have any fixed quota. For example, in a liquidity crisis the central bank, if limited to a quota system, might find it in everyone's interest to ignore it. (See also note 6.)

## References

Bhagwati, Jagdish. "On the Equivalence of Tariffs and Quotas." In *Trade Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler*, edited by R.E. Balduring et al. Amsterdam: North Holland, 1965.

Coase, Ronald H. "The Problem of Social Cost." *Journal of Law and Economics* 3 (October 1960): 1-44.

Gorton, Gary, and George Pennacchi. "The Opening of New Markets for Bank Assets." In *The Changing Market in Financial Services*, edited by R. Alton Gilbert. Norwell, Mass.: Kluwer Academic Publishers, 1992.

Grossman, Sanford J., and Joseph E. Stiglitz. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70 (June 1980): 393-408.

Hayek, F.A. "The Use of Knowledge in Society." *American Economic Review* 35 (September 1945): 519-30.

Kane, Edward. "Accelerating Inflation, Technological Innovation, and the Decreasing Effectiveness of Banking Regulation." *Journal of Finance* 36, no. 2 (1981): 355-67.

Kupiec, Paul. "On the Ramifications of a Securities Transaction Tax for the Function and Efficiency of Capital Markets." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, 212, October 1992.

Mulherin, J. Harold, Jeffry M. Netter, and James A. Overdahl. "Prices Are Property: The Organization of Financial Exchanges from a Transaction Cost Perspective." *Journal of Law and Economics* 34 (October 1991, part 2): 591-644.

Rob, Rafael. "Pollution Claim Settlements under Private Information." *Journal of Economic Theory* (1989): 307-33.

Roberds, William. "The Rise of Electronic Payment Networks and the Future Role of the Fed with Regard to Payment Finality." Federal Reserve Bank of Atlanta *Economic Review* 78 (March/April 1993): 1-22.

Rotemberg, Julio, and Garth Saloner. "Tariffs versus Quotas with Implicit Collusion." *Canadian Journal of Economics* 22, no. 2 (1989): 237-44.

Smith, Stephen D. "Private versus Public Efficiency in Securities Markets." *Journal of Banking and Finance* (forthcoming, 1993).

FEDERAL RESERVE BANK OF ATLANTA

Public Affairs Department
104 Marietta Street, N.W.
Atlanta, Georgia 30303-2713
(404) 521-8020