R E V I E W

Vol. 74, No. 5

September/October 1992

THE FEDERAL RESERVE BANK of ST. LOUIS

In This Issue . . .

Innovation has always been a hallmark of the financial services indus-
try. Recent innovations range from new mechanical devices and new
contractual arrangements to new operational procedures. In response to
the pace and diversity of financial innovation, researchers have studied
the economics of this innovation. One approach focuses on improving
financial products by applying the principles of finance theory to the
process of contract design in securities markets. A second approach fo-
cuses on the reasons why financial innovation occurs, examining the in-
centives for people to develop new financial contracts or technologies.

In the first article in this *Review*, "Two Faces of Financial Innovation,"
Mark D. Flood uses two case studies to illustrate how financial innova-
tions are developed to meet a perceived demand for new financial serv-
ices. The two cases presented illustrate how market forces can spell
failure for product designs that do not incorporate the principles of
financial theory and success for those that do.

* * *

Although the notion of a credit crunch is not new, its widespread use
with reference to the latest recession and recovery has been the con-
ventional wisdom in many policy circles. In the second article in this is-
sue, "The Recent Credit Crunch: The Neglected Dimensions," Kevin L.
Kliesen and John A. Tatom take a historical view of credit crunches and
their relevance to the nation's cyclical performance. They explain the
origins of the credit crunch concept and point to the consistent misap-
plication of the concept to cyclical credit market developments. For ex-
ample, they show that movements of interest rates and of interest rate
spreads in recessions generally have not provided support for the credit
crunch hypothesis. The most recent case is no exception.

According to the authors, the recent decline in the growth of credit is
mostly attributable to two factors that are inextricably linked. First, the
demand for credit, especially by business, is cyclical. When economic ac-
tivity ebbs in a recession, the demand for credit falls as well. Second,
because bank lending to business is typically short term, it is used to
finance short-term assets such as inventory. Thus when businesses ex-
pect their sales to slow or decline in a recession, they no longer need to
add to their stocks of inventory. Accordingly, business inventories and
the demand for bank credit to finance them are reduced. Similarly, in
the early stages of an economic expansion, when businesses add to in-
ventory holdings, they typically rely relatively more on internal cash
flow and other sources of financing, than on bank loans. Only later do
businesses begin to expand bank borrowing in line with their burgeoning
financial requirements.

The authors maintain that in the final analysis, the conventional wisdom espoused by credit crunch theorists is not helpful in assessing recent movements of interest rates, business loans and business inventories.

* * *

Vector autoregressive (VAR) models are frequently used to estimate dynamic relationships for economic data. Initially the VAR was not intended to describe structural mechanisms because it represents a reduced form. An economic interpretation of these empirical models, however, does require a set of structural assumptions.

In the third article in this issue, "Structural Approaches to Vector Autoregressions," John W. Keating reviews the principles of VAR analysis with particular emphasis on two structural VAR approaches. The two approaches impose restrictions on contemporaneous and long-run behavior to identify a structural model from the reduced form. The author provides empirical examples with a standard group of macroeconomic variables to illustrate these two approaches. The empirical long-run structural model yields results that are generally consistent with the theoretical model, whereas the results for the contemporaneous model are frequently inconsistent with theory. These findings suggest that long-run structural VAR models are superior; however, additional investigation should determine whether this result is robust or simply a special case.

* * *

In the final article in this *Review*, "The Misuse of the Fed's Discount Window," Anna J. Schwartz, a senior research associate at the National Bureau of Economic Research, asks not only whether the Fed's discount window continues to serve any useful function, but also whether its operation may have unintended adverse effects. The discount rate, for example, draws widespread—but in Schwartz's view, misdirected—attention as an indicator of the thrust of monetary policy. Schwartz is also concerned about the many, albeit unsuccessful, attempts to extend the Fed's lender-of-last-resort function to nonbank firms. Most important, however, is her concern that in recent years the discount window has extended credit to banks that were insolvent or near insolvency—a practice that she argues has increased taxpayer losses associated with reimbursing insured depositors at failed banks. Although many analysts will find some of her arguments controversial, Schwartz raises a number of issues worthy of further research.

* * *

Mark D. Flood

Mark D. Flood is an economist at the Federal Reserve Bank of St. Louis. David H. Kelly and James P. Kelley provided research assistance.

Two Faces of Financial Innovation

INNOVATION HAS ALWAYS been a hallmark of the financial services industry. Indeed, the history of finance can be organized as a chronicle of innovations. We can trace this history from the introduction of coinage in the Greek state of Lydia in the 7th century B.C., through the various ploys to circumvent the Christian and Islamic bans on usury in the medieval era, through the development of modern systems of insurance in the 18th and 19th centuries, on up to more timely innovations such as foreign currency exchange warrants and interest-rate swaps.[1]

Recent innovations have ranged from the commonplace — the automatic teller machine (ATM), for example — to the arcane — exchange-traded options to buy futures contracts on municipal bond index funds. Indeed, the range of innovation, from new mechanical devices to new contractual arrangements to new operational procedures, is so broad as to confound concise definition of the term.[2] Moreover, innovations have recently arrived at a frenetic pace; in the narrow field of exchange-traded futures, for example, Silber (1981) lists 102 new contracts introduced in the United States in the 1970s alone.

In response to the pace and diversity of financial innovation, economists have studied the economics of innovation. Two largely distinct — but not necessarily inconsistent — approaches to the subject have developed over the last 20 years. One approach, motivated by the imperfect success rate for new securities, has focused on improving financial products by applying the principles of finance theory to the process of contract design in securities markets.[3] The second approach, motivated largely by central bankers' concerns about the effect of innovations on monetary policy, has focused on the reasons why financial innovation occurs. This approach has examined the incentives for people to develop new financial contracts or technologies.[4] Using two case studies, this paper illus-

---

[1] See Goldsmith (1987) for some descriptions of financial institutions and instruments throughout history.

[2] Schumpeter (1939) defines an innovation as a change in the shape of the production function. He goes on to categorize innovations as being either "process" or "product." Process refers to innovations that permit an existing product or service to be provided more cheaply. Product refers to innovations that introduce a product or service that was previously unavailable. The ATM, for example, is a mixture of both types. An ATM provides many routine services, such as accepting deposits and disbursing withdrawals, and usually does so more cheaply. At the same time, an ATM can provide services that were previously unavailable, such as nighttime and weekend withdrawals.

[3] Some examples are Allen and Gale (1988), Chen and Kensinger (1990), Johnston and McConnell (1989), and Silber (1981).

[4] Some examples are Goodhart (1986), Kane (1984), Podolski (1986), Rasche (1988), and Simpson (1984).

trates how financial innovations arise to meet a perceived demand for new financial services. The contrasting experience of the two cases shows how market forces can spell failure for product designs that do not attend to the principles of financial theory and success for those that do.

WHY INNOVATION OCCURS

In general, individuals do not innovate out of a spirit of magnanimity. Indeed, we shall assume, as we do with other economic behavior, that financial innovations are created in anticipation of material gain. Most theories of the incentives to innovate can be understood in terms of a cost-benefit analysis: new potential profits are the incentives to innovate. These arise when a change occurs that makes possible either a reduction in costs, an increase in revenues, or both. For simplicity, such changes are usually treated as occurring exogenously to the financial services industry, even when this depiction is not entirely accurate.

On the cost-reduction side of the cost-benefit interpretation, exogenous technological change is the force most often cited as producing the potential cost reductions that can induce innovation. As we shall see below, the transaction costs and the costs of market illiquidity are two factors that frequently affect the production and success of financial innovations. Advances in computing power, for example, have lowered the cost of such accounting-intensive products as brokered deposits and mutual funds. Other products relying on rapid calculation and decision, such as portfolio insurance and index arbitrage transactions, have similarly been made feasible by increases in computer speed. The ATM, which reduces bank operating costs by efficiently executing much of a teller's drudgery, was made possible by gains in both computing power and miniaturization.

Some of these innovations can also illustrate the other side of the incentive to innovate — the potential for increased revenues. Activities like index arbitrage would be inconceivable without some form of reliable, high-speed calculation; computers can thus be seen either as reducing

the potential cost of the activity from infinity or as increasing potential revenues above zero.

More commonly cited as forces that can generate potential sources of new revenue are government policy and inflation.[5] The tax code, in particular, offers numerous incentives. Taxpayers innovate to exploit loopholes and avoid assessments (for example, tax-free municipal bond mutual funds). If taxpayers are successful in using financial innovations to lower their tax bills, government may seek to re-legislate to close loopholes and expand the scope of taxation. Completing the cycle — in what Kane (1977) has called the "regulatory dialectic" — taxpayers find new incentives in the revised tax code and innovate again. Broad-based macroeconomic factors can also motivate innovation. For example, inflation combined with deposit interest rate ceilings in the late 1970s and early 1980s to produce new kinds of bank deposits — super-NOW accounts and money-market mutual funds — which function essentially as interest-bearing checking accounts. More recently, the Coffee, Sugar and Cocoa Exchange experimented unsuccessfully with consumer price index (CPI) futures, which were to have allowed investors to hedge against inflation.

WHY PARTICULAR INNOVATIONS SUCCEED

Creators of financial innovations are obviously interested in whether their innovations will succeed in the marketplace. Success, of course, is not automatic. For example, as Marton reports, "some exchange officials privately admit that they put out new futures products pretty much the way a cook tests his spaghetti strands — by flinging them against the wall to see if any of them stick."[6] Because innovation is not costless, however, it is important that innovators have a more systematic approach to innovation than primitive trial and error.

To examine the successfulness of specific innovations, we must first have some way of measuring success. Because we have assumed that people innovate in hopes of profit, a measure of success should either measure the inno-

---

[5]See, for example, Kane (1984) or Miller (1986).

[6]See Marton (1984), p. 239. Alchian (1977), pp. 30-32, discusses the role of trial and error in the process of innovation, and he equates success with survival. He notes (p. 31) that "the available evidence [that the necessary condi-

tions are met under which trial and error converges to a profit-maximizing equilibrium] seems overwhelmingly unfavorable." For our purposes, defining success as survival is not precise enough to be useful, since even failures will survive at least briefly before their demise.

vator's profits directly or, at least, be correlated with those profits. In many cases, a measure of popularity can proxy for success. For example, an exchange that introduces a new security might measure success by dollar trading volume in the new contract. Given a measure for success, innovators can set about designing a product to maximize that measure.

Some of the factors leading to a successful innovation are illustrated in the following example. The economist Milton Friedman, anticipating a devaluation of the British pound in the early 1970s, discovered he could not speculate on his beliefs, because no bank would allow him to sell the pound short. Hearing of his plight (and presuming that his predicament was not unique), Leo Melamed and the Chicago Mercantile Exchange (CME) launched the International Monetary Market (IMM). The IMM trades, among other things, foreign currency futures, which allow investors to speculate on devaluations (or appreciations) of the pound. This innovation has been successful because the CME found a trade that investors wanted to make, but could not, and it devised an instrument that allows them to do so cheaply and reliably (an exchange-traded foreign currency futures contract).[7]

A more general application of this recipe for success must answer several questions. For example, how is an innovator to know which as yet non-existent security investors want to trade? A full answer to this requires considerable insight into investor demand. Famous economists may offer suggestions, but this process is not always reliable. Milton Friedman, for example, also advocated the failed CPI futures contract.[8]

A secondary problem is knowing whether investors can already make a certain type of trade. While the answer might seem readily obtainable, the existence of substitutes is not always obvious. In the futures markets, for example, hedging a commodity position with a futures contract on a close substitute (cross-hedging) may be adequate or even superior to hedging directly.[9] Identifying these possibilities and how they might be used can be quite subtle. Such

subtleties can play a role in determining the success or failure of a new financial product. These issues play a prominent role in our first case study.

Nonetheless, some rules of thumb do exist for identifying likely successes in futures markets: contracts based on commodities that are easily standardized, have large price volatility, and have enough suppliers and demanders to create a liquid market.[10] These rules are far from foolproof, however. Moreover, they do not generalize automatically to other types of innovations. Indeed, codification of the elements of successful innovation as a collection of rules is almost certainly impossible; it is also beyond the scope of this paper. Instead, we illustrate with two case studies the more general ideas of the incentives to innovate and the application of financial principles to real innovations.

CASE 1, A FAILURE: CANADIAN COIN FUTURES

The first innovation we consider is a futures contract on bagged Canadian silver coins, introduced by the IMM. This innovation was a failure. After 13 months of meager trading, the IMM discontinued the unpopular and unprofitable contract. Why did this contract fail? There is an answer that is consistent with both our theoretical rationales for successful financial innovation and with the facts: A good cross-hedge existed in the much more liquid silver futures market. There, hedgers could achieve similar results at lower cost.

A Description of the Instrument

On October 1, 1973, the IMM opened trading in a new futures contract on Canadian silver coins. The purchaser of a contract promised to pay a certain future amount in U. S. dollars (USD) at a specific future maturity date; in exchange, the purchaser would receive future delivery of five bags of Canadian silver coins (dimes, quarters or half-dollars), with each bag worth 1,000 Canadian dollars (CAD) at face value. Different denominations could not be

---

[7]See Miller (1986), p. 464. Similar speculation was possible in the interbank market for forward foreign exchange. Individual investors, like Friedman, are generally excluded from this market, however.

Innovations like this, that expand the possibilities for exchange, are generally Pareto-improving. That is, they can make everyone involved better off. See, for example, Flood (1991).

[8]See Friedman (1984).

[9]A useful concept in this regard is "redundancy." If the price of a good is always a fixed multiple of the price of another good, so that the price changes for the two are always perfectly correlated, then one of the goods is said to be redundant.

[10]See Black (1986), pp. 5-12.

mixed within a bag, and the coins were to have been minted before 1967, with the exception that a bag could contain up to 2 percent 1967 or 1968 coins containing at least 50 percent silver.[11]

The contract was presumably intended to afford commercial banks holding vault inventories of silver coinage the opportunity to hedge their positions. In the mid-1960s, both the United States and Canada had phased out the use of silver in their coinage. In an application of Gresham's Law — "Bad money drives out good" — both banks and private investors hoarded silver coins for the bullion content, using instead the new, non-silver coinage as a transactions medium.[12]

The hoarded Canadian silver coins might be valuable for two reasons. First, they were Canadian currency and, therefore, could be exchanged for goods and services in Canada. Second, they contained substantial amounts of silver, a mineral with numerous industrial uses.[13] Therefore, their USD value could fluctuate with changes in the USD/CAD exchange rate or with changes in the USD price of silver. Banks wishing to protect themselves against such price fluctuations could simply sell their stores of the coins, or they could sell the new futures contracts, which would lock in a future sale price for their inventory. As we shall see, however, indirect hedges were also available.

Redundancy

We are interested in factors contributing to the failure of the IMM's futures contract on the coins. One way to explain this failure would be to find some other security that provides the same risk allocation more cheaply. In other

words, we can explain failure by establishing the existence of an efficient cross-hedge for the futures contract. It turns out that, subject to some caveats, we can demonstrate that the coinage contract was indeed redundant.

In doing this, it would be helpful to understand theoretically why the price of the cross-hedging instrument should be correlated with the price of the coinage futures contract. With this in mind, we examine the dual role (as either silver or currency) of the coins more closely.

At any time, the coins could be used for one of two purposes: as a stock of raw silver for industrial purposes, or as a medium of exchange for transaction purposes. The coins clearly could not be used for both purposes simultaneously: industrial use would require melting the coins; monetary use would require not melting them. Instead, if the coins were removed from bank vaults and put to use, we should expect them to go to the more valuable of the two uses.

These relationships are presented in figure 1. This surface is a plot of the value of coins as a function of the USD/CAD exchange rate and the USD value of silver. Given coordinates for the USD value of silver and the USD value of Canadian dollars, the height of the surface at that point is the value of the bagged Canadian coins. The graph formalizes the notion that the silver coins are worth the larger of two values: their value as silver or Canadian currency. The axes are scaled to correspond to contract specifications in the futures markets. The units on the CAD axis give the USD value of CAD 100,000, the contract size for a foreign exchange futures contract. Similarly, the silver axis gives the USD value of 5,000 troy oz. of silver, the contract size for a silver futures contract at the Chicago

---

[11]See IMM (1973a), p. 22. Because the payoffs described here are all in U. S. dollars, it may be helpful to imagine that the hedging institution is a U. S. bank. Regardless, the nationality of the parties involved should not affect in any way the pricing relationships discussed below.

[12]The new coins consisted of a copper core clad in a copper-nickel alloy; they contained no silver. The Board of Governors of the Federal Reserve (1970) ruled that vault inventories of U. S. silver coins held by commercial banks for their own account could still be counted as part of required reserves, even though they were being hoarded and therefore would not circulate. Coins to which the bank did not have "the full and unrestricted right" — for example, coins held for safekeeping in the name of a speculating depositor — would not satisfy reserve requirements.

[13]IMM (1973a), p. 20, suggested that the numismatic value of the coins might also be a factor. The possibility that bagged coins might have significant numismatic value is

remote, however. Collectors grade coins according to quality (e. g., proof, uncirculated, very fine, etc.). With rare exceptions, coins of recent minting are priced above face value, only if they are rated as proof or uncirculated. The fact that the coins subject to the futures contract were poured loose into bags implies that they could not be rated proof or uncirculated. Indeed, the "sweating" of coins, i. e., shaking them loose in a bag to rub off silver shavings, is a traditional means of debasing a currency. The ABA (1965), p. 8, for example, claimed that it was "obvious that it will be years and years before these coins have any value as collectors' items."

Figure 1
Theoretical Relationship Between Spot Prices
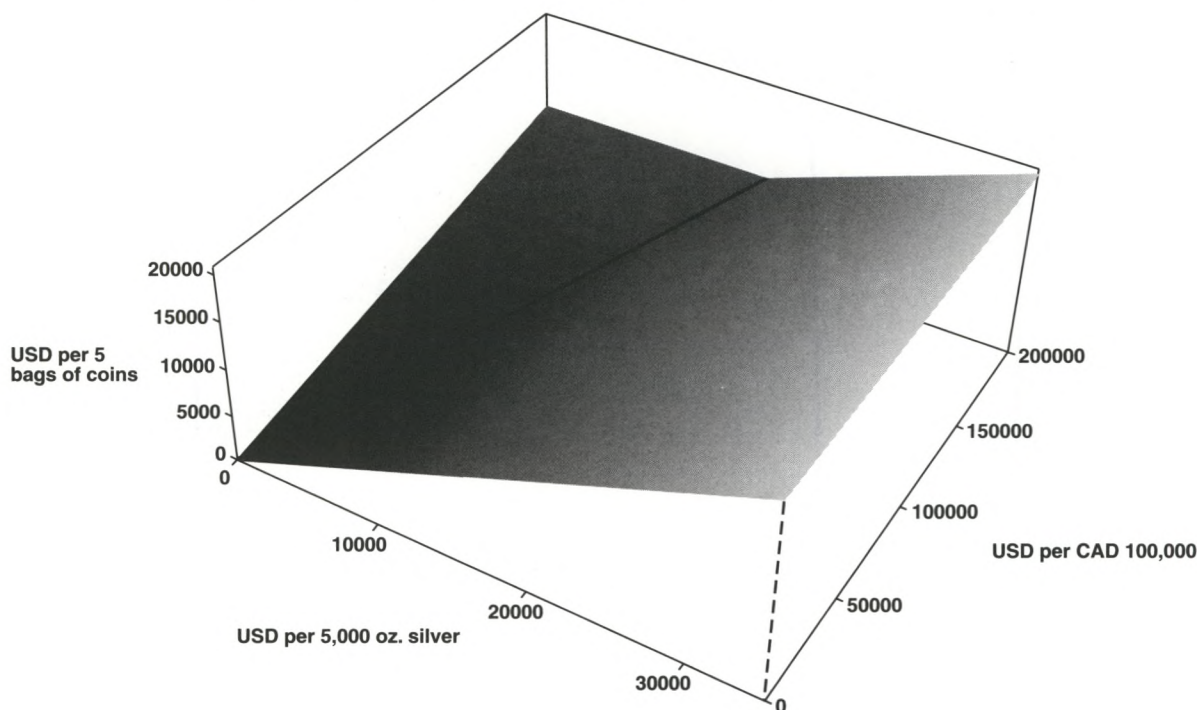
Board of Trade (CBOT). The surface is a graph of the function:

$$V_{ct} = \max \{0.05V_{dt}, 0.579V_{st}\},$$

where V_{ct} is the USD value of the five bags of coins at time t, V_{dt} is the USD value of CAD 100,000, and V_{st} is the USD value of 5,000 troy oz. of silver.[14]

While this equation describes the relationships among spot prices for the three commodities, it does not address the price of the coinage futures contract directly. Two facts are relevant to our investigation at this point: First, futures contracts for both Canadian dollars and silver were actively traded at the time; second, the price of any futures contract must converge to the corresponding commodity spot price on the maturity date of the futures contract. The latter condition holds because, at maturity, arbitrage ensures that a contract for future delivery effectively becomes a contract for spot delivery.[15]

---

[14]The coefficients, 0.05 and 0.579, are simply the proportions of CAD 100,000 and 5,000 troy oz. of silver, respectively, present in five bags of coins. The contract requires that the coins are worth CAD 5,000 at face value, or 5 percent of CAD 100,000. The contract also specifies that "The coins must bear a minting date of 1966 or earlier, **except that an individual bag may contain up to 2% 1967 or 1968 coins containing at least 50% silver** ... The gross weight of each bag, including bag, seal, and tag shall not be less than 50.75 pounds [avoirdupois] for each [CAD] $1,000 face amount;" emphasis in the original, IMM (1973b), p. 2. (Coins minted prior to 1967 were 80 percent silver by weight.) Deducting 0.75 pounds for the weight of the bag, and assuming that banks applied Gresham's law, that is, they chose their coins to minimize

the silver content of each bag within the constraints of the contract, we find that each bag (CAD 1,000 face value) contained 39.7 lbs., or 578.9 troy oz. of silver. Thus, five bags contained 57.89 percent of 5,000 troy oz. of silver.

[15]For example, if the price of a maturing futures contract significantly exceeded the spot commodity price, then an arbitrageur could make unlimited, riskless profits by selling futures, buying spot and delivering the commodity. There is some margin for price discrepancies: this arbitrage is only profitable if the price difference exceeds the costs of transaction and delivery. Still, only one arbitrageur is needed to enforce the arbitrage; the delivery and transaction costs for the least-cost arbitrageur should be the only relevant ones in this context.

Figure 2
Daily Spot Prices of Silver and Canadian Dollars

U.S. $ per 100,000 Canadian dollars   U.S. $ per 100,000 Canadian dollars

$ U.S. per 5,000 oz. Silver
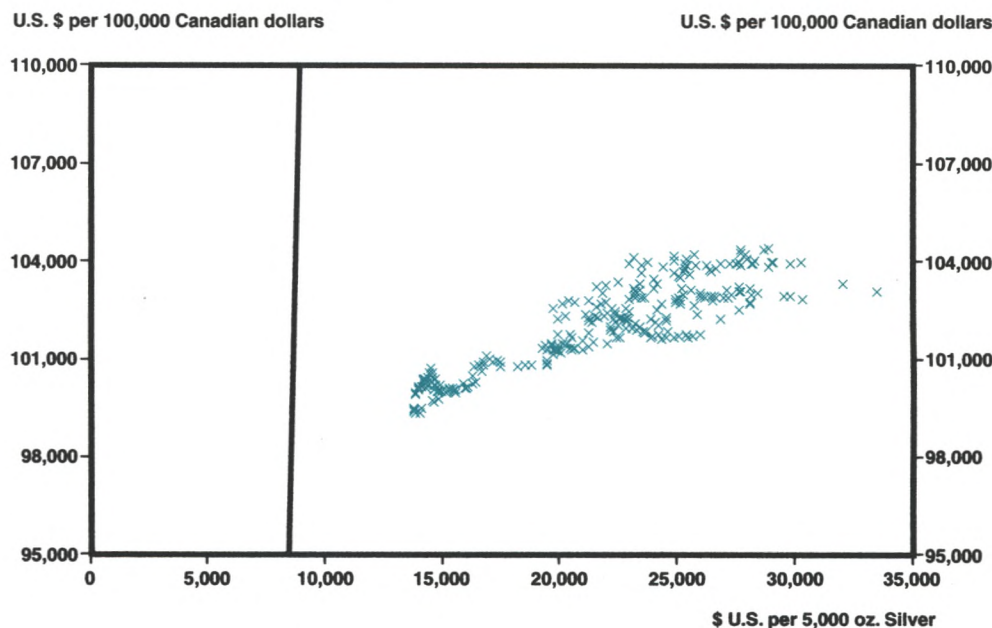
Given these relationships, we can restate our equation (approximately) in terms of the maturing futures contracts:

$$V_{cT} \approx \max \{0.05V_{dT}, \ 0.579V_{sT}\},$$

where V_{cT} is the USD value of the coinage futures at the maturity date, T, V_{dT} is the USD value of CAD futures, and V_{sT} is the USD value of the silver futures. Assuming that delivery and transaction costs are negligible makes this equation exact: $V_{cT} = \max \{0.05V_{dT}, \ 0.579V_{sT}\}$. This equation holds only at the maturity date of the futures contracts, however. At maturity, there is no uncertainty about which is larger, $0.05V_{dT}$ or $0.579V_{sT}$. Before maturity, V_{dT} and V_{sT} are uncertain.[16]

As it turns out, this latter complication is largely academic. In hindsight, over the life of the coinage futures contracts, the coins were never more valuable as currency than they were as silver. Figure 2 graphs daily observations (scaled as in figure 1) of spot silver prices and the USD/CAD spot exchange rate for October 1973 through June 1974. The line in the graph corresponds to the crease in the surface of figure 1.[17] The fact that all points fall to the right of the line means that $\max\{0.05V_{dt}, \ 0.579V_{st}\}$ was always $0.579V_{st}$ (that is, that $0.579V_{st} > 0.05V_{dt}$).

More significantly, it appears that, even before the fact, investors considered an outcome to the left of the line to be highly unlikely. If the probability of such an outcome is negligible (that is, if the probability is small that $0.05V_{dt} > 0.579V_{st}$), then it is safe to use the approximation: $\max\{0.05V_{dt}, \ 0.579V_{st}\} \approx 0.579V_{st}$. Moreover, if this approximation *always* holds in the spot markets, it should also hold for futures

---

[16]We can still establish certain relationships between the prices, however. For times t < T, the possibility that the coins might be more valuable as currency than as silver provides a price floor for the coinage futures in terms of the silver futures: $V_{ct} \geq 0.579V_{st}$. Similarly, we have a price floor for the coinage futures in terms of Canadian dollar futures: $V_{ct} \geq 0.05V_{dt}$. Finally, because holding both silver and Canadian dollars at maturity must always be prefera-

ble to holding only the more valuable of the two, we can readily establish a price ceiling: $V_{ct} \leq 0.579V_{st} + 0.05V_{dt}$.

[17]The crease in figure 1 is the set of points satisfying both $V_{ct} = 0.05V_{dt}$ and $V_{ct} = 0.579V_{st}$. Thus, it is the line defined by $0.05V_{dt} = V_{ct} = 0.579V_{st}$, or $V_{dt} = (0.579/0.05)V_{st}$. This is the line graphed in figure 2.

Figure 3
Daily Futures Prices for Silver and Canadian Coins

Price of Canadian coinage futures

Figure 3
Daily Futures Prices for Silver and Canadian Coins

◇ ◇ ◇ June contracts   O O O December contracts

Price of silver futures
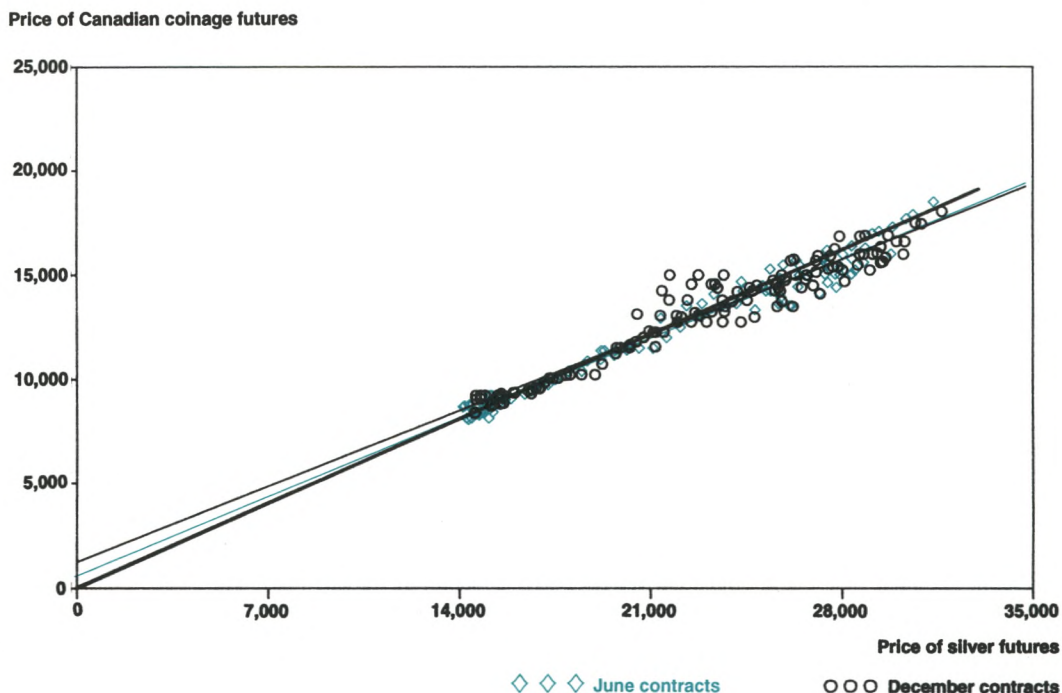
contracts; this implies we can use the approximation $V_{ct} \approx 0.579 V_{st}$.

The upshot of this is that, as long as investors could safely ignore the possibility that the coins would be more valuable as currency, a silver futures contract was a good cross-hedge for the coins. In other words, the value of the coins would be determined solely by their silver content, rather than their potential use as Canadian currency — the value of the coins should move in tandem with silver prices. Conversely, if all the points were to lie on the left side of the line in figure 2 instead of the right, the opposite condition would hold: the CAD futures contract would be the appropriate cross-hedge to

consider, and we could disregard the silver contract.

The usefulness of silver futures as a cross-hedge is confirmed in figures 3 and 4. Figure 3 plots daily price observations for silver futures and Canadian coinage futures maturing in June 1974 and December 1974. The bold line through the origin is the theoretical relationship under redundancy: $V_{ct} = 0.579 V_{st}$.

The other two lines are regression lines (lines of best fit) for the two different maturities.[18] Given our regression results, we can be more specific about the effectiveness of a cross-hedge. The coefficient of determination, or R^2 statistic, from such a regression is a standard measure of

---

[18]These regressions show that the behavior of silver futures prices and that of coinage futures prices are very similar, implying that these two potential hedging instruments are essentially indistinguishable from a pricing standpoint. An alternative would have been to demonstrate their hedging effectiveness directly, by regressing spot coin prices first on coinage futures prices and then on silver futures prices. A spot price series for Canadian silver coins was not available, however.

Figure 4
Daily Prices for December 1974 Silver and Canadian Coin Futures

Price of Canadian coinage futures　　　　　　　　　　　　　**Price of silver futures**

October 1, 1973 to October 25, 1974
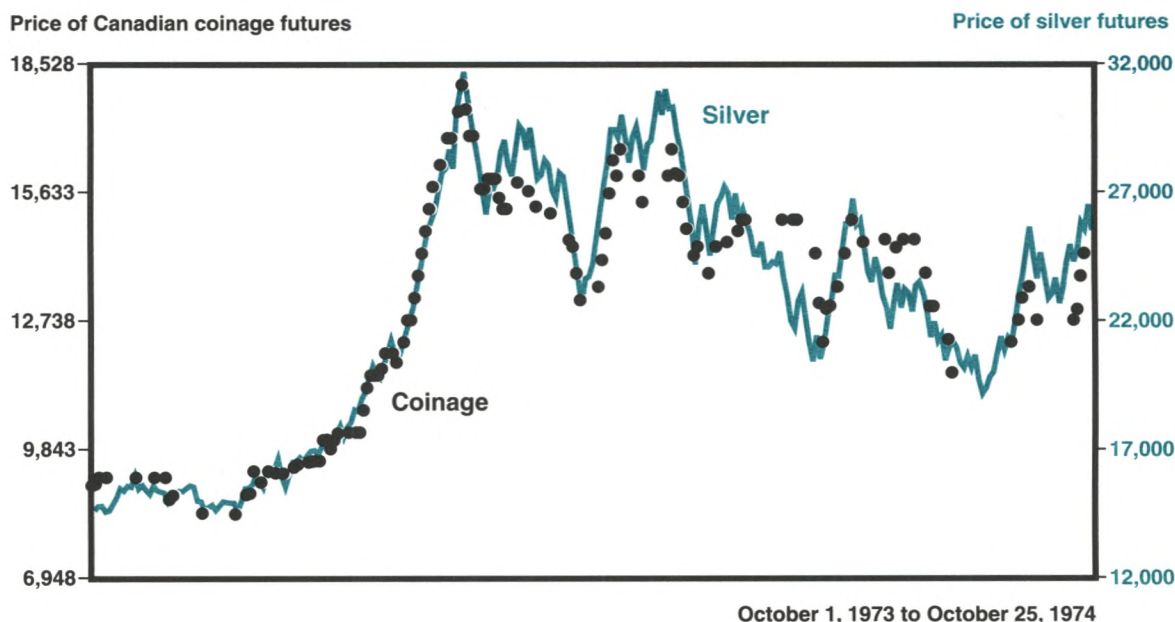
the effectiveness of a hedge. The R^2 statistic can vary from 0.00 for a useless hedge to 1.00 for a perfect hedge. For the regressions on the June and December contracts, the R^2 statistics were 0.973 and 0.933, respectively.[19] The implication is that silver futures moved closely together with coinage futures, and thus could serve as an excellent cross-hedge: Investors could have achieved almost identical results with the silver contract as with the coinage futures. This is confirmed in figure 4, which shows the same relationship plotted as time series of the two prices. Silver and coinage futures prices moved in near lockstep.[20]

Performance of the Instrument

Having established the existence of a serviceable cross-hedge for the Canadian coinage futures contract, we now examine the contract from the perspective of the innovator. From this perspective, the contract is successful to the extent that it profits the innovator. The innovator in this case was an organized futures exchange, the IMM, owned by the members of the exchange. The membership of a futures exchange consists of its traders, who benefit from an increase in trading volume through higher

[19]We can compare these numbers to the hedging effectiveness of CAD futures. As expected, a regression of the price of coinage futures on the price of CAD futures yields lower R^2 statistics, implying that silver futures were a better cross-hedge. The R^2 was 0.568 for the June contracts and 0.715 for December.

The use of the R^2 originated with Ederington (1979), pp. 163-64. The standard measure is based on regressions of returns on returns, rather than prices on prices, as are shown in the figure. Because the coinage futures had no recorded price on most days (due to no trading), however, it was not possible to construct a daily return series. An imperfect alternative is to calculate returns from one price observation to the next, producing a time series of returns with irregular holding periods. The R^2 statistics for these

regressions are 0.385 for the June contract and 0.419 for the December contract.

In comparison, Johnston and McConnell (1989) report on Government National Mortgage Association (GNMA) Collateralized Depository Receipt (CDR) futures contracts. During the most successful years of the contract, 1980-82, the hedging effectiveness (R^2) of the futures contract for the underlying GNMA securities over five-day holding periods ranged from .85 to .94. During the contract's leanest years, 1983-85, the R^2 ranged from .54 to .62 percent, making it a less effective hedge than a Treasury bond futures contract.

[20]Because of low volume in the coinage futures market, there were many days with no bids or offers submitted and, therefore, no posted settlement prices.

turnover or through greater liquidity.[21] For our purposes, we can proxy for profits by measuring trading volume in the contract.

Illiquidity can be costly to traders in two ways. First, traders may be forced to search actively for counterparties with whom to trade; searching consumes time and resources. Second, traders who are delayed by the search process face price risk: the price of their commodity can change before they locate a counterparty. In general, a trader would require some remuneration before bearing such a risk willingly. One of the primary economic functions of an organized exchange is to bring traders together in the same place, to obviate such search costs. There is a recursive catch in the economic logic here, however: traders may not be attracted to a market if it is not liquid, but a market may not be liquid unless traders are attracted to it.

Trading volumes for the June and December coinage contracts, however, were never large. On most days, fewer than five contracts changed hands. There were many days with no trades. Of 183 trading days for the June contract, there were 66 days when only a bid quote or only an ask was available; neither was available on 78 days. For the December contract, 92 of 310 trading days had only one side of the market present, and on 181 days neither side was available.[22] Trading volume peaked at 45 contracts on October 31, 1973; in comparison, for the CBOT's silver futures, several thousand contracts changed hands on a good day in 1973.[23] By October 1974, trading in the Canadian coinage futures had dwindled to almost nothing. The last recorded trade came on October 25, 1974, when a single December 1974 contract changed hands.

In hindsight, we have a theoretical explanation for the contract's failure. Futures contracts for Canadian silver coins should have been at-

tractive to owners of silver coins, who wanted to hedge the value of their coins against price fluctuations, and speculators, who were willing to bet they could predict those price changes more accurately than the rest of the market. A comparison of prices reveals that existing silver futures were a close substitute for the coinage futures as a hedging/speculating tool. Moreover, the long-established silver contract traded in a much more liquid market. Thus, hedgers and speculators had in the silver contract most of the benefits of the coinage contract as a hedging instrument, without most of the drawbacks associated with illiquidity. In this context, then, the silver futures should be seen as uniformly preferable to the coinage futures.

CASE 2, A SUCCESS: MARKET INDEX MUTUAL FUNDS

We turn now to a successful innovation: market index mutual funds. The size of the particular fund chosen as an example has grown steadily since its introduction in 1976. Although an index fund is little more than a repackaging of other securities, with little decision-making discretion left to the fund's management, such funds can succeed by reducing the costs of transacting for individual investors.

A Description of the Security

On August 31, 1976, the Vanguard Group introduced the "500 Portfolio," a stock market mutual fund whose specific objective is to track the Standard and Poor's 500 stock market index (S&P 500). In the language of the fund's prospectus, "the 500 Portfolio seeks to replicate the aggregate price and yield performance of the Standard & Poor's 500 Composite Stock Price Index. ... The 500 Portfolio invests in all 500 stocks in the S&P 500 Index in approxi-

---

[21]See Black (1986), pp. 19-21. There is no universally accepted definition for the term "liquidity." It is usually associated with an absence of search costs. See Demsetz (1968), Tinic (1972) and Logue (1975) for further discussion.

[22]Figure 4 includes prices for the coinage contract for many days on which trading volume was zero. The exchange posted settlement prices on many days when either bids or offers appeared but no trading occurred.

[23]A related measure of the coinage contract's lack of popularity is open interest (the sum of all traders' net long positions in the contract, also called committed contracts). The same picture of a thin and illiquid market emerges. Open interest in the June 1974 contract peaked in May 1974 at 56 committed contracts. For the December 1974 contract, the peak of 37 committed contracts was reached

in October 1973. Again, in comparison, open interest for the CBOT's silver contract reached tens of thousands of committed contracts during this time.

mately the same proportions as they are represented in the Index."[24]

There are good reasons why we might expect a market index mutual fund to be unattractive to investors. First, it is inflexible. For example, an investor in an index fund who wanted to divest any stake in, say, petroleum stocks would have to sell her entire stake in the fund. In contrast, an investor holding the same stocks directly could sell petroleum stocks without affecting her other positions. Second, a mutual fund introduces a middle man: The fund must be compensated for its investment management services, which, for an index fund, are essentially paper-shuffling. The allocation of assets in an index is set by a predetermined rule, which an investor could follow on his own.

On the other hand, there are also reasons why investors might find an index fund attractive. First, despite the inflexibility of rule-based indexes, some investors might be attracted to a specific index, so that the inflexibility of that index would not be a binding constraint for those investors. The standard capital asset pricing model, for example, concludes that all investors should hold a value-weighted portfolio of all available assets to achieve the best risk-return trade-off.[25] Although the S&P 500 does not contain *all* assets, it is well diversified, it is value-weighted, and it is widely known. Perhaps for this reason, the S&P 500 has become an industry benchmark.[26]

Assuming there is a special interest in the S&P 500 as an investment portfolio, we still must explain the popularity of a mutual fund as a preferred means of holding that portfolio. Gorton and Pennacchi (1991) construct an elaborate rationale for security baskets as a way for uninformed traders to exploit the inflexibility of

a market index to diversify themselves against losses to insider trading. A plausible, although more prosaic, reason is that mutual funds might be a way to spread the fixed costs of stock market trades over many investors, thus lowering the average cost faced by each investor. These costs can be significant. Since stock prices generally change many times in one day, the number of transactions required to maintain a theoretically exact index portfolio of 500 stocks is potentially enormous. As long as there is a non-zero fixed cost per trade, the sum of these fixed costs of maintaining the index will be proportionately large.

Redundancy

We saw in the first case that the prices of Canadian coinage futures contracts were tracked almost exactly by the prices of silver futures; the coinage futures were redundant. We now perform a similar analysis on the index fund and demonstrate that it too is redundant: the value of the fund is closely tracked by the value of the index. In this case, however, the managers of the fund have actively pursued exactly this correlation as their stated purpose. Redundancy here helps explain the innovation's success.

Figures 5 and 6 are analogous to figures 3 and 4, respectively. Figure 5 plots daily price observations for the Vanguard 500 Portfolio and the S&P 500. We see that the fund price and the index value move tightly together. Upon closer examination, the prices seem to be confined to a collection of line segments radiating from the origin and rotating downward as one moves further out. The line segments are periodically bumped downward, as maintenance fees, operating charges and dividend and capital gains distributions tend to be concentrated in

---

[24]Vanguard Index Trust (1992), p. 9. For those unfamiliar with index investments, it is important to note that a buy-and-hold investment strategy is incompatible with indexing. The composition of an index portfolio (i. e., the number of shares of each stock) depends on the prices of all the stocks in the index. Therefore, in contrast to a buy-and-hold investment, the composition of a theoretically exact index portfolio changes every time the price of any one of its component securities changes. A dynamic investment strategy is required for an index portfolio.

Vanguard is only one of many companies to offer an S&P 500 index mutual fund for investors. Its use as an example should not be interpreted as a recommendation for or against this or any other mutual fund. There are also other ways of investing in the S&P 500 index, including index futures and index options.

[25]See, for example, Fama (1976), chapters 7 and 8.

[26]There are other commonly used benchmark indexes, for example, the S&P 100, the Wilshire 5000 Index, or the Dow Jones Industrial Average.

Figure 5
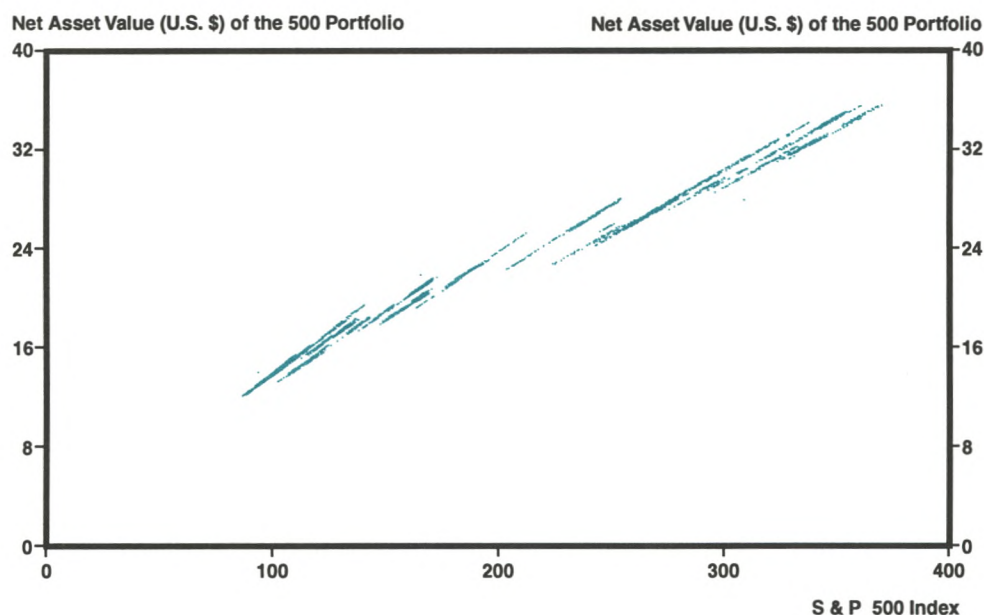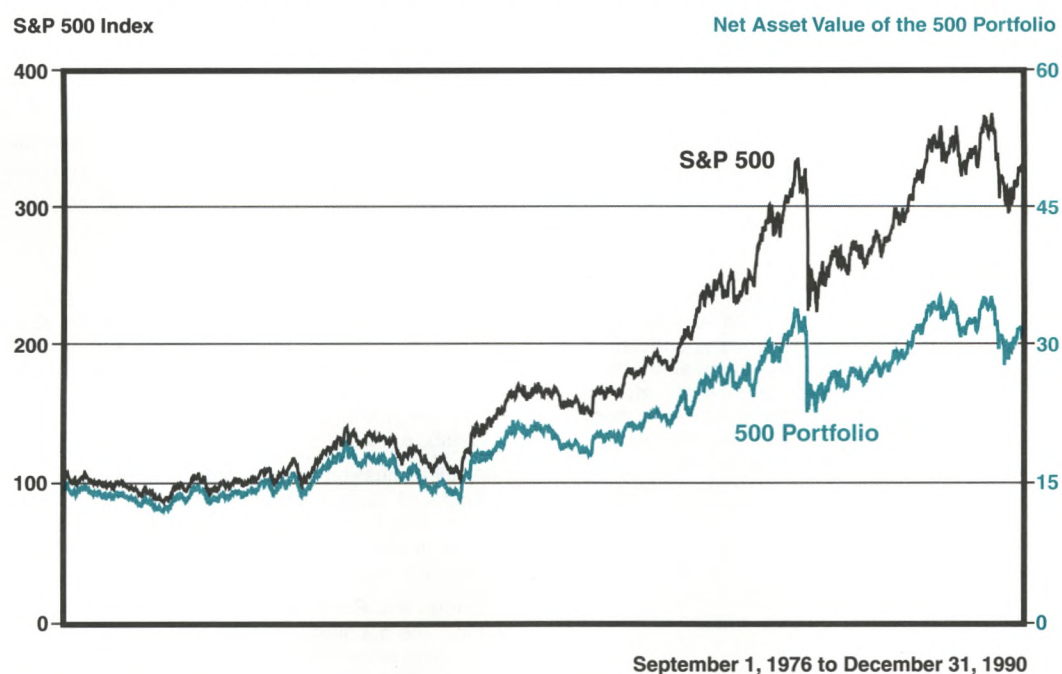## Daily Values for the S&P 500 Index and the 500 Portfolio

Net Asset Value (U.S. $) of the 500 Portfolio

Net Asset Value (U.S. $) of the 500 Portfolio

S & P 500 Index

Figure 6
Daily Values for the S&P 500 Index and the 500 Portfolio

S&P 500 Index

Net Asset Value of the 500 Portfolio

S&P 500

500 Portfolio

September 1, 1976 to December 31, 1990

cumulative year-end adjustments.[27] The line segments also tend to move successively outward over time, because of the general upward trend in stock prices over time. Even with these discontinuities, however, a regression of five-day-holding-period mutual fund returns on S&P 500 returns yields an R^2 statistic of 0.978; according to the standard measure, the mutual fund is a nearly perfect hedge for the S&P 500.[28]

Figure 6 shows the same relationship, but with the 500 Portfolio values and index values plotted as time series. Over relatively long periods of time, the value of the index grows at a slightly higher rate than the value of the mutual fund. This long-run discrepancy reflects the fact that the index is a theoretical portfolio — one that assumes transaction costs are zero.[29] The mutual fund, on the other hand, charges each shareholder a fixed quarterly account maintenance fee, plus operating expense charges equal to .2 percent of the value of the fund over the course of each year. Over time, these fees accumulate and compound to produce a discrepancy.

Performance of the Security

We turn now to the question of the innovation's success. For this, we need a measure. The innovator in this case was the fund's manager, the Vanguard Group. We therefore define a success as an innovation that profits them. For an index fund, unlike many other forms of intermediation, the manager profits only through its fees. The fund receives periodic fees from its shareholders, while incurring the costs of fund management: accounting, buying and selling stocks, receiving and disbursing payments, etc. Data on management costs were unavailable,

however, so that profits could not be measured directly.

Instead, we assume that the fund's profits are increasing with revenues over the entire range considered here. Under this assumption, profits increase with the size of the fund. We therefore use the size of the fund as our measure of success.[30] Figure 7 shows the size of the fund both nominally and in constant (CPI-adjusted) 1983 dollars. Almost since its inception in 1976, the value of the fund has grown exponentially. In nominal terms, the fund currently contains $4.346 billion, or $3.176 billion in constant 1983 dollars.

One plausible explanation of the success of an index mutual fund is transaction costs. Like many expenses, the cost to an individual investor of a stock purchase can be divided into a variable and a fixed portion: price per share times number of shares plus a brokerage commission. If large numbers of investors wish to hold the index, and if the size of a brokerage commission is fixed, or at least insensitive to the quantity transacted, then the investors can pool their transactions to reduce the total amount of commissions paid. A mutual fund is one way to achieve this pooling of investments.

Although the mutual fund falls short of the index in the long run, the relevant comparison is not with a theoretical entity, but with those alternatives that are available on a practical basis. While everyone acknowledges the presence of transaction costs, some might argue that these costs are too small in modern financial markets to make a difference. Gorton and Pennacchi, for example, suggest that "investors can costlessly replicate [these composite securities]."[31]

---

[27]Because of differences in the way the index and the mutual fund account for dividends and capital gains, tracking the net asset value of the mutual fund tends to understate its performance relative to the index. In particular, the index generally assumes that dividends, stock splits, etc., are reinvested, adjusting the index accordingly, while the mutual fund gives investors the option of collecting their dividends and capital gains. Unfortunately, sufficient information was not available to make accurate compensating adjustments to the net asset values of the fund.

[28]Performing the same analysis with daily returns, the R^2 statistic is 0.859. The drop in significance is attributable to the fact that returns for one-day holding periods are smaller than for five-day holding periods, relative to noise factors such as rounding errors.
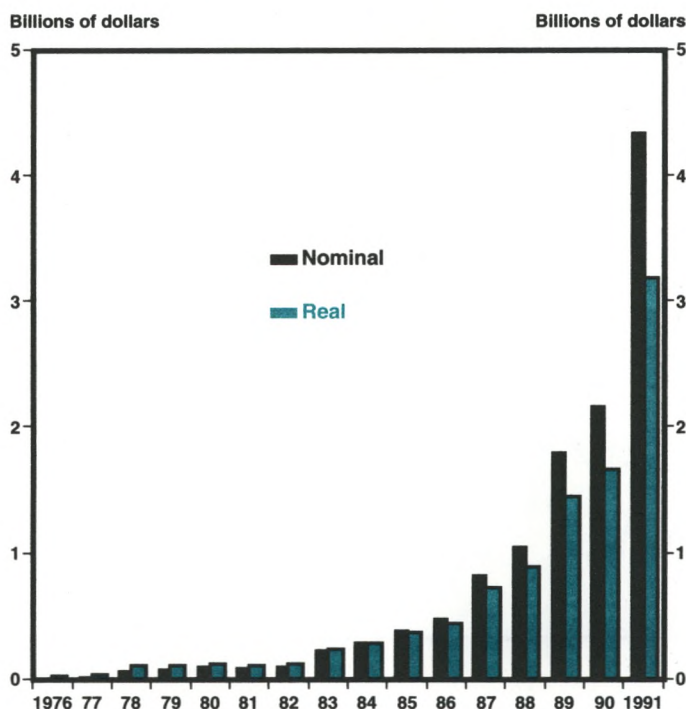
[29]As noted in footnote 27, the mutual fund time series presented here tends to understate its performance relative to the index. If we were able to compensate for disburse-

ments, the compensated mutual fund price path would lie above the net asset values shown here, but still below the curve for the S&P 500. The average annual return calculated for the mutual fund on a compensated basis was 13.9 percent over the period 8/31/76 to 12/31/91, compared with 14.4 percent for the index. See Vanguard Index Trust (1992), p. 16.

[30]A full discussion of the relationship between the size of the fund and Vanguard's profits, and therefore of Vanguard's incentives to innovate, would require consideration of issues of returns to scale and market structure that are beyond the scope of this paper.

[31]Gorton and Pennacchi (1991), p. i. They go on to state (p. 2) that "the popularity of such composite securities seems puzzling since consumers, on their own, can apparently accomplish the same resulting cash flow by holding a diversified portfolio of the same securities in the same proportions."

Figure 7
Size of the 500 Portfolio

Billions of dollars   **Billions of dollars**

■ Nominal

■ Real

1976 77 78 79 80 81 82 83 84 85 86 87 88 89 90 1991

To demonstrate the potential role of lower transaction costs in the success of the mutual fund, consider the performance of an index portfolio managed directly by an individual investor trading through a brokerage house and facing realistic brokerage commissions. To make this example plausible, we consider pre-tax returns on an S&P 500 portfolio for an investor making a modest number of trades each day and able to transact at standard brokerage commission rates. In particular, we consider an initial net investment of $1 million in the S&P 500 index made on September 1, 1976. Assume that our investor is able to track the index (approximately) by making 10 trades per day at an average cost of $30 per trade.[32] Brokerage commissions on this portfolio thus absorb $300 per day.

Figure 8 compares the results for this portfolio with those for the mutual fund over the life of the fund. The results are straightforward: over the course of almost 15 years, the value of the index held directly has fallen to about $860,000 in nominal terms, while the mutual fund has posted a 119.2 percent gain. If we were to adjust for inflation and dividends, the poor performance of the direct portfolio would be even more striking. The result is that investors who would otherwise have to trade at standard brokerage rates can hold the index more cheaply as a mutual fund. Indeed, given such transaction costs, it appears likely that an index portfolio managed by an individual investor would have negative *ex ante* returns. If this were the cheapest means of doing so, no one would want to hold an indexed portfolio.
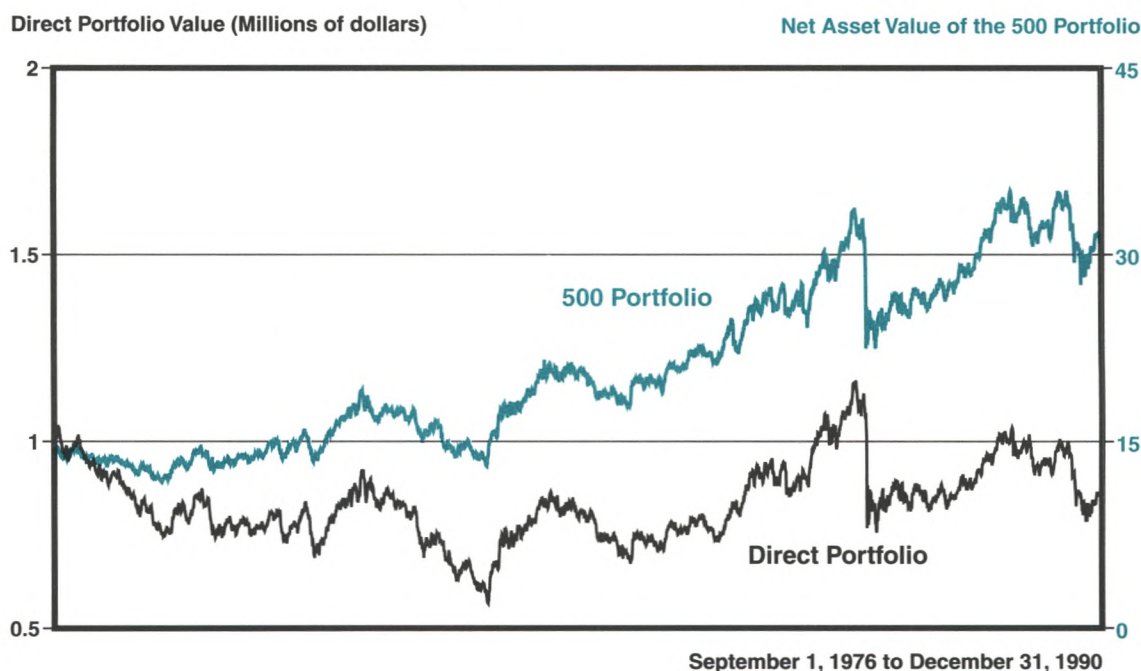
CONCLUSIONS

The foregoing has presented two case studies of financial innovations. One innovation, Canadi-

[32]Brokerage commission schedules are complex and vary considerably from one broker to another. Some of the factors that can affect the commission on a trade are volume discounts, odd-lot trading premia, and negotiated commissions for very active customers. The average $30 fee used here is based on a commission rate of 1 percent of the total price of each transaction, an average share price of $30 per share, and trading in single round lots of 100 shares.

Figure 8
An Index Portfolio Subject to Transaction Costs

Direct Portfolio Value (Millions of dollars)   Net Asset Value of the 500 Portfolio

September 1, 1976 to December 31, 1990

an silver coin futures, failed. The other, a market index mutual fund, has succeeded. Together, the two case studies represent an experiment of sorts. We have examined two innovations with a common feature: Both were redundant in the technical sense that their price movements were closely tracked by the price movements of other securities. One might say that the first failed because it was redundant, while the latter has succeeded for the same reason.

This conclusion is more compelling if we state it as the following more general proposition: Given two securities with redundant prices (that is, two that are perfect hedges for one another), investors will be drawn to the one with the lower transaction and liquidity costs. If there is no investor clientele for which a redundant security is the cheaper to employ, then that security will fail.

When stated in these terms, the conclusion seems obvious. Nonetheless, this proposition is not universally observed. It requires an explicit acknowledgement of the fact that transaction and liquidity costs can be significant factors in the financial marketplace. This runs contrary to the common assumption in financial economics that capital markets are "perfect," which implies, among other things, that transaction costs are zero. While such an assumption may be appropriate in certain applications, a full understanding of the behavior of financial markets and innovations requires an appreciation of these various hindrances to exchange. Thus, to the extent that financial frictions such as transaction and liquidity costs represent real resource drains on an economy, successful financial innovations should normally be regarded as welfare-enhancing. By replacing cumbersome or inefficient modes of exchange, successful innovations can make everyone involved better off.

REFERENCES

Alchian, Armen A. *Economic Forces at Work* (Liberty Press, 1977).

Allen, Franklin, and Douglas Gale. "Optimal Security Design," *Review of Financial Studies 1* (Fall 1989), pp. 229-63.

American Bankers Association. "Facts About the New U. S. Coins," (ABA, 1965).

Black, Deborah G. "Success and Failure of Futures Contracts: Theory and Empirical Evidence," *Monograph Series in Finance and Economics,* Salomon Brothers Center for the Study of Financial Institutions (New York, 1986).

Board of Governors of the Federal Reserve. "Interpretation of Regulation D: Currency or Coin Held Principally for its Numismatic or Bullion Value," *Federal Reserve Bulletin* (December 1970), p. 942.

Chen, Andrew H., and John W. Kensinger. "Creating Contingent Liabilities: Master Craftsmanship in Financial Engineering," in *Game Plans for the '90s: Proceedings of the 26th Annual Conference on Bank Structure and Competition, May 9-11, 1990* (Federal Reserve Bank of Chicago, 1990), pp. 199-223.

Demsetz, Harold. "The Cost of Transacting," *Quarterly Journal of Economics* (February 1968), pp. 33-53.

Ederington, Louis H. "The Hedging Performance of the New Futures Markets," *Journal of Finance* (March 1979), pp. 157-70.

Fama, Eugene F. *Foundations of Finance* (Basic Books, 1976).

Flood, Mark D. "An Introduction to Complete Markets," this *Review* (March/April 1991), pp. 32-57.

Friedman, Milton. "Financial Futures Markets and Tabular Standards," *Journal of Political Economy* (February 1984), pp. 165-67.

Goldsmith, Raymond W. *Premodern Financial Systems, A Historical Comparative Study* (Cambridge University Press, 1987).

Goodhart, Charles. "Financial Innovation and Monetary Control," *Oxford Review of Economic Policy* (Winter 1986), pp. 79-102.

Gorton, Gary, and George Pennacchi. "Security Baskets and Index-Linked Securities," NBER working paper No. 3711 (May 1991).

International Monetary Market of the Chicago Mercantile Exchange, Inc. *A Guide to Silver Coin Futures Trading* (IMM, 1973a).

_____. "Chapter 15: Canadian Silver Coins," *Rules of the International Monetary Market of the Chicago Mercantile Exchange* (CME, 1973b).

_____. *International Monetary Market Year Book 1973-1974* (IMM, 1974).

Johnston, Elizabeth Tashjian, and John J. McConnell "Requiem for a Market: An Analysis of the Rise and Fall of a Financial Futures Contract," *Review of Financial Studies 2* (1989), pp. 1-23.

Kane, Edward J. "Good Intentions and Unintended Evil: The Case against Selective Credit Allocation," *Journal of Money, Credit and Banking* (February 1977), pp. 55-69.

_____. "Microeconomic and Macroeconomic Origins of Financial Innovation," in *Financial Innovations: Their Impact on Monetary Policy and Financial Markets* (Kluwer-Nijhoff, 1984), pp. 3-20.

Logue, Dennis E. "Market-Making and the Assessment of Market Efficiency," *Journal of Finance* (March 1975), pp. 115-23.

Marton, Andrew. "How much is too much?" *Institutional Investor* (August 1984), pp. 238-50.

Miller, Merton H. "Financial Innovation: The Last Twenty Years and the Next," *Journal of Financial and Quantitative Analysis* (December 1986), pp. 459-71.

_____. *Financial Innovations and Market Volatility* (Basil Blackwell, 1991).

Podolski, T. M. *Financial Innovation and the Money Supply* (Oxford: Basil Blackwell, 1986).

Rasche, Robert H. "Monetary Policy and Financial Deregulation in the United States," *Kredit und Kapital* (Heft 3, 1988), pp. 451-68.

Schumpeter, Joseph A. *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process, Volume I* (McGraw-Hill, 1939).

Silber, William L. "Innovation, Competition, and New Contract Design in Futures Markets," *Journal of Futures Markets* (Summer 1981), pp. 123-55.

_____. "The Process of Financial Innovation," *American Economic Review* (May 1983), pp. 89-95.

_____. "Towards a Theory of Financial Innovation," in William L. Silber, ed., *Financial Innovation* (Lexington Books, 1975), pp. 53-85.

Simpson, Thomas D. "Changes in the Financial System: Implications for Monetary Policy," *Brookings Papers on Economic Activity* (1:1984), pp. 249-65.

Tinic, Seha M. "The Economics of Liquidity Services," *Quarterly Journal of Economics* (February 1972), pp. 79-93.

Vanguard Index Trust. *Prospectus: March 15, 1992; Revised April 10, 1992* (Valley Forge: The Vanguard Group, 1992).

Kevin L. Kliesen and John A. Tatom

*Kevin L. Kliesen is an economist and John A. Tatom is an
assistant vice president at the Federal Reserve Bank of
St. Louis. James P. Kelley provided research assistance. This
article was written while Tatom was a visiting economist at
the Austrian National Bank. Tatom received useful comments
in seminars on this paper at the Bank of the Netherlands,
the Swiss National Bank and the Institute of Advanced
Studies in Vienna.*

The Recent Credit Crunch: The Neglected Dimensions

CONVENTIONAL WISDOM HAS IT THAT a credit crunch occurred in the U.S. economy in 1990-92, causing, or at least contributing to, the latest recession and jeopardizing the strength of the recovery. Much has been written about the causes and consequences of the credit crunch and about remedies for it.[1]

While the term is widely used, the precise definition of a credit crunch is not widely agreed upon. Credit crunches have in common, however, a slowing in the growth of—or an outright decline in—the quantity of credit outstanding, especially business loans at commercial banks. Analysts who espouse the credit crunch theory typically have a more specific definition in mind. In their view, a credit crunch arises from a reduction in the supply of credit. Accordingly, this article uses the term "credit crunch" to refer only to a reduction in the supply of credit. It addresses the credit crunch hypothesis by examining the existence and implications of competing potential sources of a decline in credit, including the recent

behavior of interest rates, interest rate spreads and commercial bank business loans. This paper suggests that recent movements in short-term interest rates and changes in relevant interest rate spreads cast doubt upon the conventional credit crunch view.

WHAT IS A CREDIT CRUNCH?

The traditional notion of a credit crunch originally involved the process known as disintermediation— a decline in savings-type deposits at banks and savings and loans that result in a decline in bank lending.[2] Episodes of disintermediation occurred when market interest rates, especially rates on Treasury bills and commercial paper, rose above Regulation Q interest rate ceilings at these financial institutions. As this occurred, depositors withdrew their funds from banks and savings and loans to invest at higher open market rates, and bank credit, especially for business loans, fell.

---

[1]The Chairman of the Federal Reserve System, Alan Greenspan, has expressed concern over the slowing in credit growth and the extent to which it was induced by bank regulators' attempts to raise bank capital. See Greenspan (1991). Other Federal Reserve officials who have expressed concern over the credit crunch during this period include LaWare (1991), Forrestal (1991) and Syron (1991). Concern over a potential global credit crunch has been raised by the Bank for International Settlements (1991) and Japan's Economic Planning Agency [see Reuters (1991a)]. Concern for a national crunch has also surfaced in France [see Reuters (1991b)]. Also see O'Brien and Browne (1992).

[2]See Kaufman (1991) for a discussion of the origin of this term in his work with Sidney Homer and for a brief sketch of the history of credit crunches. For detailed analysis of previous credit crunches, see Wojnilower (1980).

The phrase "credit crunch" was coined in mid-1966 when the Federal Reserve's monetary policy became more restrictive; the Fed wanted to slow the growth of demand for goods and services in order to fight inflation.[3] In 1966, the Fed's actions to slow the growth of money and credit were reinforced by allowing short-term interest rates to rise above the Regulation Q ceiling rate on bank deposits. As a result, depositors withdrew their funds from regulated deposits to seek higher market rates. The reduction in bank deposits, in turn, limited banks' ability to lend and their supply of credit. What made this event significant was the Fed's refusal to accommodate the rise in short-term market interest rates by raising the Regulation Q ceiling rates at banks and savings and loans, as it had done in the past.[4] Since financial deregulation in the early 1980s ended interest rate ceilings, such regulatory-induced disintermediation can no longer occur.

A more encompassing view of a credit crunch is based on any non-price constraint on bank lending, not simply on disintermediation. In its recent application, the source of this constraint has presumably been the response by bankers to increased regulatory oversight and their own reaction to recent deterioration of bank asset values and profitability. Increased savings and loan failures, as well as increased capital requirements, may also have played a part.[5]

This broader definition of a credit crunch has been summarized by the Council of Economic Advisers (1992):

A credit crunch occurs when the supply of credit is restricted below the range usually identified with prevailing market interest rates and the profitability of investment projects. (p. 46)

A credit crunch, either through disintermediation, overzealous regulators or banks' unwillingness to lend for some other reason, is therefore usually thought of as a supply phenomenon. The flow of credit, however, results from the interaction of both credit supply and credit demand. In other words, while supply considerations could certainly result in a decline in credit flows, a reduction in the demand for credit could produce the same result. Fortunately, economic theory indicates a criterion for assessing which of these is the dominant source of a change in the quantity of credit.

CREDIT CRUNCHES: THEORY AND EVIDENCE

Figures 1a and 1b illustrate the typical analysis of the credit market aspects of a credit crunch using the supply and demand framework for credit flows. In this framework, the quantity of credit demanded varies inversely with the cost of credit—that is, the interest rate—given other factors that influence overall demand for credit. Conversely, the quantity supplied of credit increases with the interest rate, given the other factors that influence credit supply decisions.[6] We will examine two scenarios. The first illustrates the credit crunch hypothesis, namely, a reduction in the quantity of credit supplied resulting from reduced bank willingness to lend. The second scenario offers an alternative. In this case, a reduction in the demand for bank credit occurs. As will be explained below, this could be the result of a decline in business demand for credit associated with a reduction in inventory investment.[7]

Case I: Supply-induced decline in credit. In figure 1a, an equilibrium in the credit market

[3]As a result of this policy action, interest rates rose and credit became more scarce. For example, in the third quarter of 1966, the interest rate on three-month Treasury bills rose above 5 percent for the first time in more than 30 years; the 5.04 percent average rate during the quarter was up sharply from the 4.59 percent rate in the previous quarter. The growth of the money stock (M1) had already slowed from a 7.3 percent annual rate in the two quarters ending in the first quarter of 1966 to a 4.3 percent rate in the second quarter of 1966. This was followed by a decline at a 1.2 percent rate in the third quarter and a 1.2 percent rate of increase in the last quarter of 1966. See Burger (1969) and Gilbert (1986) for discussions of this episode.

[4]According to Burger (1969), p. 24, the Fed previously accommodated the rise in rates in July 1963, November 1964 and December 1965 by raising the Regulation Q ceiling. This behavior is also discussed by Wojnilower (1980). One flaw with the disintermediation view is that a decline in intermediation through banks does not reduce the *total* supply of

credit unless some of the funds removed from banks are not funneled into other credit instruments like Treasury bills or commercial paper.

[5]The latter idea is closely associated with Syron (1991), who has termed the reduction in lending as a "capital crunch." Greenspan (1991) has also supported elements of this argument, as has the Council of Economic Advisers (1992). Also see Bernanke and Lown (1991) for a different perspective on this hypothesis.

[6]Figures 1a and 1b are drawn conditionally on an expected rate of inflation. That is, the standard assumption that expected inflation is not a cyclical phenomenon during business recessions is employed. Therefore, in this framework, a change in a nominal variable is also a change in a real variable.

[7]Figures 1a and 1b are not meant to imply that bank credit determines the level of economic activity. For alternative explanations about the linkage between bank credit and economic activity, see the shaded insert on p. 24.

Figure 1a
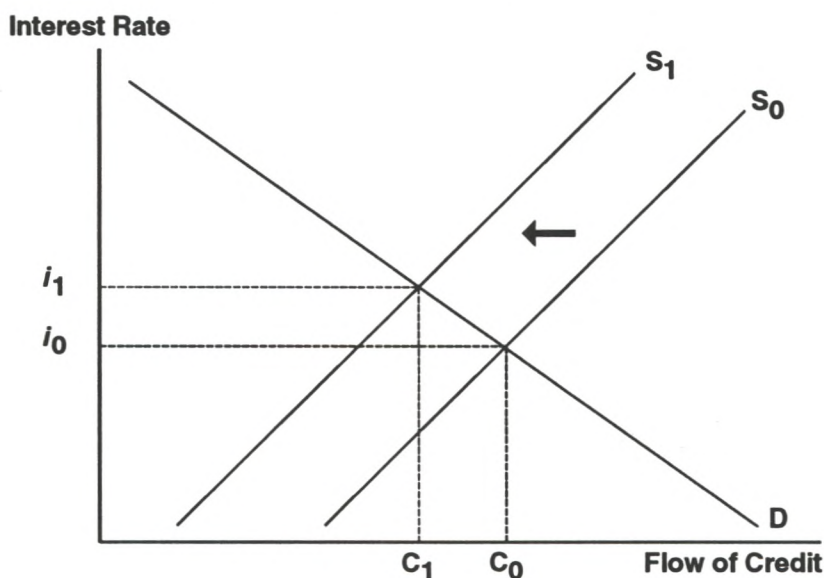## Decline in the Supply of Credit
Case I: Reduced willingness to lend

Figure 1b
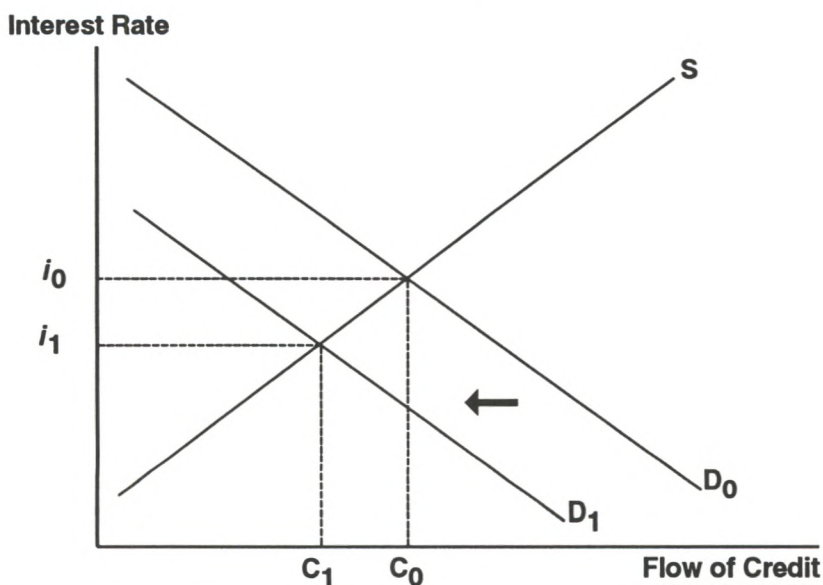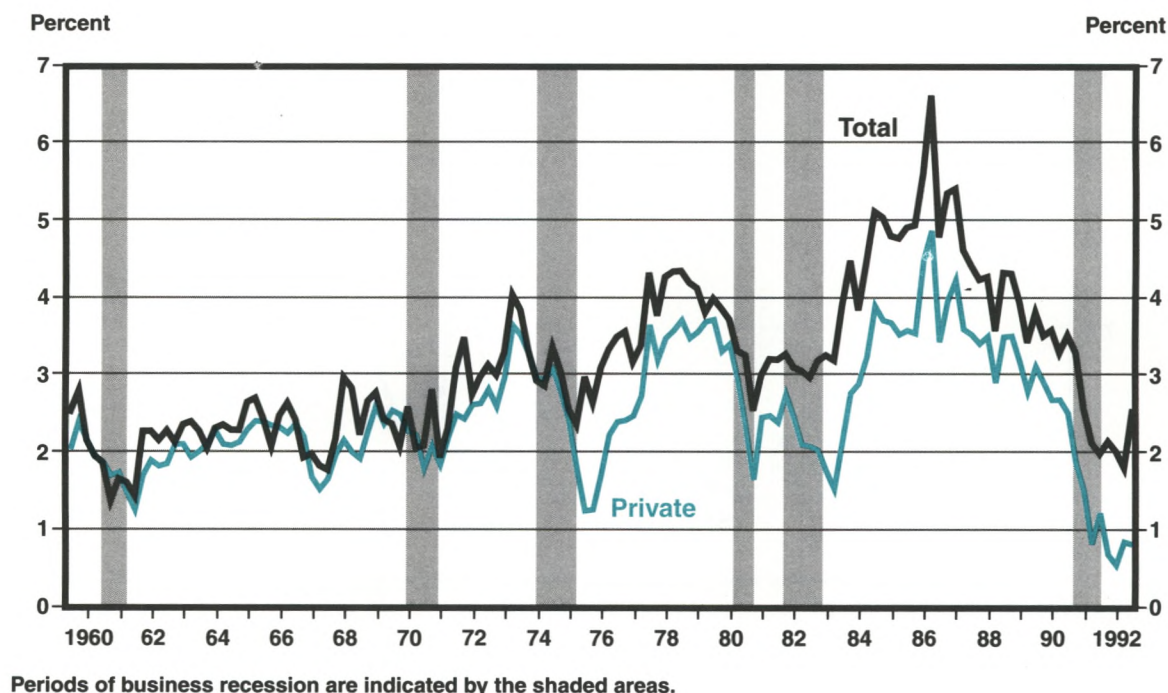## Decline in the Demand for Credit
Case II: Reduced loan demand by businesses

Figure 3
Short- and Long-Term Interest Rates

Percent   Quarterly Data   Percent
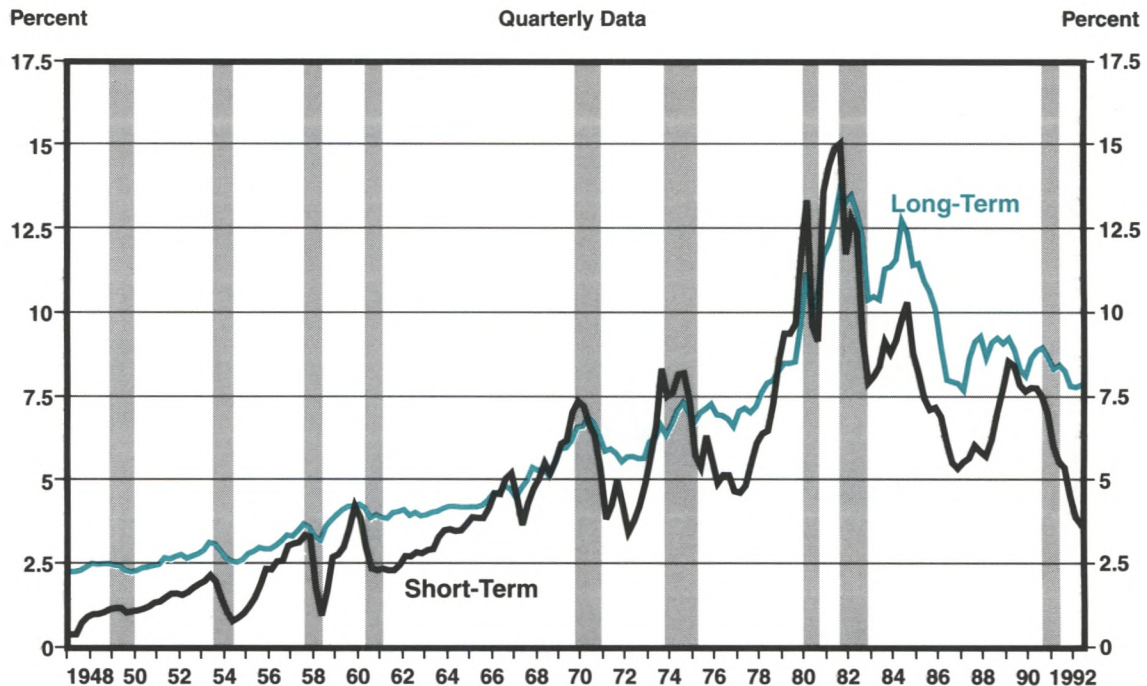
Periods of business recession are indicated by the shaded areas.

private domestic nonfinancial debt (excluding federal debt) and the change in total domestic nonfinancial debt—both expressed as a percent of gross domestic product (GDP). Typically, in recessions, the rate of debt accumulation declines relative to GDP.[9] This reflects the fact that credit growth tends to be cyclical, especially the growth of private credit; typically, credit growth slows relative to GDP during recessions and rises during expansions.[10] Thus, it is not possible to ascertain from figure 2 whether the decline in credit in each instance caused the recession or

was merely a reflection of the recession. To determine that, one must look at interest rate movements over the business cycle.

Evidence: Interest Rate Movements

The two scenarios shown in figure 1 provide us with a stark contrast: In figure 1a, the decline in credit results in a rise in the interest rate, while in figure 1b, the decline in credit is associated with a decline in the interest rate. A rise in interest rates, however, is not typical in recessions. Figure 3 shows a long-term interest rate

[9] The cyclical behavior of public sector deficits and overall credit demand and its implication for interest rates is discussed in more detail in Tatom (1984 and 1985).

[10] Another factor influencing the private demand for credit, particularly business credit, is business' use of internally generated funds to finance investment. The significance of the cyclical behavior of these funds is developed by Gilbert and Ott (1985). Internal funds include retained earnings and depreciation allowances. When a firm's investment exceeds its internally supplied funds, it must turn to external financing to bridge this gap. Conversely, when internal funds exceed investment flows, a firm does not necessarily

require external financing. For nonfinancial corporations, the ratio of fixed investment (plant and equipment) and inventory investment to internal funds is greater than one during cyclical expansions, but usually falls below one during recessions, exhibiting the same cyclical nature of credit demand. During the seven recessionary periods from 1953 to 1982, the average ratio fell from 1.24 at the business cycle peak to 1.01 at its trough. A similar pattern developed in the most recent recession. In the third quarter of 1990, this ratio stood at 1.15; by the second quarter of 1991, it had fallen to 0.89. Thus, internal funds were relatively more abundant for financing investment, so less external financing, including bank credit, was demanded.

Table 1
The Federal Funds Rate and the Business Cycle

| Business cycle peak/trough | Federal Funds Rates (percent) | | |
	At business cycle peak/ trough	At closest peak[1]	At closest trough[1]
III/1957-II/1958	3.24 / 0.94	SAME	SAME
II/1960-I/1961	3.70 / 2.00	3.99 (−2)	1.68 (+2)
IV/1969-IV/1970	8.94 / 5.57	8.98 (−1)	3.86 (+1)
IV/1973-I/1975	10.00 / 6.30	10.57 (−1)	5.41 (+1)
I/1980-III/1980	15.07 / 9.83	SAME	SAME
III/1981-IV/1982	17.59 / 9.28	17.79 (−1)	8.66 (+1)
III/1990-II/1991	8.16 / 5.86	9.73 (−5)	3.77 (+4)[2]

[1] The number of quarters earlier (−) or later (+) in which the closest federal funds rate peak or trough occurs is indicated in parentheses.

[2] Latest data available.

(the yield on long-term government bonds) and a short-term interest rate (three-month Treasury bills) since 1947. Typically, both long- and short-term interest rates decline during recessions, although the long-term rate is much less cyclical than the short-term rate. Indeed, short-term interest rates sometimes reach a peak before the business cycle peak. In the latest instance, in particular, the three-month Treasury bill rate peaked at 8.54 percent in the first quarter of 1989, six quarters before the business cycle peak.

An alternative measure of short-term interest rate movements is the interest rate in the federal funds market. Briefly, the fed funds market is the market in which banks compete daily for excess reserves to meet their level of required reserves. Since the fed funds market is heavily influenced by Federal Reserve policy actions (through open market operations), some analysts believe that the behavior of this key short-term rate can provide direction about the relative tightness of bank credit.[11] Accordingly,

since credit crunches are allegedly the result of Federal Reserve policy actions to reduce credit availability, the interest rate in the federal funds market may be a more useful indicator.

Table 1 shows the average federal funds rate at the business cycle peak and trough and the nearest peak and trough of the rate itself. The fed funds rate typically peaks before or at the business cycle peak, suggesting that the type of supply shift shown in figure 1a does not occur during recessions. Like the three-month Treasury bill rate, the fed funds rate typically falls before and during recessions.

There are periods, of course, when short-term interest rates rise and the flow of credit declines. Figures 2 and 3 support the earlier discussion which indicated that such developments occurred in 1966 and, to a certain extent, over stretches during the period 1986-88. Interest rates and the demand for credit, however, tend to be procyclical, both when the economy is in recession

[11] Since the Fed and banks create money largely by acquiring debt instruments (loans and investments in debt issued by governments, firms and individuals), analysts typically use the terms money and credit interchangeably. Some analysts, like Kaufman, regard the principal channel of influence of monetary policy to be its effect on credit, not on the money stock or other measures of monetary assets. Kahn (1991) suggests that the linkage between bank loans and monetary aggregates is weak, so that a slowing in monetary growth could not have caused bank loan growth

to fall in 1990 nor could an increase in money growth have raised bank lending. Walsh (1991) argues that the recent credit crunch reflects a decline in the demand for credit, not in its supply, and that traditional monetary policy actions can change the supply of credit independently of bankers' willingness to make loans. Some analysts look at the excess reserve holdings of commercial banks as an indicator of a credit crunch, but Haubrich (1991) finds this measure an unreliable indicator of credit conditions.

Bank Credit and Economic Activity

Many analysts attribute a central, causal role in business cycle developments to bank credit movements, regardless of the reasons for such movements. For example, some researchers suggest that it is the interplay of credit and real economic activity that provides banks and monetary institutions with a potentially direct role in business cycle developments.[1] In particular, they view the growth of the money stock as purely passive, responding to the general movements of the economy, while disruptions to credit markets have real consequences for output and employment decisions. Thus, in this view, a decline in credit supply is critically important in causing and maintaining recessionary conditions.[2]

A related view assigns a special role to banks in extending credit and promoting economic activity. This view emphasizes that many firms are relatively small and have limited access to organized financial markets (for example, the commercial paper market). Accordingly, these firms' ability to expand is constrained by their access to bank credit.[3] The special role of banks, then, is to provide objective evaluations and monitoring services for the continuing viability and credit-worthiness of this relatively large sector of the economy. Thus, one dollar of bank credit is not a perfect substitute for one dollar of other credit, like a direct personal loan or the proceeds from selling commercial paper.

Bank credit carries with it the banker's certification of credit-worthiness and the banker's implicit contract for future monitoring services, the provision of financial advising and other financial services. Furthermore, extensions of bank credit to firms provide information to potential customers, financiers and other suppliers about the firm's economic prospects. In this view, not only does the supply of credit play a unique role, independent of monetary policy developments, but bank credit is also the principal linchpin for the influence of financial market developments on real economic activity.

Whether a slowing in the growth of business loans is considered a source of recessionary pressures on economic activity or simply a reflection of the recession is important to both business cycle analysts and policymakers. If the supply of credit plays a central role and is not simply a reflection of monetary aggregate movements, then policymakers may need tools to operate specifically on the credit supply. If credit movements do not change independently or exert an independent influence, however, traditional monetary policies—namely, open market operations—can reliably address cyclical problems; credit market conditions will provide no more than useful supporting information.

---

[1]This is known as the "credit view" of the transmission mechanism. Gertler (1988) reviews the literature on the effects of credit market shocks. See also Gertler and Hubbard (1988) and Bernanke (1986). Bernanke (1983) and Hamilton (1987) argue for an independent role of credit in explaining the Great Depression. The credit view is an extension of the approach taken by real business cycle (RBC) researchers. See, for example, Plosser (1991).

[2]Analyses of the unique role of bank credit sometimes focus on changes arising from changes in reserve requirements. Since reserve requirement changes affect bank credit, given deposits, it is easy to infer that bank credit can change independently of movements in money. Such analyses ignore the role of the Federal Reserve. A change in bank credit because of a change in reserve requirements, however, results in an equal and offsetting change in credit supplied by Federal Reserve Banks, as the Fed accommodates the change in demand for required reserves available by buying or selling securities or other assets. Accordingly, there is no change in the net total of credit associated with a given money stock and adjusted monetary base. Thus, movements in credit or money during such episodes are not especially unique compared with those in periods when reserve requirements do not change.

[3]A related argument stresses segmented markets, so that relatively small firms have no access to organized capital markets. Blinder and Stiglitz (1983) discuss the independent function of bank loans. James (1987) provides evidence supporting the view that bank loan extensions raise the value of firms. Judd and Scadding (1981) were the first analysts to model an independent role for bank loans to cause changes in the money stock. See Anderson and Rasche (1982), however, for a critical discussion of their model. More recently, analysts have focused on the structure of loan contracts and on loan commitments as a mechanism for avoiding credit rationing. See Duca and VanHoose (1990) for a model of the effects of loan commitments on optimal monetary policy.

and when it is not.[12] In the latest instance, in fact, interest rates and credit flows were declining well before the economy entered the recession. Thus, the evidence presented in this section is consistent with the hypothesis that, during episodes that have been characterized as credit crunches, the factors that affect the demand for credit tend to outweigh any possible effects from factors influencing the supply of credit.[13] The appropriate interpretation of credit market developments during the latest "credit crunch" is that illustrated in figure 1b.

CORROBORATING EVIDENCE: INTEREST RATE SPREADS

Economists have long pointed to the informational content of interest rate spreads for economic activity.[14] This observation stems from the fact that certain credit market instruments have similar characteristics and are more or less substitutable. Thus, if the interest rate on one instrument changes relative to its substitute, this may provide information about the behavior of credit market participants. For example, evidence of unusual bank lending behavior might be obtained by looking at the difference between bank lending rates and the rates banks pay for deposits or between rates available at banks and those available elsewhere.

This "pricing" behavior of banks has a direct application in examining credit market conditions. If banks become more reluctant to lend, then the interest rate at which they lend (bank loans)

should rise relative to the interest rate at which they borrow (deposits). This is a poor test, however, because periods when credit crunches are believed to occur also tend to be periods of recessions, when the bankruptcy rate and default rates on business loans rise. Thus, the spread between a bank's lending and borrowing rates should rise to compensate for these risks. Therefore, while interest rate spreads cannot provide definitive support for the existence of a credit crunch, they can provide some evidence about the comparability of the most recent episode by comparing the recent spreads with earlier ones.

Figure 4 shows the quarterly interest rate spreads for some relatively risky short- and long-term securities from 1947 to the present. The short-term spread measures the prime rate relative to the interest rate on three-month Treasury bills, while the long-term spread is the excess of riskier BAA bond yields over AAA bond yields.[15] Thus, a rise in the spread indicates an increase in risk.

During recession periods, the spread on risky assets rises, especially the short-term spread, because—as figure 4 shows—that is where most of the risk is. The rise in rates banks charge on loans that are inherently more risky in recessions—relative to their costs on insured deposits—need not reflect a new reluctance to lend beyond that arising from recession risk. Compared with the rise in spreads (on risky assets) in previous recessions, however, the recent spreads did not rise unusually. In fact, the long-term spread actually fell in 1991 before the recession ended.

[12]As pointed out by Gilbert (1976), one must distinguish between the demand for long-term credit and the demand for short-term credit when discussing the cyclical nature of credit demand. The demand for certain types of long-term credit tends to increase toward the end of a recession. (For example, many firms undertake bond and equity financing to lengthen the maturity of their capital structure). This rise reflects a shift away from short-term credit and not a rise in total credit demand. In effect, firms not only reduce credit demand, they also restructure their financing, shifting toward longer-term financing like bonds and especially new equity.

[13]Schreft (1990) discusses the role of credit policy in the 1980 recession. In this case, however, a rise in money demand, especially in its currency component, contributed to cyclical developments; for example, see Tatom (1981). Another recent example of a credit crunch that differs from the stereotype discussed here is the effect on world credit markets of the economic reformation of the Eastern and Central European economies. In this instance, a rise in their demand for capital is expected to raise the total demand and the real rate of interest in world credit markets. While an excess demand for credit will initially occur, and the interest rate is expected to rise, the total quantity of credit does not fall. See the Bank for International Settlements (1991) for a discussion of this credit crunch, although they do not refer to the situation as

a credit crunch and they emphasize the effects of an earlier decline in world saving in their analysis. Also see Sesit (1991) for some doubts concerning this analysis. The Council of Economic Advisers (1991) argues that real interest rates were boosted by European developments in late 1990. Such a shift in the overall demand for credit and rise in the real interest rate would be expected to create the usual credit crunch conditions in other economies as credit is diverted to Eastern Europe. This argument has not been raised, however, in the context of the 1990-91 U.S. credit crunch.

[14]For example, Brunner and Meltzer (1968) emphasized the role of credit and asset prices, or interest rate spreads, in the transmission of monetary policy over the business cycle. See Bernanke (1990) for a recent study of interest rate spreads.

[15]Bernanke (1990) points out that the spread between the BAA bond rate and the AAA bond rate is a measure of default risk.

Figure 4
Selected Interest Rate Spreads
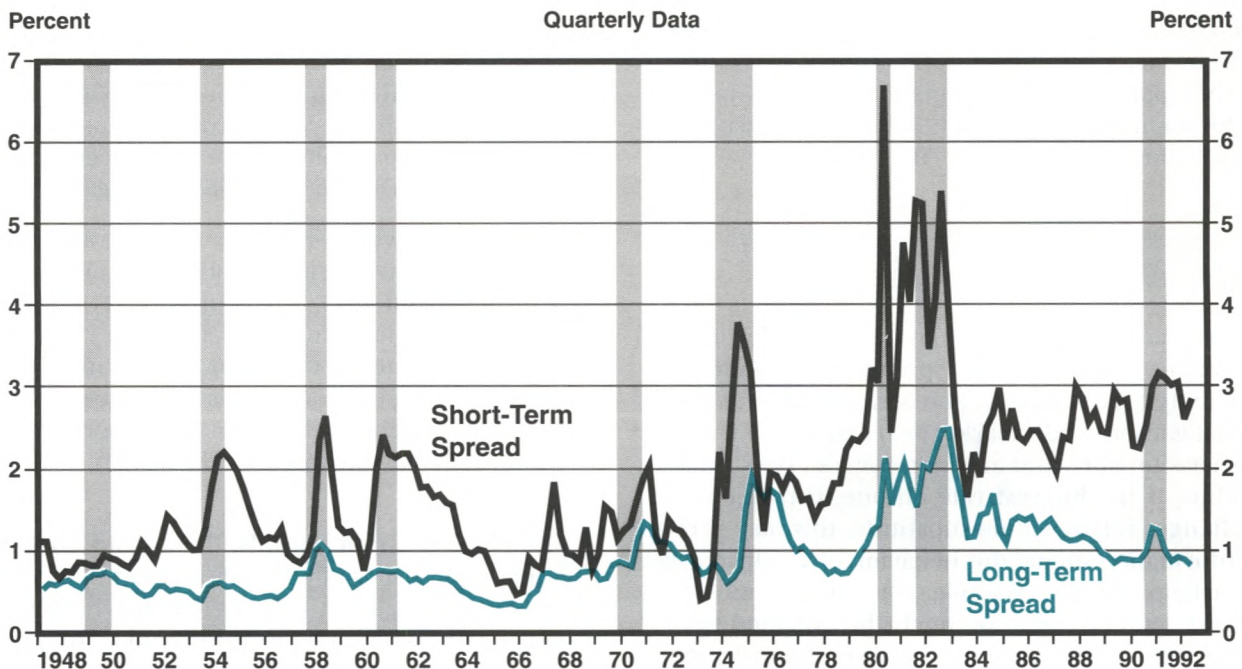
Long-term spread is the excess of the yield on BAA-rated bonds over AAA-rated bonds.
Short-term spread is the prime rate less the yield on 3-month Treasury bills.

Periods of business recession are indicated by the shaded areas.

Moreover, the short-term spread was lower than in the three previous recessions. While figure 4 shows that the price of risk rises in recessions, it provides insufficient evidence of bank unwillingness to lend.

Although bank willingness to lend is certainly a function of default risk, a bank's demand for funds to intermediate is also a factor. Figure 5 shows the spread between banks' lending and borrowing rates, where the prime rate is the lending rate and the interest rate paid on large negotiable certificates of deposit (CD) is used for the bank borrowing rate. A rise in this spread could reflect an increased reluctance to lend like that envisioned by credit crunch analysts, but this reluctance may simply reflect an assessment

of the risk arising from a recession.[16] This spread rose from 196 basis points in III/1990—when the recession started—to about 265 basis points at the end of the recession; it rose slightly more in the fourth quarter of 1991.

The rise in the spread and its recent levels, however, are both smaller than the peak spreads observed in the 1980 and 1981-82 recessions.[17] While the spread in 1990-91 was higher than in the 1969-70 and 1973-75 recessions, the rise in the spread during each of the four previous recessions was larger than the rise during the most recent recession. Therefore, if an increase in the spread between the prime rate and the CD rate is associated with an increased unwill-

[16]There is a distinct negative relationship between the prime rate and CD rate spread in figure 5 and the growth rate of C&I loans over the (available) sample period I/1970 to II/1992. The correlation between these two measures is -0.43, which is statistically significant at the 1 percent level.

[17]The difference between movements in inflationary expectations in the most recent recession and in the previous recessions may explain part of the difference. Changes in inflationary expectations, however, should affect both interest rates in the same direction.

FEDERAL RESERVE BANK OF ST. LOUIS

Figure 5
Spread Between the Prime Rate and the Rate on 3-Month Certificates of Deposit
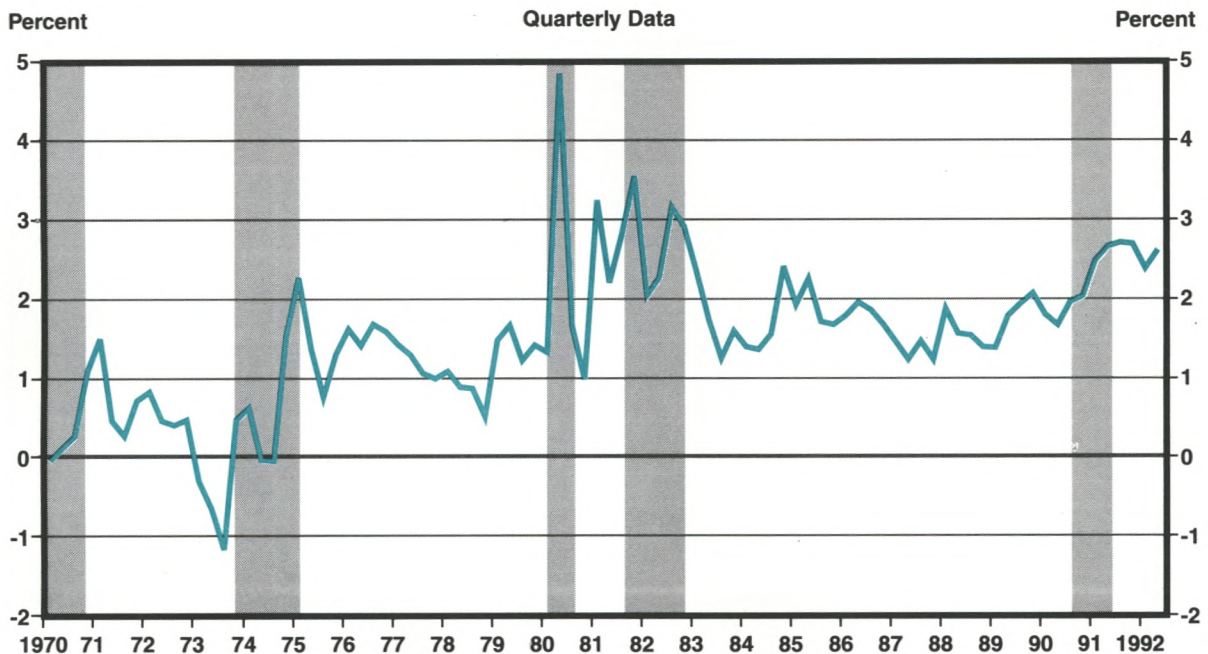
Periods of business recession are indicated by the shaded areas.

ingness to lend, then in the most recent case such unwillingness was smaller than usual.[18]

Another related measure that could indicate a rise in bankers' reluctance to lend is the spread on banks' borrowing rates compared to rates available to lenders elsewhere. Figure 6 shows how bank borrowing rates change relative to other safe, short-term rates during recessions. Because the intermediation process occurs through the initial issuance of bank liabilities (that is, a bank attracts deposits that it will then relend), a bank's demand for funds to intermediate should show up in the CD rate.

Figure 6 shows that the spread between the three-month CD rate and the three-month Treasury bill, although somewhat volatile, always declines from peak to trough. Furthermore, this spread typically falls sharply late in recessions and for a while thereafter. For example, this spread fell through most of 1970, at the beginning and end of the 1973-75 recession, at the trough in 1980 and during much of the 1981-82 recession.[19] Generally, however, this spread appears to be relatively higher at some point in each recession than it was at the business cycle peak.[20] Following a brief surge in the fourth quarter of 1990, the CD rate declined at the end of the recession and

---

[18]It is possible that the spread for some borrowers, or in some sectors or regions, rose unusually in this recession, perhaps reflecting some risk beyond those typically associated with recessions.

[19]See Gilbert (1976) for a discussion of the earlier episode and the typical weakness of loan demand early in a recovery.

[20]Figure 6 also shows that the rate on CDs generally increases at some point early in a recession. This rise probably occurs

because of the perceived riskiness of CDs due to bank failure; however, since the "too-big-to-fail" doctrine, and especially since the savings and loan bailout, CDs are about as safe as Treasury bills. Thus, this spread—as shown in figure 6—had moved to relatively low levels even before the latest recession. The decline in the spread immediately before the recession is consistent with an earlier decline in bank loan growth.

Figure 6
Spread Between the Rate on 3-Month Certificates of Deposit and the Rate on 3-Month Treasury Bills

Percent　　　　　　　　　　Quarterly Data　　　　　　　　　Percent
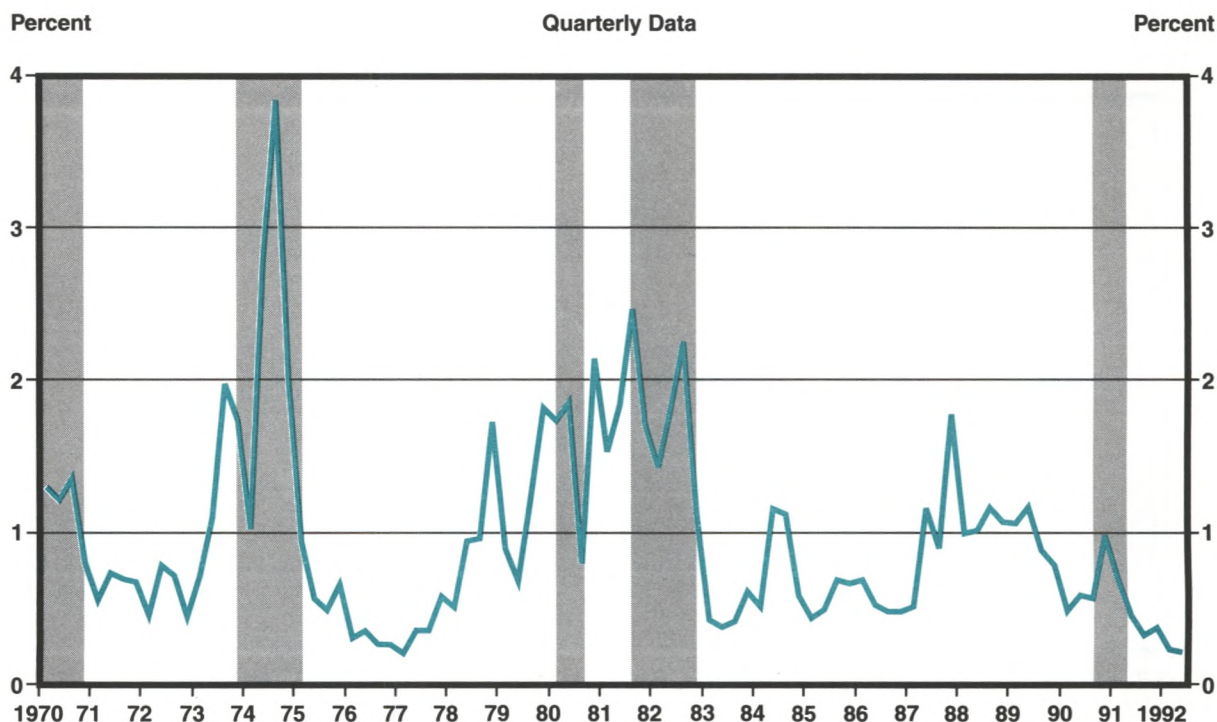
Periods of business recession are indicated by the shaded areas.

early in the recovery—just as it did previously. By the third quarter of 1991, the spread between the CD rate and the Treasury bill rate had fallen to its lowest level since 1976-77. This narrowing of the spread (that is, a relative decline in the CD rate) and its relatively low level are consistent with reductions in banks' demand for funds—either because loan demand was weaker or because bankers were reluctant to lend. In either case, however, the recent decline in this spread is much smaller than the peak-to-trough decline in the four previous recessions.

The evidence on bank pricing indicates that the spread between the lending (prime) and borrowing (CD) rate widened during the recent recession, as it typically does, although not by as much. It also indicates that the prime rate rose and the CD rate fell relative to the Treasury bill rate. The decline in the latter was not unusual, nor was it unusually large. Since it is not unusual for bank margins to rise in recessions or for the CD rate to fall at

the end of recessions, these arguments, while consistent with the credit crunch hypothesis, do not support the view that bankers have been less willing to make business loans. The critical issue in reaching this conclusion is whether interest rate spreads reacted unusually in the most recent recession. As shown in figures 4 through 6, they did not.

So far we have focused on the specifics of a credit crunch—as hypothesized by a simple model for the demand and supply of credit—and on the evidence from interest rate spreads. This evidence questions the supply-side argument behind the credit crunch hypothesis and shows the recent episode was not unusual relative to earlier instances when, others have argued, credit crunches occurred. This leaves open the issue of the source of movement in the demand for business credit at banks during these periods. The nexus between the demand for business loans and business inventories is explored below.

COMMERCIAL AND INDUSTRIAL LOANS: CAUSE FOR CONCERN?

The U.S. economy officially entered a recession in the third quarter of 1990, when real GDP contracted at a 1.6 percent rate. Output continued to contract in the fourth quarter of 1990 and the first quarter of 1991, declining at a 2.9 percent rate over the three-quarter period. One of the factors blamed for the recession was the unusual weakness of business loans.[21] Even after signs of recovery began to emerge in the spring and summer of 1991, commercial and industrial loans at banks (hereafter, business loans) remained weak. For example, nominal commercial bank loans to business fell from $642.5 billion in the fourth quarter of 1990 to $620.7 billion in the fourth quarter of 1991, after rising at only a 0.1 percent rate over the previous year. By the first quarter of 1992, business loans fell to $612.8 billion; they dropped further in the second quarter to $602.8 billion. As a result, concern about whether the recession had ended or whether the economy would have a double-dip, with real GDP resuming its earlier decline, continued well into the winter of 1991-92.

There are at least two reasons why analysts are concerned about business loan growth. The first is the concern raised by proponents of the credit crunch view of the recent recession: slow growth of bank lending could reflect unusual structural problems in banking.[22] Second, slow growth in business loans is an indication that business activity is not expanding. If businesses are reluctant to expand, the potential for economic recovery is jeopardized. What is absent from the discussion, however, is the fact that business loans typically grow more slowly in recessions. This is examined in greater detail in the next section.

WHY DOES THE FLOW OF CREDIT DECLINE IN A RECESSION?

When sales slow, firms have an incentive to reduce production and employment to avoid an accumulation of undesired inventory. Such a reduction in output and employment constitutes a typical recession. But firms also alter their other investment decisions during recessions. For instance, firms also reduce their demand for new plant and equipment based on lower desired output levels and growing excess capacity. As a result, overall investment and its financing tend to decline during recessions.

The Decline in Inventory Investment in Recessions

The role of inventory investment in recessions is especially important. Indeed, one principal type of recession is called an inventory recession because of the central role of changes in inventories.[23] In an inventory recession, an *unanticipated* decline in sales growth leads to an undesired build-up of inventory followed by adjustments to production and employment. As firms reduce inventory to eliminate the initial excess, inventory investment becomes negative; however, such investment must eventually be restored to continue meeting the slower pace of expected sales. This eventual rise in inventory investment implies that some firms' production and sales have risen, thereby setting in motion an overall cyclical expansion. Thus, an inventory recession is characterized by, first, an initial rise in inventories relative to sales (before, or in, the initial stage of a recession), second, a subsequent decline in inventory investment to a negative pace, and finally, a rebound in inventory investment before or at the recession's end.

---

[21]See the references in footnote 1. For differing analyses, see Brenner and Schmidt (1991), Corcoran (1992), Furlong (1991) and Bacon and Wessel (1991), Heinemann (1991), Jordan (1992), Meltzer (1991), Passell (1991) and Prowse (1991). Syron (1991) attributes the weakness in bank loans and economic activity to a shortage of capital induced by higher capital requirements and bank losses. Parry (1992) argues that policy-related changes in the real cost of intermediation have been appropriate, even if they have permanently changed the extent of bank intermediation.

[22]The Council of Economic Advisers (1991) discusses several reasons for the decline in bank credit growth. It notes, however, that the substitution of other debt should have offset the decline in bank lending. Strongin (1991) also stresses that credit reductions have been offset by increased equity financing. Bernanke and Lown (1991) point to an absence of business loan growth. They indicate that other sources of business loans show a decline that is unusually large for

them during recessions, reinforcing the view that the overall demand for business loans fell, not the supply. They also provide evidence that New England banks' efforts to raise capital had a relatively small impact on bank lending, although not necessarily on business loans. Feldstein (1992) argues that the imposition of risk-based capital standards has restricted banks' ability to intermediate.

[23]The cyclical behavior of inventory investment and inventory recessions are described in more detail in Tatom (1977). The first effort to formally model the inventory cycle is Metzler (1941). Blinder and Maccini (1991) provide a recent review of the state of research on inventories.

Figure 7
Change in Business Inventories

Billions of 1987 dollars　　　　Quarterly Data　　　　Billions of 1987 dollars

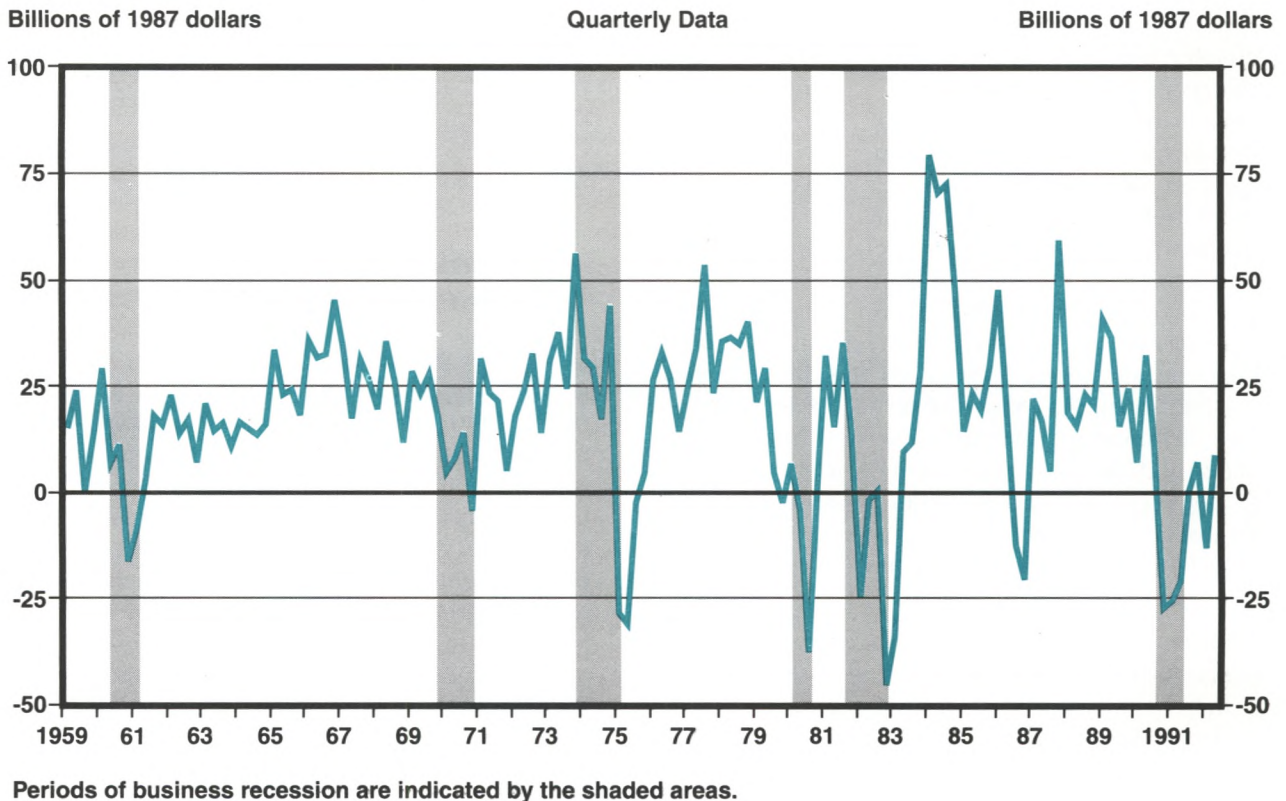**Periods of business recession are indicated by the shaded areas.**

Figure 7 shows the change in inflation-adjusted business inventories since 1959. Inventory investment does not always rise unusually at the business cycle peaks. Indeed, in half the instances shown, including the latest, inventory investment falls at the business cycle peak. Also, it is uncommon for inventory investment to register increases at the end of a recession or in the trough quarter. It is not unusual for it to rise in the first quarter of the recovery. Such a rise occurred in seven of the past eight recessions, although, in four of these cases, inventory investment remained negative in the quarter following the business cycle trough. Therefore, while all recessions do not conform to the stereotypical inventory recession pattern, there is no question that movements in inventory investment play a central role in recessions. The decline in overall investment and real GDP

in recessions is concentrated most heavily in their inventory investment component.

Table 2 shows that the decline in the constant-dollar change in business inventories accounts for much of the business cycle peak-to-trough decline in real GDP. Excluding the relatively large swings in inventory investment (as a share of the decline in GDP) in the 1960-61 and 1969-70 recessions, the decline in inventory investment in the most recent recession (48.2 percent) was fairly typical. It exceeded that in three of the previous eight recessions, although inventory investment already had declined rather substantially from early in 1989 to the business cycle peak in III/1990. For the recent period of decline in real GDP (II/1990 to I/1991), the decline in inventory investment of 54.6 percent of the production decline exceeded that in four of the previous eight recessions.

Table 2
The Decline in Inventory Investment in Recessions

Recession peak-trough	Change in real inventory investment[1]	Change in real GDP[1]	Column 1 as a percent of column 2
IV/1948-IV/1949	$−28.3 billion	$−20.9 billion	135.4%
III/1953-II/1954	−11.1 (−18.4)	−37.4 (−43.3)	29.7 (42.5)
III/1957-II/1958	−20.1 (−22.5)	−44.9 (−53.1)	44.8 (42.4)
II/1960-I/1961	−15.7 (−45.5)	5.7 (−15.8)	−275.4 (288.0)
IV/1969-IV/1970	−22.5 (−19.8)	−1.8 (−25.0)	1250.0 (79.2)
IV/1973-I/1975	−84.7	−135.1	62.7
I/1980-III/1980	−44.3 (−10.7)	−97.3 (−98.2)	45.5 (10.9)
III/1981-IV/1982	−80.6 (−35.0)	−104.9 (−110.1)	76.8 (31.8)
III/1990-II/1991	−31.6 (−57.9)	−65.5 (−106.0)	48.2 (54.6)

[1]Prior to 1960, data are expressed in 1982 dollars. From 1960 to the present, data are in 1987 dollars. Numbers in parentheses correspond to the data for the respective peak-to-trough periods for real GDP: II/53-II/54, III/57-I/58, I/60-IV/60, III/69-II/70, I/80-II/80, III/81-III/82 and II/90-I/91.

The Linkage Between Business Loans & Business Inventories

Inventory decisions are also central to business loan behavior during recessions.[24] Since banks tend to hold short-term liabilities, which in large part are payable on demand, they have a strong incentive to hold relatively short-term loans. Thus, bank loans and lines of credit to business are principally related to business financing of short-term assets, such as inventories. Inventory assets are crucial because they are superior collateral to receivables. Moreover, the value of receivables can disappear more easily than that of inventory in the event of default; inventory also can be taken over and liquidated more easily.

Figure 8 shows the stock of business inventories and business loans since 1959, both measured in nominal terms.[25] Business loans and business inventories move together over time. For example, both rise slowly until 1973, then accelerate sharply until late 1974. At the end of the 1973-75 recession and during the early quarters of the recovery, business loans declined along with business

inventories. The growth rates of each series appear to slow in the 1980 recession and at the end of the 1981-82 recession. Inventory growth is unusually slow from early 1982 to the end of 1986 compared with loan growth, however. Over this period, business loans rose from about 44 percent of inventories to about 59 percent, reflecting the slowing in inventory growth.

Since the end of 1986, both business loans and inventories have grown at about the same rate, keeping business loans at about 60 percent of inventories. Both measures declined in the recent recession, following a slowing in growth in 1989 and 1990. For example, from the first quarter of 1987 to the first quarter of 1989, business loans rose at a 5.9 percent annual rate, while business inventories rose at a 7.9 percent rate. During the same interval, overall nominal final sales in the U.S. economy grew at a 7.9 percent rate.[26] Over the next six quarters, final sales growth slowed to a 5.7 percent rate, while inventory growth slowed to a 4.3 percent rate and business loans slowed to a 3.0 percent rate

[24]For a textbook treatment of the interplay between business loans and business inventories, see Goldfeld (1966).

[25]The business loan series begins in 1959, but is only available for the last Wednesday of each month until 1973. After 1973, the data are quarterly averages of Wednesday data. The stock of business inventory is end-of-quarter data, based on estimates of the change in business inventory. The value

of business inventory typically is much larger than that of business loans.

[26]Final sales is the sum of gross private domestic fixed investment, personal consumption expenditures, net exports and government purchases. GDP is the sum of final sales and the change in business inventories.

Figure 8
Business Loans and Business Inventories

Quarterly Data Billions of dollars

Periods of business recession are indicated by the shaded areas.
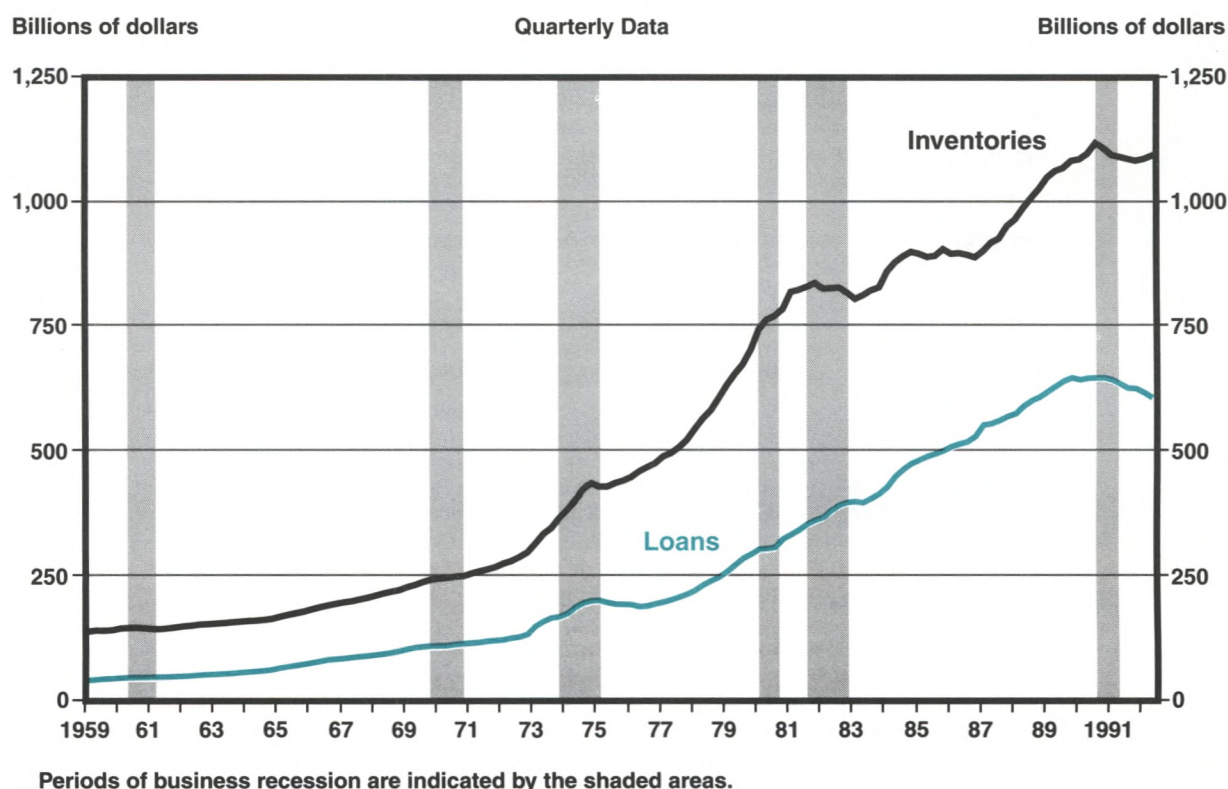
of advance. Finally, during the recent recession, from the third quarter of 1990 to the second quarter of 1991, final sales growth slowed to a 3.1 percent rate and inventory fell $27.6 billion, or at a 3.3 percent rate. The decline in business loans over the recession totaled $11.9 billion— a 2.5 percent rate of decline.

The link between nominal business loans and nominal business inventories is more systematic than the simple upward trends in figure 8 might suggest. Quarter-to-quarter changes in business loans are statistically related to quarter-to-quarter changes in the stock of inventories in a significant and positive fashion.

Figure 9 shows the change in business inventories and the change in business loans (both in current dollars); the two series are expressed as percentages of GDP to scale the data, but this has no effect on the close visual relationship between the two series. The correlation coefficient for these changes over

the period II/1959-II/1992 is 0.52, which is statistically significant at the 5 percent level.

Stronger evidence for the relationship between business loans and business inventories can be obtained from causality analysis. In simple terms, causality analysis examines the statistical direction of influence from one variable to another; in particular, it assesses whether there is a statistically significant, temporal sequence between changes in one measure and another. This issue is addressed in the appendix. The results there support the hypothesis that changes in business loans are significantly influenced by changes in business inventories. When inventory investment falls, as it does in every recession, it is not surprising, therefore, to see an accompanied weakness in commercial bank business loans.

CONCLUSION

The recent decline in the growth of business loans at commercial banks reflects normal

Figure 9
Quarterly Changes in Business Inventories and Business Loans (Percent of GDP)

Percent

Percent
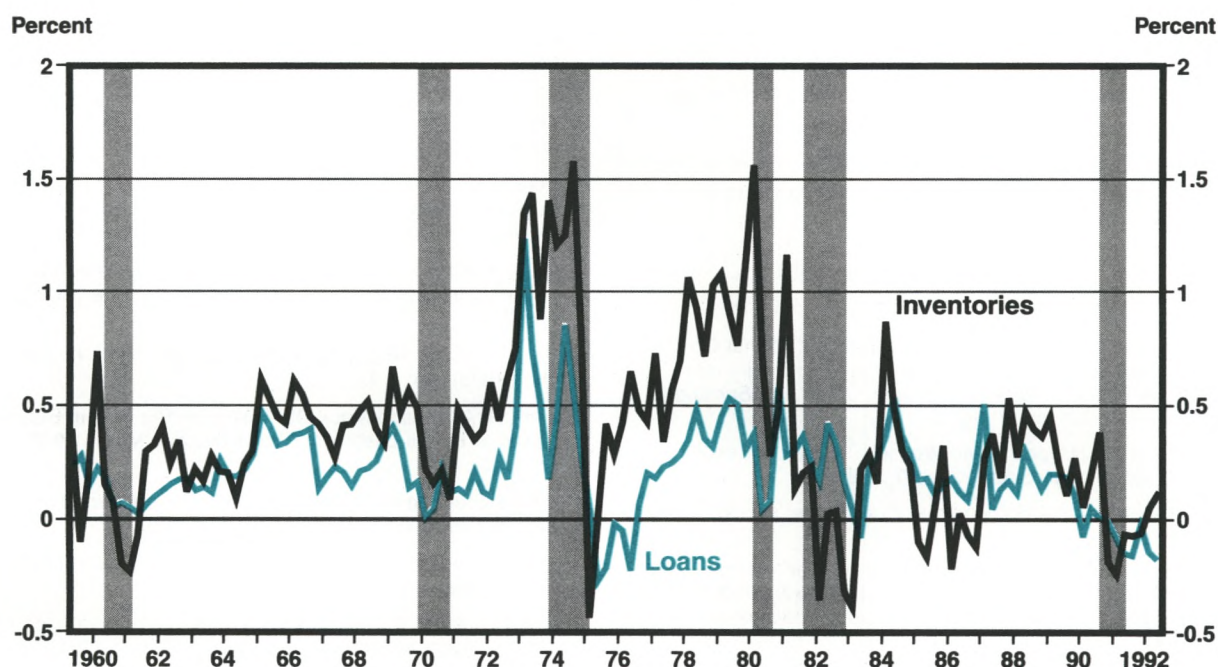
Periods of business recession are indicated by the shaded areas.

cyclical phenomena. While this decline has been referred to as a credit crunch, it is unlikely to have occurred because bankers were unusually reluctant to make business loans, as is sometimes suggested. Instead, as in earlier credit crunches/recessions, the decline most likely originated on the credit demand side.

No doubt there are individual cases in which supply factors have been important in reducing credit availability. Indeed, some researchers have alleged that such occurrences explain, to a small extent, business loan slowings in some parts of the country owing to changes in bank capital requirements or other regulatory changes. These analyses generally do not control for the normal cyclical phenomena addressed here, however.

The decline in business loan growth in recessions is due, in large part, to the cyclical nature of business loan demand. Bank loans are typically short-term collateralized loans, so that the prime commercial asset that is financed by bank credit

is inventories. The evidence presented here suggests that business loans and business inventory holdings are very closely related statistically, so that business loans and inventory move up or down in tandem. Since businesses typically reduce their desired inventory holdings during recessions, business loans at banks tend to decline as well.

An additional consideration is the recent movement in interest rates and interest rate spreads. As in earlier periods of so-called credit crunches, the recent decline in business loans has been accompanied by reductions in interest rates, particularly short-term rates. This behavior is inconsistent with a shortage of credit from a simple supply and demand perspective. While interest rate spreads for "risky" credit have risen recently, including the difference between bank lending and borrowing rates, this is also a normal cyclical phenomenon. This spread, as well as that between the prime rate and the three-month Treasury bill rate, have not been unusually large

compared with previous recessions, nor has their *increase* been unusually large. Although the spread between the large CD rate and three-month Treasury bill rate has fallen, this too is not unusual in a recession nor in the initial stages of a recovery.

To summarize, the theory and evidence presented here suggests that recessions cause inventory demand and the growth of business borrowing to slow. To the extent that this argument and evidence characterize recent developments, the solution to the recent decline in credit growth is likely to be found, as usual, in a restoration of business inventory accumulation and its financing.

REFERENCES

Anderson, Richard G., and Robert H. Rasche. "What Do Money Market Models Tell Us about How to Implement Monetary Policy?" *Journal of Money, Credit and Banking* (November 1982), pp. 796-828.

Bacon, Kenneth H., and David Wessel. "Wary Lenders," *Wall Street Journal*, September 30, 1991.

Bank for International Settlements. *61st Annual Report* (June 10, 1991).

Bernanke, Ben S. "On the Predictive Power of Interest Rates and Interest Rate Spreads," Federal Reserve Bank of Boston *New England Economic Review* (November/December 1990), pp. 51-68.

_____. "Alternative Explanations of the Money-Income Correlation," *Real Business Cycles, Real Exchange Rates and Actual Policies*, Carnegie-Rochester Conference Series on Public Policy (Autumn 1986), pp. 49-100.

_____. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression," *American Economic Review* (June 1983), pp. 257-76.

Bernanke, Ben S., and Cara S. Lown. "The Credit Crunch," Brookings Papers on Economic Activity (2:1991), pp. 205-47.

Blinder, Alan S., and Louis J. Maccini. "Taking Stock: A Critical Assessment of Recent Research on Inventories," *Journal of Economic Perspectives* (Winter 1991), pp. 73-96.

Blinder, Alan S., and Joseph E. Stiglitz. "Money, Credit Constraints, and Economic Activity," *American Economic Review* (May 1983), pp. 297-302.

Brenner, Joel Glenn, and Susan Schmidt. "Bankers Say There's No 'Credit Crunch'," *Washington Post*, October 12, 1991.

Brunner, Karl, and Allan H. Meltzer. "Liquidity Traps for Money, Bank Credit, and Interest Rates," *Journal of Political Economy* (January/February 1968), pp. 1-37.

Burger, Albert E. "A Historical Analysis of the Credit Crunch of 1966," this *Review* (September 1969), pp. 13-30.

Corcoran, Patrick J. "The Credit Slowdown of 1989-91: The Role of Demand," paper presented at the 28th Annual Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, May 6-8, 1992.

Council of Economic Advisers. *Economic Report of the President* (Government Printing Office, 1992).

_____. *Economic Report of the President* (Government Printing Office, 1991).

Duca, John V., and David D. VanHoose. "Loan Commitments and Optimal Monetary Policy," *Journal of Money, Credit and Banking* (May 1990), pp. 178-94.

Feldstein, Martin. "Revise Bank Capital Standards Now," *Wall Street Journal*, March 6, 1992.

Forrestal, Robert P. "Policy Implications of a Credit Crunch," speech delivered at the Conference on "Credit Crunches—Causes and Cures," Wellington, New Zealand, August 16, 1991.

Furlong, Fred. "Financial Constraints and Bank Credit," Federal Reserve Bank of San Francisco *Weekly Letter* (May 24, 1991).

Gertler, Mark. "Financial Structure and Aggregate Economic Activity: An Overview," *Journal of Money, Credit and Banking* (August 1988), pp. 559-88.

Gertler, Mark, and R. Glenn Hubbard. "Financial Factors in Business Fluctuations," National Bureau of Economic Research, Working Paper No. 2758 (November 1988).

Gilbert, R. Alton. "Requiem for Regulation Q: What It Did and Why It Passed Away," this *Review* (February 1986), pp. 22-37.

_____. "Bank Financing of the Recovery," this *Review* (July 1976), pp. 2-9.

Gilbert, R. Alton, and Mack Ott. "Why the Big Rise in Business Loans at Banks Last Year?" this *Review* (March 1985), pp. 5-13.

Goldfeld, Stephen M. *Commercial Bank Behavior and Economic Activity* (North Holland: Amsterdam, 1966).

Greenspan, Alan. "Statements to Congress," *Federal Reserve Bulletin* (May 1991), pp. 300-310.

Hamilton, James D. "Monetary Factors in the Great Depression," *Journal of Monetary Economics* (March 1987), pp. 145-69.

Haubrich, Joseph G. "Do Excess Reserves Reveal Credit Crunches?" Federal Reserve Bank of Cleveland *Economic Commentary* (July 15, 1991).

Heinemann, H. Eric. "The 'Credit Crunch' Is a Red Herring," *Christian Science Monitor*, October 1, 1991.

James, Christopher. "Some Evidence on the Uniqueness of Bank Loans," *Journal of Financial Economics* (December 1987), pp. 217-35.

Jordan, Jerry L. "The Credit Crunch: A Monetarist's Perspective," paper presented at the 28th Annual Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, May 7, 1992.

Judd, John P., and John L. Scadding. "Liability Management, Bank Loans, And Deposit 'Market' Disequilibrium," Federal Reserve Bank of San Francisco *Economic Review* (Summer 1981), pp. 21-44.

Kahn, George A. "Does More Money Mean More Bank Loans?" Federal Reserve Bank of Kansas City *Economic Review* (July/August 1991), pp. 21-31.

Kaufman, Henry. "Credit Crunches: The Deregulators Were Wrong," *Wall Street Journal*, October 9, 1991.

LaWare, John P. "Setting the Global Scene: A Global Credit Crunch?" speech delivered at the Conference on "Credit Crunches—Causes and Cures," Wellington, New Zealand, August 15, 1991.

Meltzer, Allan H. "There Is No Credit Crunch," *Wall Street Journal*, February 8, 1991.

Metzler, Lloyd A. "The Nature and Stability of Inventory Cycles," *Review of Economic Statistics* (August 1941), pp. 113-29.

O'Brien, Paul Francis, and Frank Browne. "A Credit Crunch? The Recent Slowdown in Bank Lending and Its Implications for Monetary Policy," Organization for Economic Cooperation and Development Working Paper No. 107, 1992.

Parry, Robert T. "The Problem of Weak Credit Markets: A Monetary Policymaker's View," Federal Reserve Bank of San Francisco Weekly Letter (January 3, 1992).

Passell, Peter. "Capital Crunch: Capitol Quandary," New York Times, November 6, 1991.

Plosser, Charles I. "Money And Business Cycles: A Real Business Cycle Interpretation," in Michael T. Belongia, ed., Monetary Policy on the 75th Anniversary of the Federal Reserve System (Kluwer, 1991), pp. 245-74.

Prowse, Michael. "The Credit Crunch as Scapegoat," Financial Times, November 18, 1991.

Reuters News Service. "Japan Predicts Capital Crunch," International Herald Tribune, December 4, 1991a.

_____. "France Sees No 'Crunch' in Lending Slowdown," International Herald Tribune, December 12, 1991b.

Schreft, Stacey L. "Credit Controls: 1980," Federal Reserve Bank of Richmond Economic Review (November/December 1990), pp. 25-55.

Sesit, Michael R. "Fears of a Global Credit Crunch Are Overdone, Some Analysts Say," Wall Street Journal Europe, November 5, 1991.

Strongin, Steven. "Credit Flows and the Credit Crunch," Federal Reserve Bank of Chicago, Chicago Fed Letter (November 1991).

Syron, Richard F. "Are We Experiencing a Credit Crunch?" Federal Reserve Bank of Boston New England Economic Review (July/August 1991), pp. 3-10.

Tatom, John A. "Two Views of the Effects of Government Budget Deficits in the 1980s," this Review (October 1985), pp. 5-16.

_____. "A Perspective on the Federal Deficit Problem," this Review (December 1984), pp. 5-17.

_____. "Energy Prices and Short-Run Economic Performance," this Review (January 1981), pp. 3-17.

_____. "Inventory Investment in the Recent Recession and Recovery," this Review (April 1977), pp. 2-9.

Walsh, Carl E. "The Credit Crunch and The Real Bills Doctrine," Federal Reserve Bank of San Francisco Weekly Letter (May 3, 1991).

Wojnilower, Albert M. "The Central Role of Credit Crunches in Recent Financial History," Brookings Papers on Economic Activity (2:1980), pp. 277-339.

Appendix
Business Loans and Business Inventories: Some Statistical Evidence

The relationship between inventory decisions and the growth of business loans can be examined using a straightforward causality test, which tests the statistical significance of the temporal sequence between measures that are hypothesized to be related. The growth rate of business loans ($\dot{L} = 400\ \Delta\ln L$) and business inventories ($\dot{I} = 400\ \Delta\ln I$) can be examined to see (1) whether one measure "causes" the other; (2) whether they each cause the other (bi-directional causality) or (3) whether they are statistically independent.[1] This is done by examining the statistical significance of past values of one measure in explaining the other, controlling for the time series properties of the other.

Considered alone, the growth in business loans ($\dot{L}$) and in business inventories ($\dot{I}$) are first- and third-order autoregressive series, AR1 and AR3, respectively, which means that current values of each are highly related to their own past value one quarter earlier and three quarters earlier, respectively, but not to earlier changes.

---

[1] The statistical analysis presented here uses variables measured in nominal terms because the credit crunch hypothesis concerns the effects of nominal bank credit and the latter finances, in part, nominal inventory. When the procedures are performed on the same variables measured in constant-dollar terms, the results are essentially the same and the conclusions are not altered.

To conduct the causality tests, up to eight past values of each variable were added to the autoregressive model of the other to see if one variable is statistically significant in explaining future values of the other.

For loan growth during the period from II/1959 to II/1992, the only statistically significant relationship between the two measures is:

$$(1) \quad \dot{L}_t = 2.124 + 0.236 \, \dot{I}_{t-1} + 0.565 \, \dot{L}_{t-1}$$
$$ \quad \quad (2.77) \quad (2.47) \quad \quad (7.13)$$
$$ \quad \overline{R}^2 = 0.46 \quad S.E. = 5.586 \quad D.W. = 1.85$$

This equation indicates that inventory growth causes business loan growth, because the coefficient on the change in inventory (0.236) is significantly different from zero at a 5 percent level according to the relatively high value of the t-statistic given in parentheses. The adjusted R^2 and Durbin-Watson (D.W.) statistics suggest, respectively, that there is a strong relationship and that the first lag value of the dependent variable ($\dot{L}_{t-1}$) is sufficient for removing any problematical serial correlation in the errors of the estimated equation.[2] No other lagged value of $\dot{I}$ or $\dot{L}$ or combination of lagged values is statistically significant.

While the evidence presented here indicates that changes in inventories result in changes in business loans, there is nonetheless some evidence of reverse causality—although the effect is ephemeral in nature. In particular, the best time series representation for inventory growth is an AR(3) model. The inventory equation similar to equation 1, for the period I/1960 to II/1992, is:

$$(2) \quad \dot{I}_t = 1.157 + 0.560 \, \dot{I}_{t-1} - 0.079 \, \dot{I}_{t-2} + 0.274 \, \dot{I}_{t-3}$$
$$ \quad (1.88) \quad (6.32) \quad (-0.80) \quad (3.31)$$
$$ \quad \quad + 0.264 \, \dot{L}_{t-1} - 0.216 \, \dot{L}_{t-2}$$
$$ \quad \quad \quad (3.76) \quad \quad (-3.02)$$
$$ \quad \overline{R}^2 = 0.55 \quad S.E. = 4.200 \quad D.W. = 1.92$$

Since the lagged terms on the change in business loans are approximately equal in magnitude and opposite in sign, it is possible that only a transitory effect is present, so that a rise in loan growth has no effect on inventory growth after two quarters.

To test this hypothesis, the coefficients on the two lags of business loan growth were constrained to sum to zero. The resulting estimate is:

$$(3) \quad \dot{I}_t = 1.318 + 0.580 \, \dot{I}_{t-1} - 0.067 \, \dot{I}_{t-2}$$
$$ \quad (2.33) \quad (6.98) \quad (-0.69)$$
$$ \quad \quad + 0.281 \, \dot{I}_{t-3} + 0.241 \, \Delta\dot{L}_{t-1}$$
$$ \quad \quad (3.42) \quad \quad (3.96)$$
$$ \quad \overline{R}^2 = 0.55 \quad S.E. = 4.189 \quad D.W. = 1.94$$

A test of the restriction (that is, regressions 3 and 2, respectively) yields an $F_{1,124} = 0.42$, which is not statistically significant. Thus, one cannot reject the hypothesis that there is a transitory causal link from increased loan growth to increased inventory growth. After two quarters, however, a change in business loan growth has no statistically significant effect on business inventory growth.[3] These results suggest that policies designed to increase bank lending, taken alone, are unlikely to raise inventory investment, which is one of the principal effects expected by proponents of the credit crunch/recession linkage. Moreover, these results also reaffirm the behavior of business loans during and immediately following a recession.[4]

---

[2] Both business loans and inventory are integrated of order one and are cointegrated according to tests that are not reported here.

[3] Equations 1 and 2, or 1 and 3, were also estimated using the seemingly-unrelated-regression method to allow for contemporaneous correlations of the error terms. This does not affect the causality conclusions. The inclusion of the lagged residual from an estimated cointegrating vector for the levels also does not alter these results.

[4] According to Gilbert and Ott (1985), business loans at large banks typically remain at their trough level during the first year of the recovery, before giving way to moderate growth in the second year.

John W. Keating

John W. Keating, assistant professor, Department of Economics, Washington University in St. Louis, was a visiting scholar at the Federal Reserve Bank of St. Louis while this paper was written. Richard I. Jako provided research assistance.

Structural Approaches to Vector Autoregressions

T HE VECTOR AUTOREGRESSION (VAR) model of Sims (1980) has become a popular tool in empirical macroeconomics and finance. The VAR is a reduced-form time series model of the economy that is estimated by ordinary least squares.[1] Initial interest in VARs arose because of the inability of economists to agree on the economy's true structure. VAR users thought that important dynamic characteristics of the economy could be revealed by these models without imposing structural restrictions from a particular economic theory.

Impulse response functions and variance decompositions, the hallmark of VAR analysis, illustrate the dynamic characteristics of empirical models. These dynamic indicators were initially obtained by a mechanical technique that some believed was unrelated to economic theory.[2] Cooley and LeRoy (1985), however, argued that this method, which is often described as atheoretical, actually implies a particular economic structure that is difficult to reconcile with economic theory.

This criticism led to the development of a "structural" VAR approach by Bernanke (1986), Blanchard and Watson (1986) and Sims (1986). This technique allows the researcher to use economic theory to transform the reduced-form

VAR model into a system of structural equations. The parameters are estimated by imposing contemporaneous structural restrictions. The crucial difference between atheoretical and structural VARs is that the latter yield impulse responses and variance decompositions that can be given structural interpretations.

An alternative structural VAR method, developed by Shapiro and Watson (1988) and Blanchard and Quah (1989), utilizes long-run restrictions to identify the economic structure from the reduced form. Such models have long-run characteristics that are consistent with the theoretical restrictions used to identify parameters. Moreover, they often exhibit sensible short-run properties as well.

For these reasons, many economists believe that structural VARs may unlock economic information embedded in the reduced-form time series model. This paper serves as an introduction to this developing literature. The VAR model is shown to be a reduced-form for a linear simultaneous equations model. The contemporaneous and long-run approaches to identifying structural parameters are developed. Finally, estimates of contemporaneous and long-run structural VAR models using a common set of macroeconomic variables are presented. These models

---

[1] A VAR can be derived for a subset of the variables from a linear structural model. Furthermore, it is a linear approximation to any nonlinear structural model. The accuracy of the VAR approximation will depend on the features of the nonlinear structure.

[2] A Choleski decomposition of the covariance matrix for the VAR residuals.

---

for unit roots, their existence is controversial. VARs can accommodate either side of the debate, however.[8]

The VAR is a general dynamic specification because each variable is a function of lagged values of all the variables. This generality, however, comes at a cost. Because each equation has many lags of each variable, the set of variables must not be too large. Otherwise, the model would exhaust the available data.[9] If all shocks have unit roots, equation 6 is estimated. If all shocks are stationary, equation 5 is used.[10] If some shocks have temporary effects while others have permanent effects, the empirical model must account for this.

Recently, Blanchard and Quah (1989) have estimated a VAR model where some variables were assumed to be stationary while others had unit roots. Alternatively, King, Plosser, Stock and Watson (1991) use a cointegrated model, where all the variables are difference stationary but some linear combinations of the variables are stationary. The stationary linear combinations are constructed by cointegration regressions prior to VAR estimation. They impose the cointegration constraints using the vector error-correction model of Engle and Granger (1987). Sometimes unit-root tests combined with theory suggest the coefficients for stationary linear combinations. Shapiro and Watson (1988), for example, present evidence that the nominal interest rate and inflation each have a unit root while the difference between these two variables is stationary. They impose the cointegration constraint by selecting this noisy proxy for the real rate of interest as a variable for the model.

Unrestricted versions of the VAR model (and the error-correction model) are estimated by ordinary least squares (OLS) because Zellner (1962) proved that OLS estimates of such a system are consistent and efficient if each equation has precisely the same set of explanatory variables. If the underlying structural model provides a set of over-identifying restrictions on the reduced form, however, OLS is no longer optimal.[11] The simultaneous equations system in

a contemporaneous structural VAR, however, generally does not impose restrictions on the reduced form.

An alternative approach of Doan, Litterman and Sims (1984) estimates the VAR in levels with a Bayesian prior placed on the hypothesis that each time series has a unit root. The Bayesian VAR model permits more lags by imposing restrictions on the VAR coefficients, reducing the number of estimated parameters (called hyper-parameters). The reduction in parameters contributes to the Bayesian model's propensity to yield superior out-of-sample forecasts compared with unrestricted VARs with symmetric lag structure.

CONTEMPORANEOUS STRUCTURAL VAR MODELS

It is clear from equations 5 and 6 that, if the contemporaneous parameters in A and D were known, the dynamic structure represented by the parameters in C(L) could be calculated from the estimated VAR coefficients, that is, $C(L) = A\beta(L)$. Furthermore, the structural shocks, ε_t, could be derived from the estimated residuals, that is, $\varepsilon_t = D^{-1}Ae_t$. Because the coefficients in A and D are unknown, identification of structural parameters is achieved by imposing theoretical restrictions to reduce the number of unknown structural parameters to be less than or equal to the number of estimated parameters of the variance-covariance matrix of the VAR residuals. Specifically, the covariance matrix for the residuals, Σ_e, from either equation 5 or equation 6 is

$$(7) \ \Sigma_e = E[e_t e_t'] = A^{-1}DE[\varepsilon_t\varepsilon_t']D'A'^{-1} = A^{-1}D\Sigma_\varepsilon D'A'^{-1},$$

where E is the unconditional expectations operator, and Σ_ε is the covariance matrix for the shocks.

An OLS estimate of the VAR provides an estimate of Σ_e that can be used with equation 7 to obtain estimates of A, D and Σ_ε. The contemporaneous structural approach imposes restrictions on these three matrices. There are n^2 elements in A,

---

[8] Alternatively, the unit root could result from parameters in the dynamic structural model.

[9] The lag structure of a VAR can be shown to represent various sources of economic dynamics. Structural models with rational expectations predict restrictions on the VAR model. Dynamics in these models are often motivated by the costs of adjustment to desired or equilibrium positions. The lag structure of the VAR can also be motivated by dynamic processes for structural disturbances.

[10] VAR lag length is often selected by statistical criteria such as the modified likelihood ratio test of Sims (1980).

[11] A two-step structural VAR estimator will generally not be efficient if there are structural restrictions for C(L) since this implies restrictions on $\beta(L)$. For example, Sargent (1979) derives restrictions on VAR coefficients from a particular model of the term structure of interest rates under rational expectations. The full structural system is estimated by maximum likelihood.

n^2 elements in D, and $n(n+1)/2$ unique elements in Σ_ε, but only $n(n+1)/2$ unique elements in Σ_e. The maximum number of structural parameters is equal to the number of unique elements in Σ_e. Thus, a structural model will not be identified unless at least $2n^2$ restrictions are imposed on A, D and Σ_ε.

Often these restrictions are exclusion restrictions; of course, that need not be the case. Typically, Σ_ε is specified as a diagonal matrix because the primitive structural disturbances are assumed to originate from independent sources. The remaining parameters are identified by imposing additional restrictions on A and D. The main diagonal elements of A are set to unity because each structural equation is normalized on a particular endogenous variable. The main diagonal for D has this same specification since each equation has a structural shock. These normalizations provide $2n$ restrictions. Identification requires at least $3n(n-1)/2$ additional restrictions based on economic theory. Alternatively, the restrictions may be based on the contemporaneous information assumed available to particular economic agents following Sims (1986). Keating (1990) and West (1990) extend this approach by showing how rational expectations restrictions can be imposed in the contemporaneous structural VAR framework. Except for Bernanke (1986) and Blanchard (1989), existing models typically have not attempted to identify the structural parameters in D. Hence, D is usually taken to be the identity matrix, leaving at least $n(n-1)/2$ additional identifying restrictions to be imposed on A.

A two-step procedure is used to estimate structural VAR models. First, the reduced-form VAR, with enough lags of each variable to eliminate serial correlation from the residuals, is estimated with OLS. Next, a sufficient number of restrictions is imposed on A, D and Σ_ε to identify these parameters. This paper obtains the parameters in equation 7 with an algorithm for solving a nonlinear system of equations. Blanchard and Quah (1989) use this approach to estimate a structural VAR model.[12] Standard errors for the parameters, the impulse responses and the variance decompositions are calculated

using the Monte Carlo approach of Runkle (1987), which simulates the VAR model to generate distributions for these results.[13]

The identification technique used in this paper is adequate for a model in which the number of parameters is equal to the number of unique elements in Σ_e. Alternative methods are needed to estimate a model with fewer parameters. Bernanke (1986) uses the method-of-moments approach of Hansen (1982) to estimate the parameters in equation 7 and obtain standard errors. Sims (1986) estimates the system of simultaneous equations for the residuals in equation 5 using maximum likelihood.[14] Blanchard and Watson (1986) also estimate the system of equations for residuals; however, they employ a sequential instrumental variables technique in which estimated structural shocks are used as instruments in all subsequent equations.[15]

The following four-equation contemporaneous structural VAR model is used to illustrate a particular set of such identifying restrictions. The residuals from a VAR consisting of the price level (p), output (y), the interest rate (r) and money (m) are used in the model. This model is used in the empirical work which follows. Equation 8 provides three restrictions by assuming that the price level is predetermined, except that producers can respond immediately to aggregate supply shocks. Equation 9 is a reduced-form IS equation that models output as a function of all the variables in the model. This approach was taken instead of explicitly modeling expected future inflation to calculate the real interest rate and explicitly modeling the term structure of interest rates to tie the short-term rate in the model with the long-term rate predicted by theory. The IS disturbance is also a factor in the output equation. The money supply function in equation 10 allows the Fed to adjust short-term interest rates to changes in the money stock. Two restrictions are obtained from assuming that the Fed does not immediately observe aggregate measures of output and price. The last equation is a short-run money demand function specifying nominal money holdings as

---

[12]Their model is identified by long-run restrictions.

[13]The actual residuals are randomly sampled, and the sampled residuals are used as shocks to the estimated VAR. After the artificial series are generated, they are used to perform the same structural VAR analysis. After 200 replications of the model, standard errors were calculated for the parameter estimates, the impulse responses and the variance decompositions.

[14]In contrast to the typical simultaneous equations model, this approach has no observable exogenous variables.

[15]This technique requires a structural model for which there are no estimated parameters in the first equation, the second equation has one parameter, the third has two parameters, etc. While the recursive model fits this description, this technique can estimate a much broader set of models.

a function of nominal GNP and the interest rate. This specification is motivated by a buffer stock theory where short-run money holdings rise in proportion to nominal income, yielding the final restriction for a just-identified model. Each equation includes a structural disturbance.

(8) $e_t^p = \varepsilon_t^{as}$

(9) $e_t^y = A_1 e_t^p + A_2 e_t^r + A_3 e_t^m + \varepsilon_t^{is}$

(10) $e_t^r = A_4 e_t^m + \varepsilon_t^{ms}$

(11) $e_t^m = A_5(e_t^y + e_t^p) + A_6 e_t^r + \varepsilon_t^{md}$

Standard VAR tools are employed after the structural parameters are estimated. Impulse response functions and variance decomposition functions conveniently summarize the dynamic response of the variables to the shocks, which is known as the moving average representation (MAR). The MAR for the VAR is obtained by applying simple algebra to a function of the lag operator. Take the VAR model for x:

$$x_t = \beta(L)x_{t-1} + e_t,$$

and subtract $\beta(L)x_{t-1}$ from both sides of this equation:

$$x_t - \beta(L)x_{t-1} = e_t.$$

Then factor terms in x_t using the lag operator,

$$[I - \beta(L)L]x_t = e_t.$$

Multiply both sides of this equation by the inverse of $[I-\beta(L)L]$:

$$x_t = [I - \beta(L)L]^{-1}e_t.$$

Insert the expression from equation 5 for e_t into this last equation:

(12) $x_t = [I - \beta(L)L]^{-1}A^{-1}D\varepsilon_t = \theta(L)\varepsilon_t,$

where $\theta(L) = \sum_{i=0}^{\infty}\theta_i L^i$

and each θ_i is an n x n matrix of parameters from the structural model. Equation 12 implies that the response of x_{t+i} to ε_t is θ_i. Hence, the sequence of θ_i from $i=0,1,2,...$, illustrates the

dynamic response of the variables to the shocks. If the variables in x are stationary, then the impulse responses must approach zero as i becomes large.

Variance decompositions allocate each variable's forecast error variance to the individual shocks. These statistics measure the quantitative effect that the shocks have on the variables. If $E_{t-j}x_t$ is the expected value of x_t based on all information available at time $t-j$, the forecast error is:

$$x_t - E_{t-j}x_t = \sum_{i=0}^{j-1} \theta_i \varepsilon_{t-i},$$

since the information at time $t-j$ includes all ε occurring at or before time $t-j$ and the conditional expectation of future ε is zero because the shocks are serially uncorrelated. The forecast error variances for the individual series are the diagonal elements in the following matrix:

$$E(x_t - E_{t-j}x_t)(x_t - E_{t-j}x_t)' = \sum_{i=0}^{j-1}\theta_i \Sigma_\varepsilon \theta_i'.$$

If θ_{ivs} is the (v,s) element in θ_i and σ_s is the standard deviation for disturbance s (s=1,...,n), the j-steps-ahead forecast variance of the v-th variable is easy to calculate:

$$E(x_{vt} - E_{t-j}x_{vt})^2 = \sum_{i=0}^{j-1}\sum_{s=1}^{n}\theta_{ivs}^2 \sigma_s^2 \quad v = 1, 2,..., n$$

The variance decomposition function (VDF) writes the j-steps-ahead percentage of forecast error variance for variable v atributable to the k-th shock:

(13) $VDF(v,k,j) = \dfrac{\sum_{i=0}^{j-1}\theta_{ivk}^2\sigma_k^2}{\sum_{i=0}^{j-1}\sum_{s=1}^{n}\theta_{ivs}^2\sigma_s^2} \times 100.$

The same analysis can be used to derive the MAR for the VAR model in equation 6.

(14) $\Delta x_t = \theta(L)\varepsilon_t,$

where $\theta(L) = [I-\beta(L)L]^{-1}A^{-1}D$. The response of x, rather than the change in x, is frequently of greater interest to economists. These impulse responses can be generated recursively by assuming that all the elements of ε at time zero and earlier are equal to zero.[16] For example,

[16]If the pre-sample ε is nonzero, its effects are lumped together with x_0 which represents the initial conditions.

$$x_1 = x_0 + \theta_0 \varepsilon_1$$

and

$$x_2 = x_1 + \theta_0 \varepsilon_2 + \theta_1 \varepsilon_1.$$

Inserting the expression for x_1 into the x_2 equation yields:

$$x_2 = x_0 + \theta_0 \varepsilon_2 + (\theta_0 + \theta_1) \varepsilon_1.$$

Repeating this operation for all x up to x_t, yields the following:

$$x_t = x_0 + \theta_0 \varepsilon_t + (\theta_0 + \theta_1) \varepsilon_{t-1} + \ldots + (\sum_{j=0}^{t-1} \theta_j) \varepsilon_1.$$

This result is equivalently written as

$$(15) \quad x_t = x_0 + \Gamma(L)\varepsilon_t = x_0 + \sum_{i=0}^{t-1} \Gamma_i \varepsilon_{t-i},$$

where $\Gamma_i = \sum_{j=0}^{i} \theta_j$.

The response of x_{t+i} to ε_t is Γ_i. Since the differenced specification assumes that Δx is stationary, the θ_j matrix goes to zero as j gets large. This implies that Γ_i converges to the sum of coefficients in $\theta(L)$. Restrictions on this sum of coefficients are used to identify long-run structural VAR models. The variance decompositions for this model are identical to equation 13 except that θ is replaced by Γ.

In contrast to the atheoretical VAR models developed by Sims (1980), the structural approach yields impulse responses and variance decompositions that are derived using parameters from an explicit economic model.[17] Finding dynamic patterns consistent with the structural model used for identification would provide evidence in support of the theoretical model. Otherwise, the theory is invalid or the empirical model is somehow misspecified.

LONG-RUN STRUCTURAL VAR MODELS

Shapiro and Watson (1988) and Blanchard and Quah (1989) developed the alternative approach of imposing identifying restrictions on long-run multipliers for structural shocks. An advantage is that these models do not impose contemporaneous restrictions, but they allow the data to determine short-run dynamics based conditionally on a particular long-run model.[18]

If each shock has a permanent effect on at least one of the variables and if cointegration does not exist for the variables in x, the VAR in equation 6 can be estimated.[19] The impulse response function for x in equation 15 shows that the long-run effect of ε converges to the sum of coefficients in $\theta(L)$. It is obvious from the definition of $\theta(L)$ that replacing L by one yields the sum of coefficients. Hence, this sum is conveniently written as $\theta(1)$, and this matrix is used to parameterize long-run restrictions. The relationship between parameters of the structural MAR, contemporaneous structural parameters and VAR lag coefficients is given by

$$(16) \quad \theta(L) = [I - \beta(L)L]^{-1} A^{-1} D.$$

The long-run multipliers are obtained by replacing L in equation 16 with unity.

Setting L equal to unity, solving equation 16 for $A^{-1}D$ and inserting the result into equation 7 yields

$$(17) \quad [I - \beta(1)]^{-1} \Sigma_e [I - \beta(1)]^{-1'} = \theta(1) \Sigma_\varepsilon \theta(1)',$$

where the matrix $\beta(1)$ is the sum of VAR coefficients.

This equation can be used to identify the parameters in $\theta(1)$ and Σ_ε. A minimal set of restrictions on the long-run response of macroeconomic variables to structural disturbances is used to identify long-run structural VAR models. Estimates of the matrices on the left side of

---

[17]The relationship between structural and atheoretical VARs is addressed in the shaded insert at right.

[18]For example, agents may temporarily be away from long-run equilibrium positions or monetary policy may be non-neutral in the short run.

[19]Unit-root tests and cointegration tests support this assumption for the time series used in this paper. See Keating (1992) for this evidence.

The Relationship between Atheoretical and Structural VAR Approaches

Atheoretical VAR practitioners separate the residuals into orthogonal shocks by calculating a Choleski decomposition of the covariance matrix for the residuals. This decomposition is obtained by finding the unique lower triangular matrix λ that solves the following equation:

$$\Sigma_e = \lambda\lambda'.$$

This statistical decomposition depends on the sequence in which variables are ordered in x. The residuals' covariance matrix from a VAR ordered by output, the interest rate, money and the price level yields a Choleski decomposition that is algebraically equivalent to estimating the following four equations by ordinary least squares:

$$e_t^y = \nu_t^y$$
$$e_t^r = R_1 e_t^y + \nu_t^r$$
$$e_t^m = R_2 e_t^y + R_3 e_t^r + \nu_t^m$$
$$e_t^p = R_4 e_t^y + R_5 e_t^r + R_6 e_t^m + \nu_t^p.$$

Hence, each ν shock is uncorrelated with the other shocks by construction. This system implies that the first variable responds to its own exogenous shock, the second variable responds to the first variable plus an exogenous shock to the second variable, and so on. In practice, atheoretical VAR studies report results from various orderings. The total number of possible orderings of the system is n!, a number that increases rapidly with n.[1] Investigators sometimes note that certain properties of the model are insensitive to alternative orderings. Results sensitive to VAR orderings are difficult to interpret, especially if a recursive economic structure is implausible.

This atheoretical approach has been criticized by Cooley and LeRoy (1985). First, if the Choleski technique is in fact atheoretical, then the estimated shocks are not structural and will generally be linear combinations of the structural disturbances.[2] In this case, standard VAR analysis is difficult to interpret because the impulse responses and variance decompositions for the Choleski shocks will be complicated functions of the dynamic effects of all the structural disturbances. The second point attacks the claim that Choleski decompositions are atheoretical. The Choleski ordering can be interpreted as a recursive contemporaneous structural model. Unfortunately, most economic theories do not imply recursive contemporaneous systems. Such criticisms of the atheoretical approach inspired structural approaches to VAR modeling. If therory predicts a contemporaneous recursive economic structure, a particular Choleski factorization of the covariance matrix for the residuals is appropriate. But a researcher using the structural approach would not experiement with various orderings, unless these specifications were predicted by alternative theories.

---

[1] For example, 3! = 6 but 6! = 720.

[2] This result is easy to prove. The Choleski decomposition yields a system in which $e_t = R\nu_t$ but the true structural model is $e_t = A^{-1}D\varepsilon_t$, implying that the shocks from the Choleski decomposition are linear combinations of the structural distrubances; $\nu_t = R^{-1}A^{-1}D\varepsilon_t$.

equation 17 are obtained directly from the unconstrained VAR.[20] $\theta(1)$ has n^2 elements and Σ_ε has $n(n+1)/2$ unique elements. The $n(n+1)/2$ unique elements in the symmetric matrix on the left side of equation 17 is the number of parameters in a just-identified model.[21] Thus, at least n^2 identifying restrictions must be applied to $\theta(1)$ and Σ_ε. The elements of the main diagonal for $\theta(1)$ can each be set equal to one, analogously to the normalization used in the contemporaneous model. If each element of ε is assumed to be independent, then Σ_ε is diagonal. Hence, $n(n-1)/2$ additional restrictions are needed for $\theta(1)$ to identify the model.

Several alternative approaches for obtaining the structural parameters have been developed. Shapiro and Watson (1988) impose the long-run zero restrictions on $\theta(1)$ by estimating the simultaneous equations model with particular explanatory variables differenced one additional time. King, Plosser, Stock and Watson (1991) impose long-run restrictions using the vector error-correction model with some of the long-run features of the model chosen by cointegration regressions. Gali (1992) combines contemporaneous restrictions with long-run restrictions to identify a structural model. In the empirical section, we use the approach developed by Blanchard and Quah (1989).

Equations 18 through 21 present the long-run identifying restrictions used in the empirical example.[22] The time subscripts are omitted because the restrictions pertain to long-run behavior. Three restrictions come from equation 18, which specifies that aggregate supply shocks are the sole source of permanent movements in output.[23] Two more restrictions are obtained from the long-run IS or spending balance equation, 19, which specifies the interest rate as a function of output and the IS shock.[24] Note the coefficient S_1 should be negative. The final restriction comes from the money demand

function, 20, which sets real money equal to an increasing function of output, a decreasing function of the interest rate, and a money demand shock. Equation 21 allows the supply of money to respond to all variables in the model and a money supply shock.[25]

$$(18)\ y = \varepsilon^{as}$$

$$(19)\ r = S_1 y + \varepsilon^{is}$$

$$(20)\ m - p = S_2 y + S_3 r + \varepsilon^{md}$$

$$(21)\ m = S_4 y + S_5 r + S_6(m-p) + \varepsilon^{ms}$$

EMPIRICAL EXAMPLES AND RESULTS

The examples from the previous two sections are estimated to illustrate the long-run and contemporaneous identification methods. Both models use real GNP to measure output, the GNP deflator for the price level, M1 as a measure of the stock of money, and the three month Treasury bill rate determined in the secondary market as the interest rate. The data are first-differenced. Statistical tests suggest that this transformation makes the data stationary.[26] The first step is to estimate the reduced-form VAR model. The estimated variance-covariance matrix from the reduced form is used to obtain the second-stage structural estimates. Four lags are used for the VAR model and the sample-period is from the first quarter of 1959 to the third quarter of 1991.

A Long-Run Structural Model

Table 1 reports the parameter estimates for the long-run model in equations 18 through 21. The first four parameters are standard deviations for the structural shocks, and each of these

---

[20]The Bayesian approach is not employed since a unit root is required to be certain that shocks have permanent effects.

[21]An over-identified long-run model will imply restrictions on the reduced-form coefficients.

[22]This is a simplified version of the model in Keating (1992).

[23]If interest rates affect capital accumulation, then IS shocks may permanently affect output and the restrictions in equation 19 may not be appropriate. If this is the only misspecification of the model, actual money supply and money demand shocks will be identified but the empirical aggregate supply and IS shocks will be mixtures of these real structural disturbances.

[24]Technically, the IS equation should use the real interest rate, an unobservable variable. However, if permanent movements in the nominal rate and the real rate are identical, this specification is legitimate. This would be true, for example, if the Fisher equation held and if inflation followed a stationary time series process.

[25]Thus, $\theta(1)$ is:
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -S_1 & 1 & 0 & 0 \\ -S_2 & -S_3 & 1 & 0 \\ -S_4 & -S_5 & -S_6 & 1 \end{bmatrix}^{-1}$$

[26]For empirical evidence, see the unit-root tests in Keating (1992).

Table 1

Estimates for the Long-Run Model

	Parameter	Standard error
σ_{as}	.0144*	.0041
σ_{is}	.0092*	.0028
σ_{md}	.0149*	.0031
σ_{ms}	.0172*	.0042
S_1	−.1171	.3068
S_2	.8722*	.4219
S_3	−2.276 *	.7083
S_4	−1.411 *	.5778
S_5	1.569	1.020
S_6	.9048*	.3842

NOTE: An asterisk (*) indicates significance at the 5 percent level.

estimates is significantly different from zero. The coefficient in the IS equation, S_1, is negative as hypothesized, but insignificantly different from zero. Each parameter in the money demand function is statistically significant and has the sign predicted by economic theory. The coefficient on real GNP, S_2, is not statistically different from one. Parameters for the money supply equation can be interpreted as a policy reaction function in which the Fed reduces money if output rises but increases money if interest rates rise. This last effect is not statistically significant. The increase in money in response to an increase in real money may reflect the fact that the Fed has typically smoothed interest rate fluctuations in the post-war period, so that a money demand shock that raises real money will be accommodated by a comparable increase in nominal money.

The impulse responses for the long-run model are shown in figures 1 through 4. Point estimates and 90 percent confidence intervals are graphed for the variables. If the long-run parameter estimates are consistent with the theoretical model, the asymptotic properties of the impulse responses must also be consistent with the theory. Economic restrictions are not imposed on the dynamic properties of the model. The empirical aggregate supply shock raises output and real money and lowers the interest rate, the price level and nominal money. The real spending shock raises output only temporarily because of the restriction that aggregate supply shocks are the only factor in long-run output movements. The interest rate and the price level rise, while the nominal and real measures of money decline after each variable initially rises by a small amount. The money demand

shock has a strong positive effect on nominal and real money. The other effects are relatively weak, with prices falling, output temporarily falling and the interest rate temporarily rising. Money and the price level both rise in response to an increase in the money supply, which also causes a temporary decline in the interest rate and a temporary increase in output and real money. The impulse response functions provide evidence that the shocks affect each variable as theory predicts.

The variance decompositions in table 2 show the average amount of the variance of each variable attributable to each shock. Standard errors for these estimates are in parentheses. The output variance due to the supply shock one quarter in the future is only 17 percent. Eight quarters in the future, however, the estimate becomes nearly half of output's variance and, at 48 quarters, 90 percent of the variance of output is attributed to supply shocks. Variability in the price level is dominated by aggregate supply shocks, particularly in the short run. The other shocks have temporary output effects. This long-run feature is obtained because the model forces aggregate supply shocks to explain all permanent output movements. Short-run output movements are dominated by real spending shocks. This shock explains most of the interest rate variance and the variance of real money in the long run. The money supply shock has a gradual effect on output that peaks at 13 percent of the variance two years in the future. This shock accounts for a large portion of the short-run variance of the interest rate and the long-run movement in nominal money. The money demand shock has virtually no effect on output, interest rates or prices. Many theories predict that money demand shocks will not have much effect on these variables if the Fed uses the interest rate as its operating target. Money demand shocks have strong effects on nominal money and real money. In general, the results for this model are consistent with most economists' views about economic behavior, although some might be surprised by the relatively small effect on output of money supply disturbances.

Contemporaneous Structural Model

The parameter estimates for the contemporaneous model in equations 8 through 11 are reported in table 3. The coefficients in the reduced-form IS equation are all negative. The negative

Figure 1
Responses to an Aggregate Supply Shock in the Long-Run Model

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
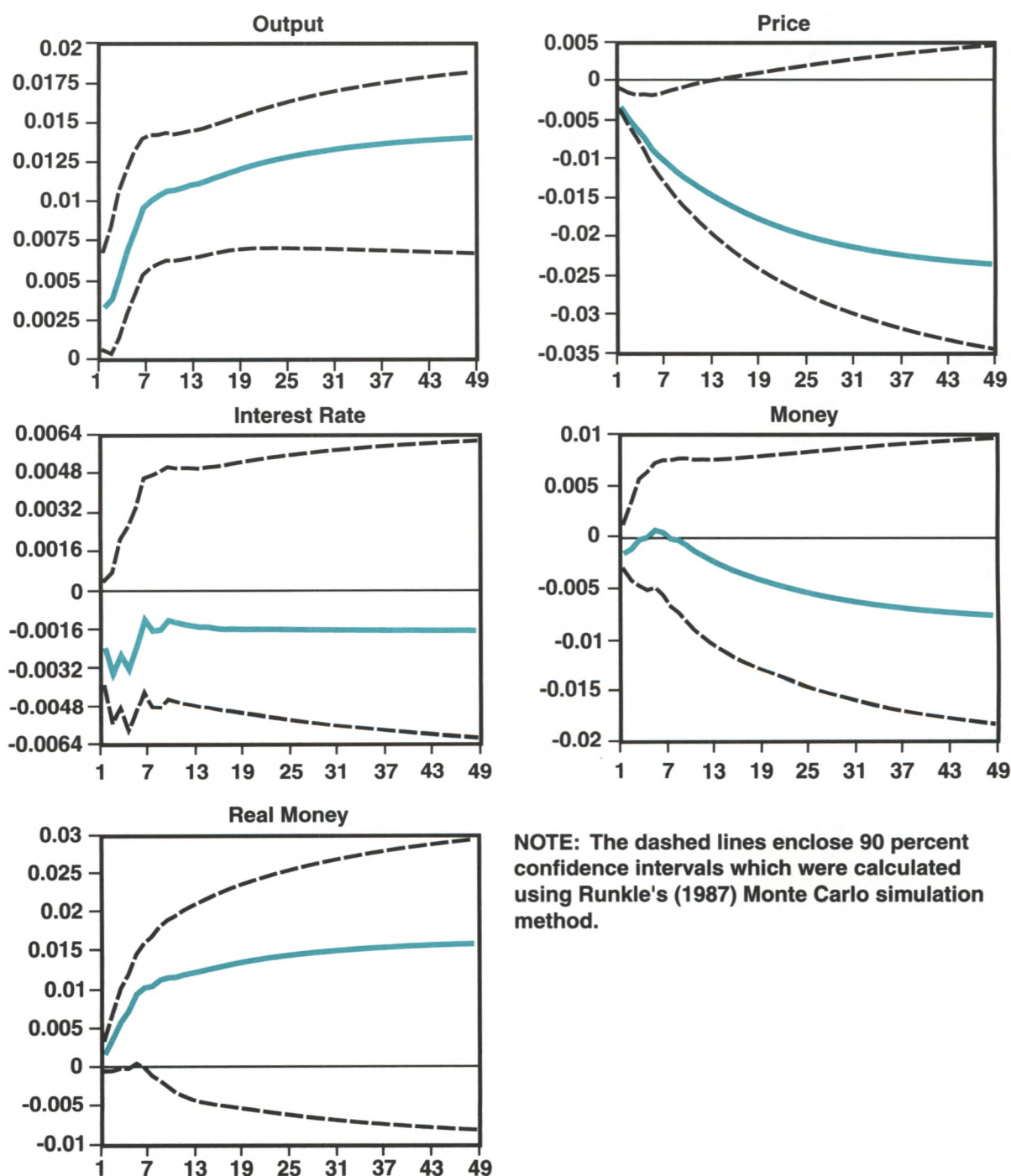
Figure 2
Responses to a Real Spending Shock in the Long-Run Model

Output

Price

Interest Rate

Money

Real Money

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
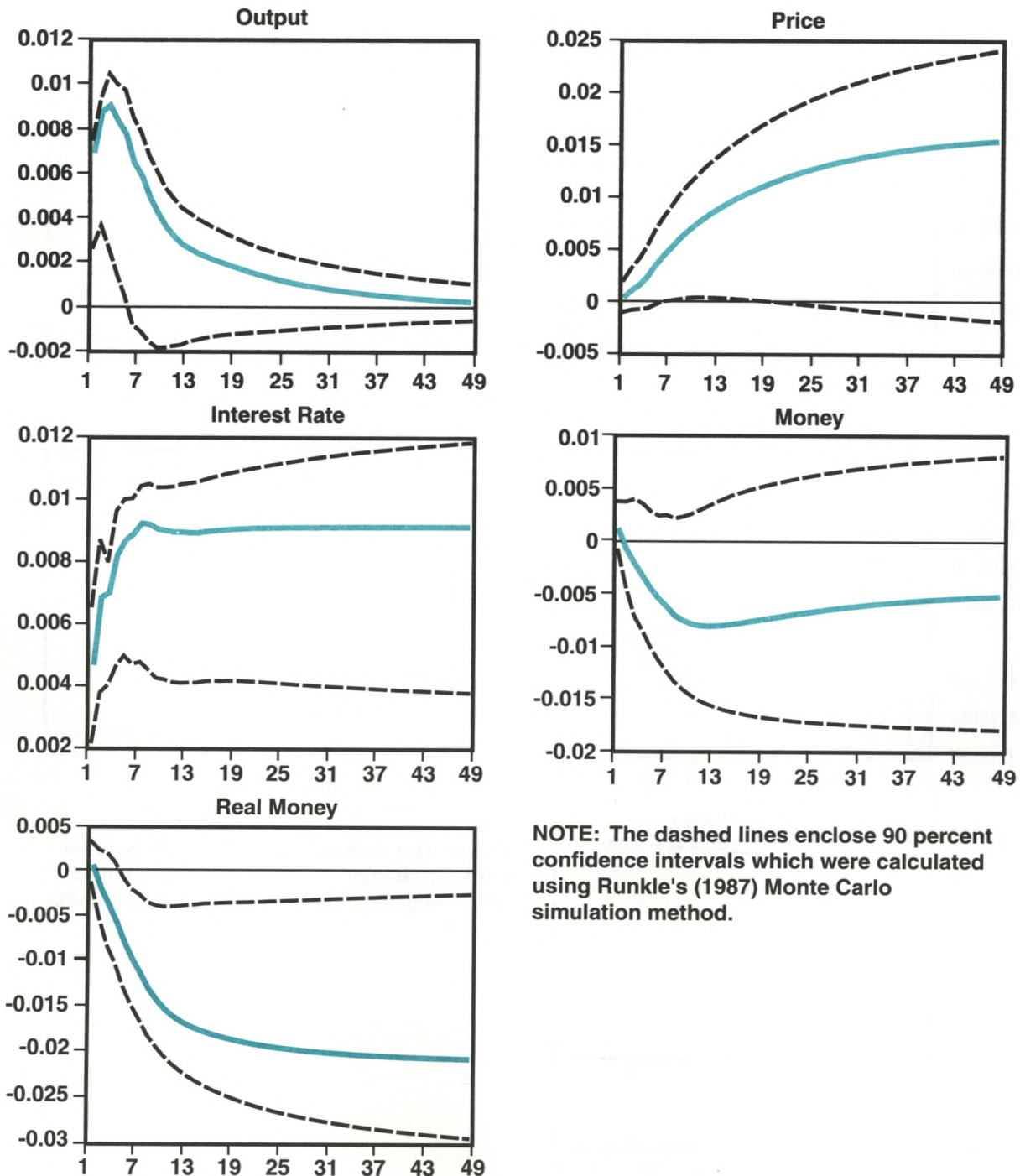
Figure 3
Responses to a Money Demand Shock in the Long-Run Model

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
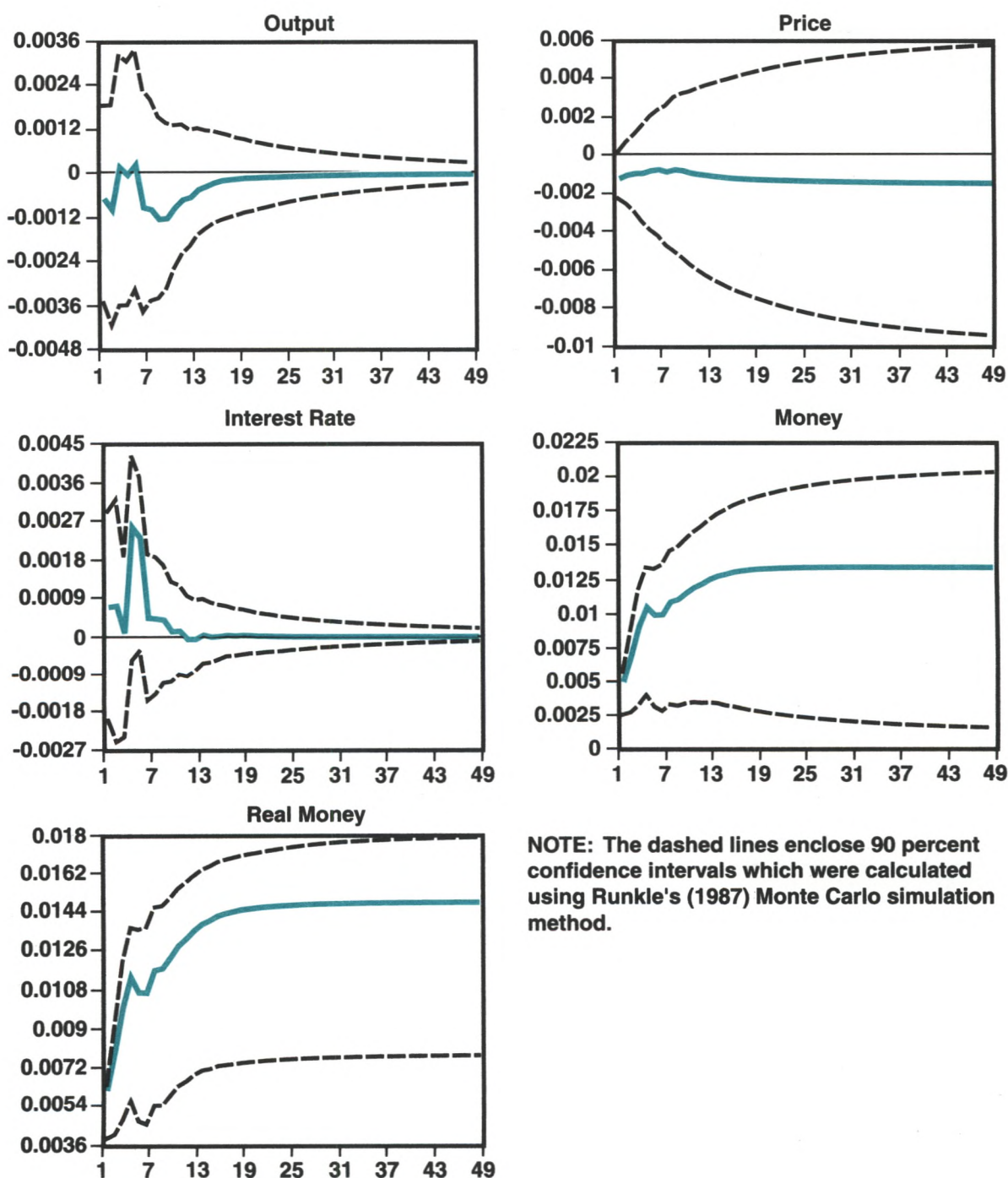
Figure 4
Responses to a Money Supply Shock in the Long-Run Model

Output

Price

Interest Rate

Money

Real Money

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
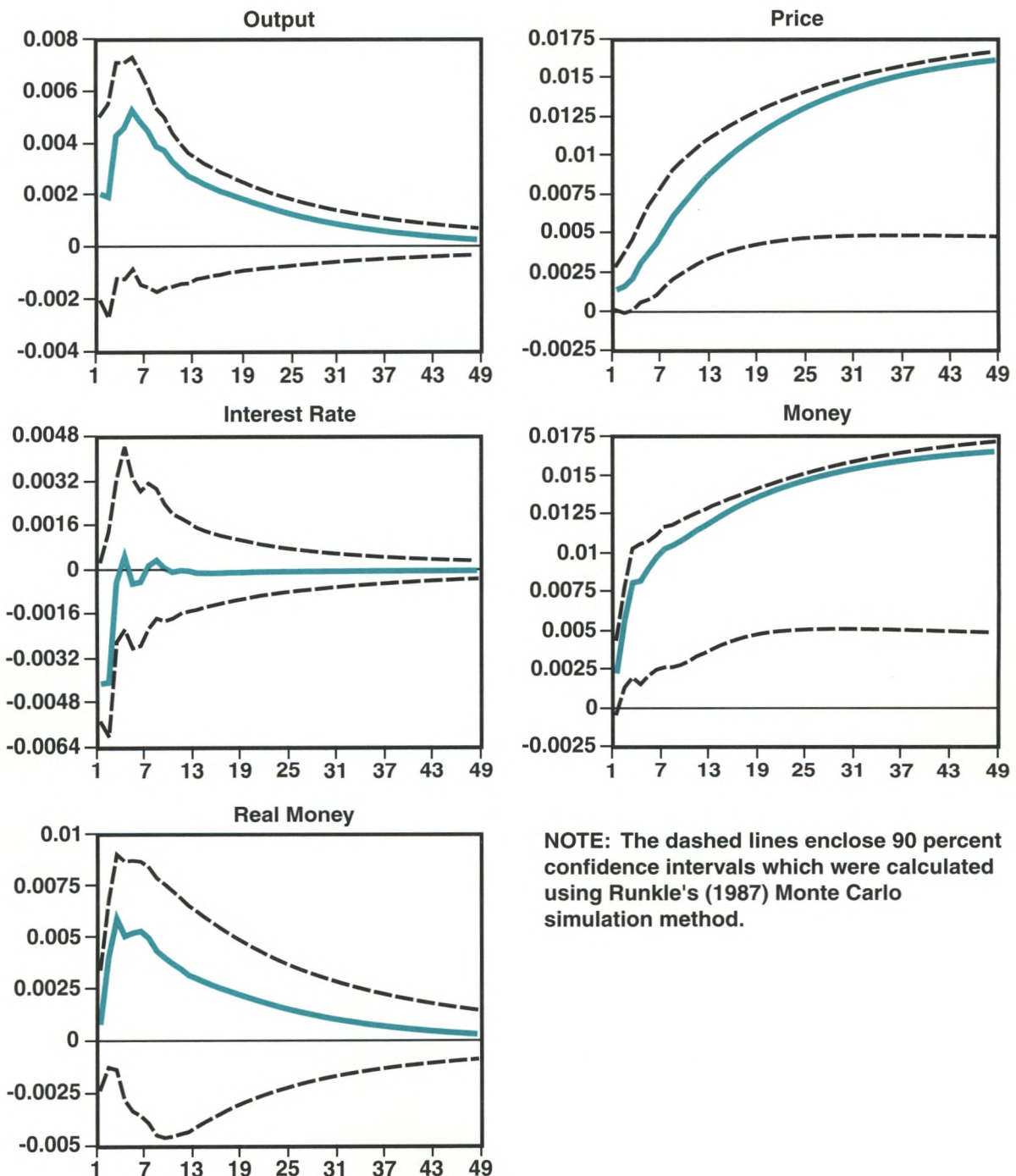
Table 2

Variance Decompositions for the Long-Run Model

Variable	Quarter(s) ahead	Aggregate supply shock	Real spending shock	Money demand shock	Money supply shock
Output					
	1	17 (24)	76 (24)	1 (8)	7 (16)
	2	16 (23)	78 (24)	1 (8)	5 (14)
	4	24 (25)	65 (24)	0 (6)	11 (14)
	8	45 (23)	42 (19)	0 (4)	13 (11)
	16	67 (15)	24 (13)	0 (2)	9 (7)
	32	84 (8)	11 (7)	0 (1)	5 (3)
	48	90 (5)	7 (5)	0 (1)	3 (2)
Interest Rate					
	1	13 (13)	49 (25)	1 (8)	37 (23)
	2	14 (13)	57 (23)	1 (7)	27 (19)
	4	13 (11)	71 (15)	3 (5)	13 (9)
	8	8 (10)	84 (12)	2 (3)	6 (6)
	16	5 (11)	91 (12)	1 (2)	3 (4)
	32	4 (13)	94 (13)	1 (1)	1 (2)
	48	4 (14)	95 (14)	0 (1)	1 (1)
Real Money					
	1	7 (11)	2 (10)	90 (17)	2 (11)
	2	12 (14)	3 (8)	73 (18)	12 (14)
	4	18 (17)	8 (12)	59 (19)	14 (14)
	8	26 (19)	24 (17)	41 (18)	9 (10)
	16	25 (19)	38 (20)	33 (18)	4 (5)
	32	26 (21)	44 (21)	29 (19)	2 (3)
	48	27 (22)	45 (21)	28 (19)	1 (1)
Money					
	1	6 (10)	5 (14)	73 (22)	15 (20)
	2	3 (9)	2 (10)	63 (22)	32 (21)
	4	1 (9)	3 (11)	59 (23)	38 (22)
	8	0 (10)	10 (15)	50 (21)	40 (21)
	16	1 (11)	14 (17)	46 (20)	39 (20)
	32	4 (13)	11 (17)	41 (19)	44 (20)
	48	6 (15)	9 (17)	39 (19)	46 (20)
Price					
	1	73 (27)	2 (12)	9 (13)	16 (23)
	2	78 (27)	4 (14)	5 (10)	12 (21)
	4	77 (28)	7 (16)	2 (8)	13 (21)
	8	69 (28)	15 (19)	1 (7)	16 (20)
	16	60 (27)	20 (21)	0 (7)	20 (19)
	32	55 (27)	22 (21)	0 (7)	23 (19)
	48	54 (27)	22 (22)	0 (7)	24 (19)

NOTE: Standard errors are in parentheses.

Table 3
Estimates for the Contemporaneous Model

	Parameter	Standard error
σ_{as}	.0038*	.0003
σ_{is}	.0086*	.0027
σ_{ms}	.0087	.0419
σ_{md}	.0087	.0324
A_1	−.1164	.3255
A_2	−.0469	.4122
A_3	−.3327	.6539
A_4	1.030	8.302
A_5	.5632	2.098
A_6	−.9397	5.737

NOTE: An asterisk (*) indicates significance at the 5 percent level.

coefficient on money would be unexpected in a structural IS equation; however, these estimates are reduced-form coefficients, not structural parameters. The coefficient on money in the interest rate equation is positive. This supports the view that the central bank attempts to stabilize money growth by raising interest rates. In the money demand equation, the coefficient on nominal spending is roughly one-half, and the interest rate coefficient is almost − 1.0. Hence, the parameter estimates in this structural model are consistent with economic theory. Unfortunately, each of these structural parameters is statistically insignificant.

Figures 5 through 8 plot the impulse responses. In contrast to the long-run model, the aggregate supply equation is normalized on the price level. Hence, an aggregate supply shock raises the price level and reduces output. The aggregate supply shock has this expected effect on prices and output. Real money also decreases. The adverse supply shock has a weak positive effect on money and the interest rate. The IS shock raises prices, output and the interest rate. Real and nominal money initially increase, although both subsequently fall. The money supply equation is normalized on the interest rate so a reduction in the supply of money raise interest rates. This shock raises the interest rate and causes a decline in nominal money, real money and the price level. Surprisingly, output rises briefly before it begins to decline. The money demand shock causes the interest rate, nominal and real money to increase while output falls. The rising price level is inconsistent

with theory, although this effect is not statistically significant. In contrast with the long-run model, there are a few unusual features in the impulse responses for the contemporaneous specification. Most of the dynamic patterns, however, are consistent with the structural model.

The variance decompositions for the contemporaneous model are shown in table 4. Many features of this table are comparable to the long-run model's results. For example, the aggregate supply shock gradually explains most of output's variability, is the most important shock for the price level and is never an important source of interest rate movements. The IS shock is the most important source of short-run output movement, and it explains most of the long-run variance of the interest rate. Some results, however, are considerably different compared with the results from the long-run model. The money demand shock has its greatest effect on output in the long run. This shock explains a large amount of the short-run variance of the interest rate but virtually none of the long-run variance of real or nominal money balances. The money supply shock has essentially no effect on output, while accounting for a large amount of the variance in real money, even in the long run, and none of the variance of prices. These results are inconsistent with most macroeconomic theories.

CONCLUDING REMARKS

This paper outlines the basic theory behind structural VAR models and estimates two models using a standard set of macroeconomic data. The results for the two specifications are often similar. The long-run structural VAR model in this paper generally provides empirical results that are consistent with the structural model. Some of the variance decompositions and the impulse responses for the contemporaneous model were inconsistent with standard macroeconomic theory. These inconsistencies pertain to the effects of money supply and money demand disturbances. Another result is that structural parameters in the long-run model are more precisely estimated than parameters in the contemporaneous model. Wherever a significant discrepancy exists between the two models, the model with long-run restrictions yields sensible results, while the results from the contemporaneous model are inconsistent with standard economic theories.

Figure 5
Responses to an Aggregate Supply Shock in the Contemporaneous Model

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
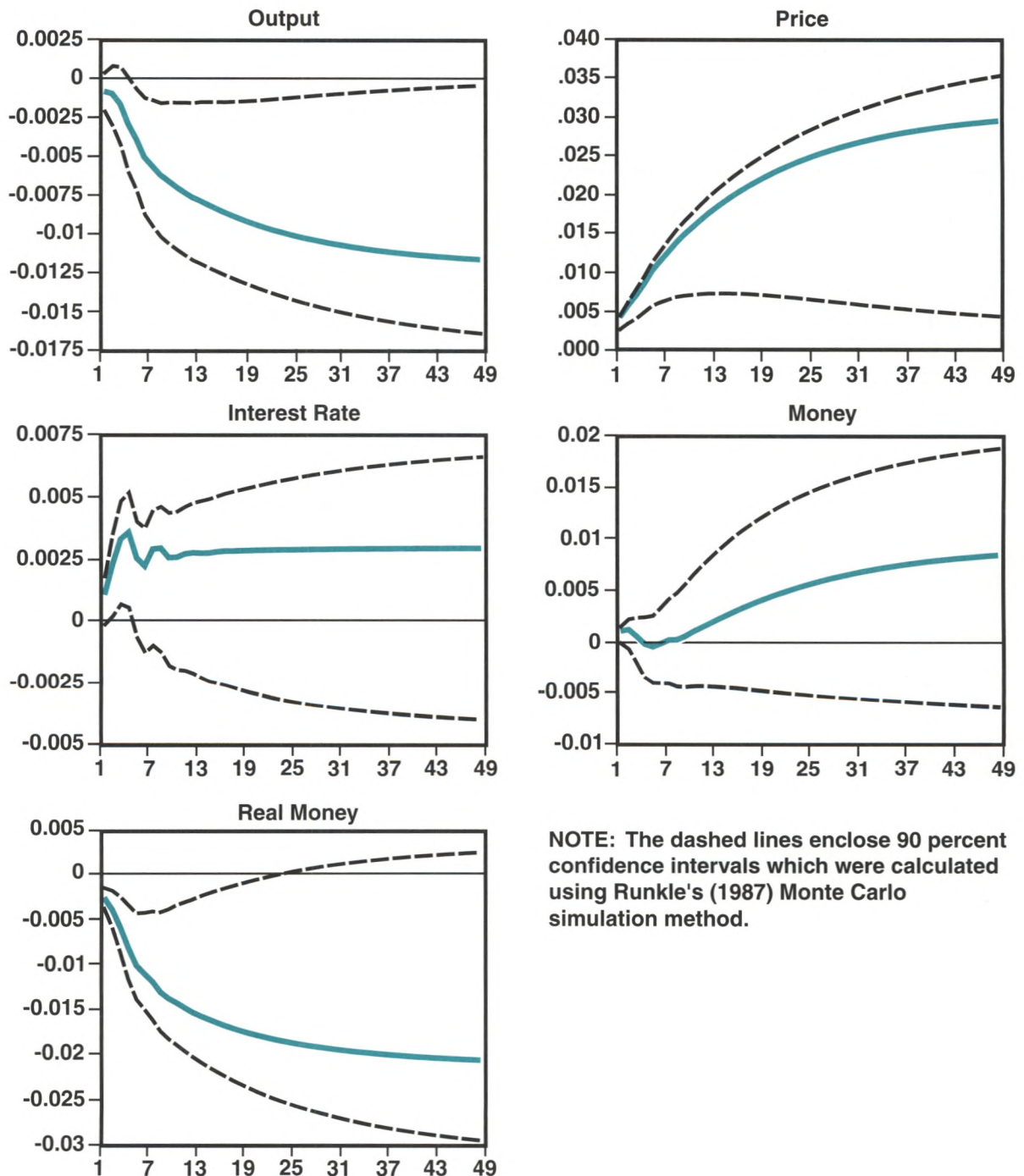
Figure 6
Responses to a Real Spending Shock in the Contemporaneous Model

Output

Price

Interest Rate

Money

Real Money

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
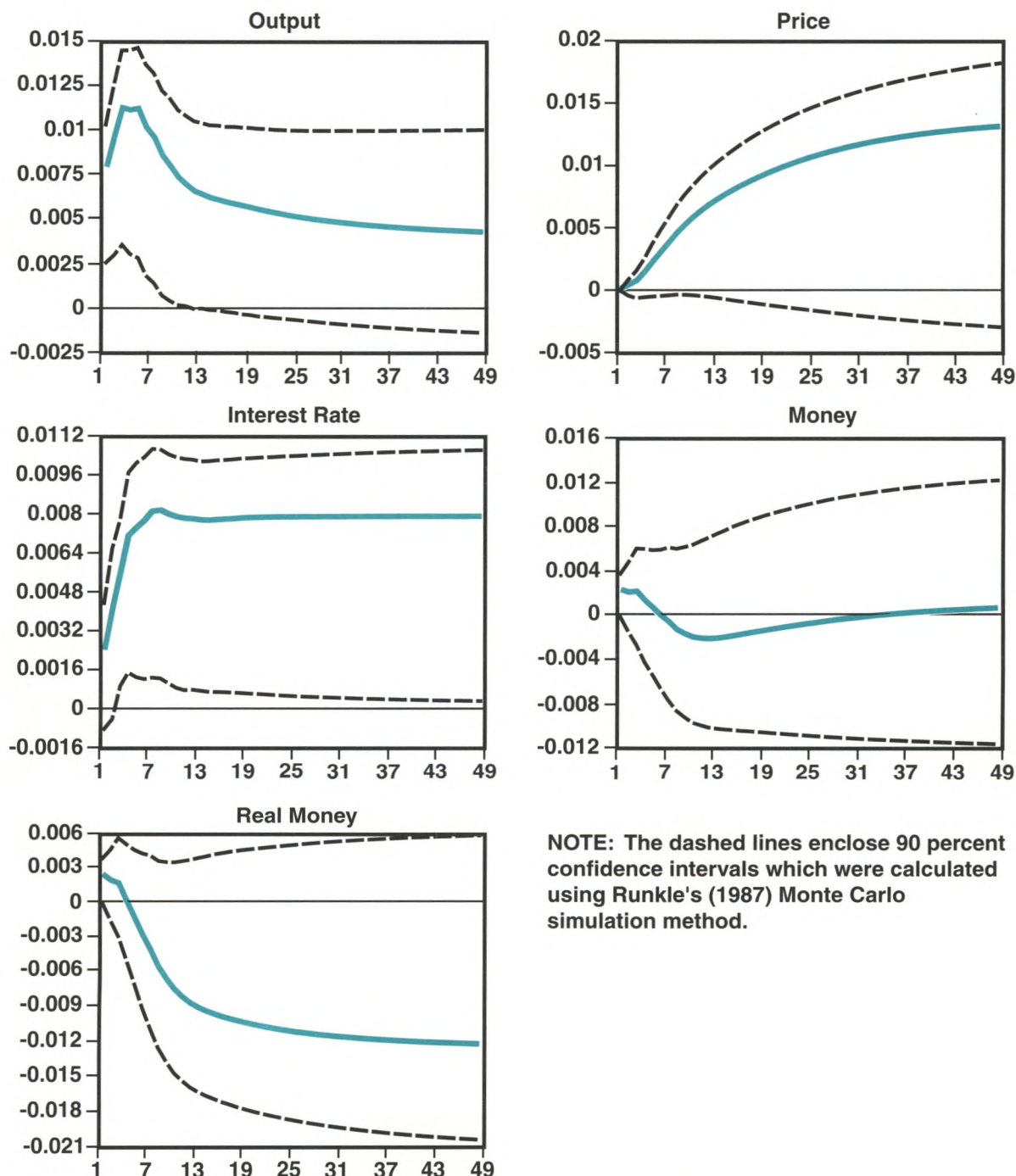
Figure 7
Responses to a Money Supply Shock in the Contemporaneous Model

Output

Price

Interest Rate

Money

Real Money

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.
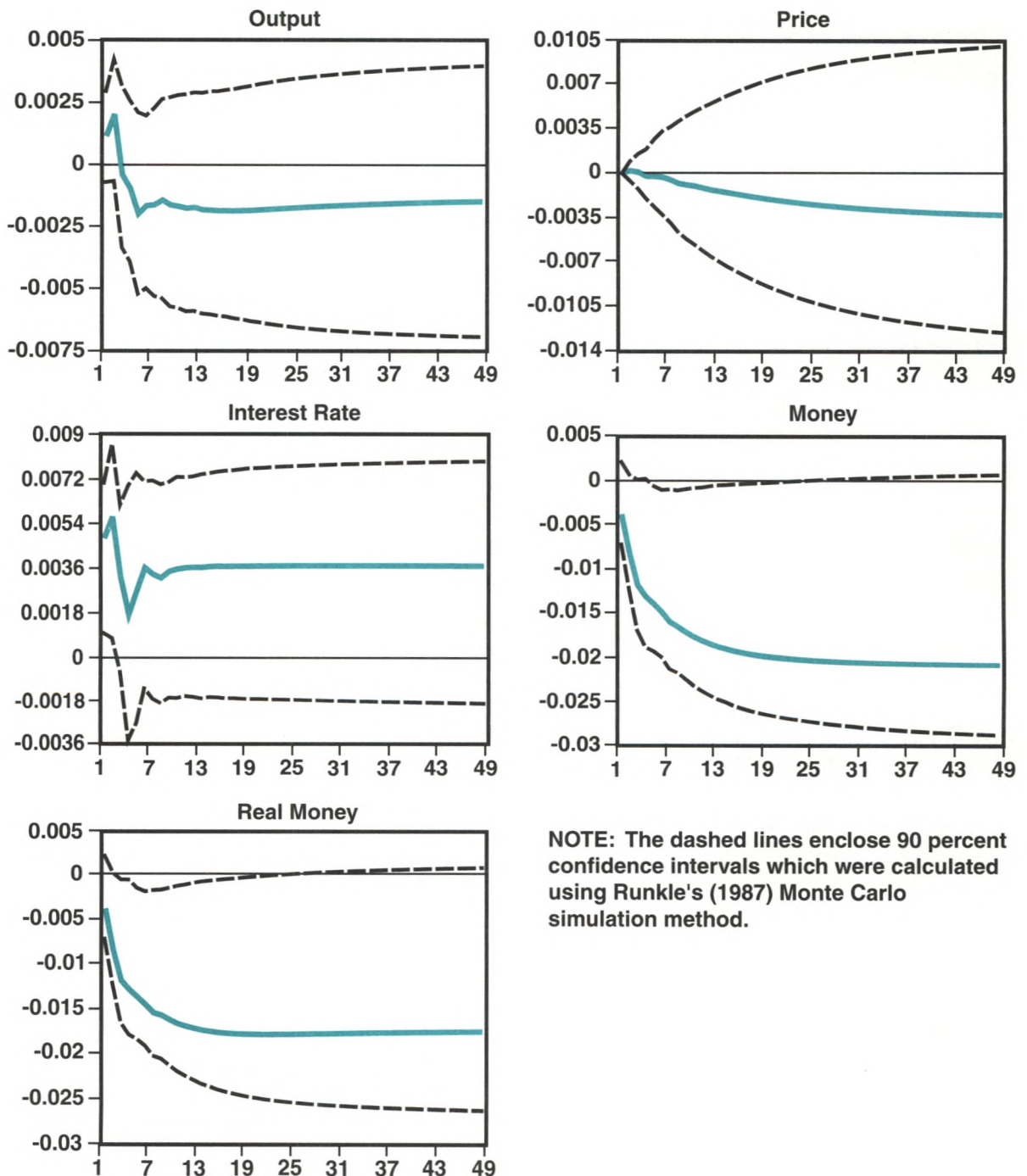
Figure 8
Responses to a Money Demand Shock in the Contemporaneous Model

Output

Price

Interest Rate

Money

Real Money

NOTE: The dashed lines enclose 90 percent confidence intervals which were calculated using Runkle's (1987) Monte Carlo simulation method.

Table 4

Variance Decompositions for the Contemporaneous Model

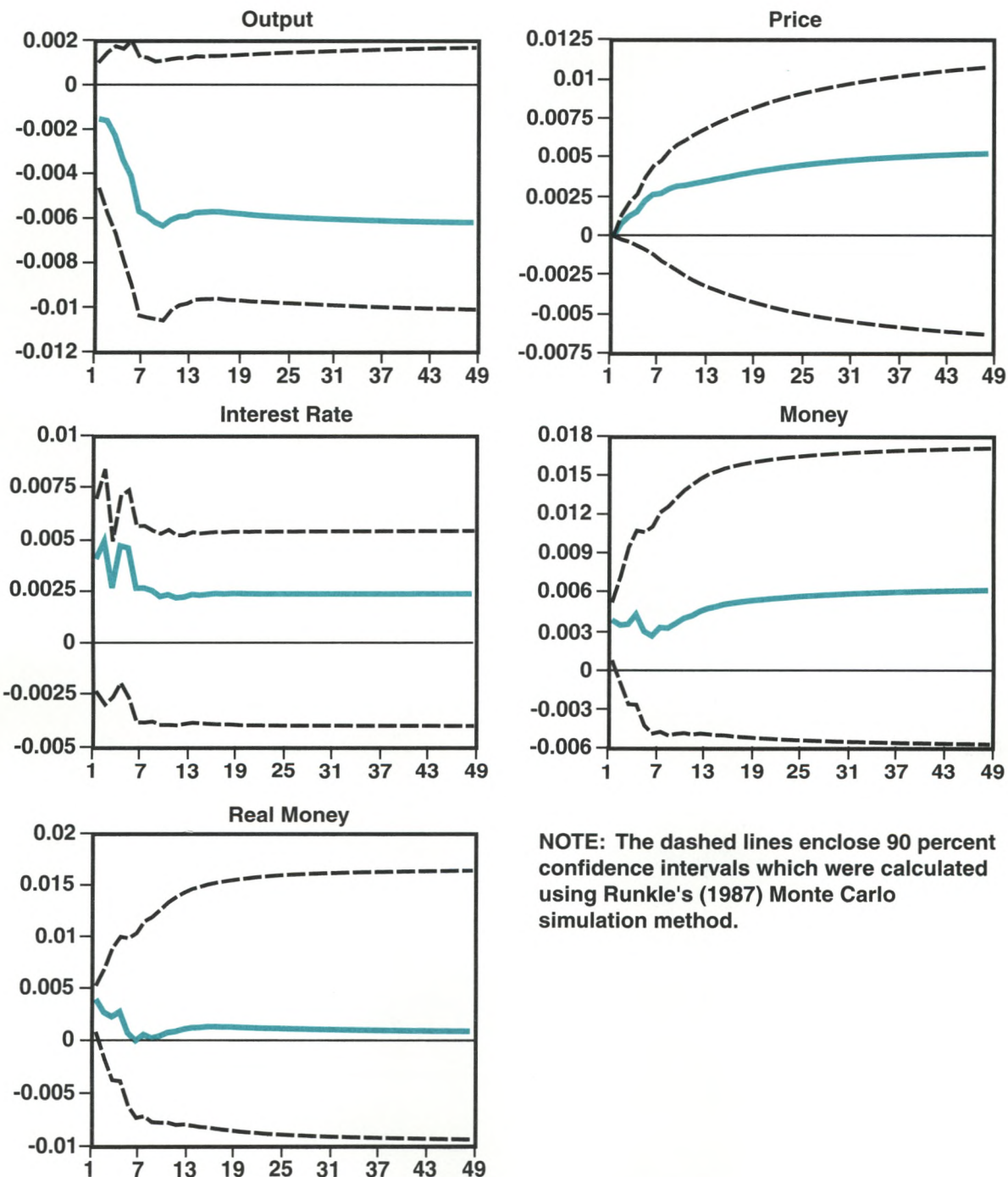| Variable	Quarter(s) ahead	Aggregate supply shock	Real spending shock	Money demand shock	Money supply shock
Output					
	1	1 (2)	94 (16)	4 (14)	2 (4)
	2	1 (3)	93 (16)	3 (13)	3 (5)
	4	3 (5)	91 (15)	5 (13)	1 (3)
	8	12 (10)	73 (19)	13 (14)	2 (5)
	16	28 (16)	51 (20)	19 (14)	2 (7)
	32	47 (21)	31 (21)	20 (14)	2 (9)
	48	55 (23)	23 (22)	20 (14)	2 (10)
Interest Rate					
	1	2 (2)	12 (12)	37 (26)	49 (31)
	2	4 (4)	17 (13)	34 (24)	44 (30)
	4	10 (7)	38 (15)	27 (18)	25 (23)
	8	8 (6)	56 (18)	19 (16)	17 (20)
	16	8 (7)	63 (20)	12 (15)	16 (20)
	32	8 (10)	67 (22)	9 (14)	16 (20)
	48	8 (11)	68 (22)	8 (14)	15 (20)
Real Money					
	1	19 (7)	11 (14)	36 (21)	34 (22)
	2	18 (8)	5 (11)	16 (18)	60 (20)
	4	24 (11)	2 (8)	6 (16)	68 (19)
	8	34 (13)	4 (7)	2 (14)	61 (19)
	16	38 (16)	10 (10)	1 (15)	51 (21)
	32	42 (20)	14 (12)	0 (15)	44 (23)
	48	45 (21)	15 (12)	0 (16)	40 (24)
Money					
	1	2 (2)	14 (15)	44 (24)	41 (31)
	2	1 (2)	7 (12)	23 (22)	68 (27)
	4	0 (2)	3 (10)	13 (21)	83 (26)
	8	0 (3)	1 (9)	7 (20)	92 (25)
	16	1 (6)	1 (11)	6 (20)	92 (25)
	32	4 (10)	1 (11)	7 (20)	89 (25)
	48	7 (13)	0 (11)	7 (20)	86 (25)
Price					
	1	100 (0)	0 (0)	0 (0)	0 (0)
	2	98 (2)	0 (1)	2 (1)	0 (1)
	4	95 (4)	2 (2)	3 (3)	0 (2)
	8	89 (7)	6 (6)	4 (5)	0 (5)
	16	84 (10)	12 (8)	4 (6)	0 (7)
	32	82 (11)	14 (9)	3 (7)	1 (9)
	48	81 (12)	15 (10)	3 (7)	1 (9)

NOTE: Standard errors are in parentheses.

These comparisons between contemporaneous and long-run specifications may not generalize to all structural VAR applications, but they suggest that long-run structural VARs may yield theoretically predicted results more frequently than VARs identified with short-run restrictions. This result is not surprising. One reason is that economic theories may often have similar long-run properties but different short-run features. For example, while output movements are driven solely by aggregate supply shocks in a typical real business cycle model, supply shocks will account for permanent output movements in Keynesian models, but every shock may have some cyclical effect. In addition, long-run structural VAR models may provide superior results because they typically do not impose contemporaneous exclusion restrictions. Keating (1990) shows that contemporaneous "zero" restrictions may be inappropriate in an environment with forward-looking agents who have rational expectations. The intuition behind this result is that any observable contemporaneous variable may provide information about future events. One implication from that paper is that different short-run restrictions can be obtained from alternative assumptions about available information. Further research should investigate other examples of contemporaneous and long-run structural VAR models to determine whether the superior performance of this paper's long-run model is a special case or a more general result.

REFERENCES

Bernanke, Ben S. "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy* (Autumn 1986), pp. 49-100.

Blanchard, Olivier Jean. "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review* (December 1989), pp. 1146-64.

Blanchard, Olivier Jean, and Danny Quah. "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* (September 1989), pp. 655-73.

Blanchard, Olivier Jean, and Mark W. Watson. "Are Business Cycles All Alike?" in Robert J. Gordon, ed., *The American Business Cycle* (University of Chicago Press, 1986).

Cooley, Thomas F., and Stephen F. LeRoy. "Atheoretical Macroeconomics: A Critique," *Journal of Monetary Economics* (November 1985), pp. 283-308.

Dickey, David A., and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association* (June 1979), pp. 427-31.

Doan, Thomas, Robert B. Litterman, and Christopher Sims. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Review* (Vol. 3, No. 1, 1984), pp. 1-100.

Engle, Robert F., and C. W. J. Granger. "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* (March 1987), pp. 251-76.

Fuller, Wayne A. *Introduction to Statistical Time Series* (Wiley Press, 1976).

Gali, Jordi. "How Well Does the IS-LM Model Fit Postwar U.S. Data," *Quarterly Journal of Economics* (May 1992), pp. 709-38.

Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* (July 1982), pp. 1029-54.

Keating, John W. "Postwar Business Cycles in a Natural Rate Model," revised version of Washington University Working Paper #159 (St. Louis, MO, July 1992).

_____. "Identifying VAR Models under Rational Expectations," *Journal of Monetary Economics* (June 1990), pp. 453-76.

Keating, John W., and John V. Nye. "Permanent and Transitory Shocks in Real Output: Estimates from Nineteenth Century and Postwar Economies," Washington University Working Paper #160 (St. Louis, MO, July 12, 1991).

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. "Stochastic Trends and Economic Fluctuations," *American Economic Review* (September 1991), pp. 819-40.

Nelson, Charles R., and Charles I. Plosser. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* (September 1982), pp. 139-62.

Runkle, David E. "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics* (October 1987), pp. 437-42.

Sargent, Thomas J. "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure," *Journal of Monetary Economics* (January 1979), pp.133-43.

Schwert, G. William. "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," *Journal of Monetary Economics* (July 1987), pp. 73-104.

Shapiro, Matthew D., and Mark W. Watson. "Sources of Business Cycle Fluctuations," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1988* (MIT Press, 1988), pp. 111-48.

Sims, Christopher A. "Are Forecasting Models Usable for Policy Analysis?" Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter 1986), pp. 2-16.

_____. "Macroeconomics and Reality," *Econometrica* (January 1980), pp. 1-48.

Sims, Christopher A., James H. Stock, and Mark W. Watson. "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* (January 1990), pp. 113-44.

Stock, James H., and Mark W. Watson. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," Federal Reserve Bank of Chicago Working Paper WP-91-4, (February 1991).

_____. "Testing for Common Trends," *Journal of the American Statistical Association* (December 1988), pp. 1097-107.

_____. "Variable Trends in Economic Time Series," *Journal of Economic Perspectives* (Summer 1988), pp. 147-74.

West, Kenneth D. "The Sources of Fluctuations in Aggregate Inventories and GNP," *Quarterly Journal of Economics* (November 1990), pp. 939-71.

Zellner, Arnold. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* (June 1962), pp. 348-68.

Anna J. Schwartz

Anna J. Schwartz is a Senior Research Fellow at The National Bureau of Economic Research. This paper was the sixth annual Homer Jones Memorial Lecture, which was presented at St. Louis University on April 9, 1992. The views expressed are those of the author and do not reflect official positions of the Board of Governors of the Federal Reserve System or the Federal Reserve Bank of St. Louis.

The Homer Jones Memorial Lecture is an annual event sponsored by the St. Louis Gateway Chapter of the National Association of Business Economists in conjunction with Washington University, St. Louis University, the University of Missouri-St. Louis and the Federal Reserve Bank of St. Louis.

Author's note: For sharing his knowledge of Federal Reserve practices, Walker F. Todd has been immensely helpful to me, but he is not responsible for the views I express in this paper. I have also benefited from comments by Allan H. Meltzer.

The Misuse of the Fed's Discount Window

I AM HONORED to have the opportunity to give the Homer Jones Memorial Lecture. In remembering him today, I pay tribute to his contributions to monetary policy making and to quantitative monetary research in his capacity as director of research at the St. Louis Fed. I got to know Homer during the period of his membership in the Shadow Open Market Committee. At our meetings he was diffident about his knowledge and insistent on scrupulous attention to statistical evidence as backing for the policy conclusions we reached. I cannot think of two more admirable qualities in an economist. It was a privilege for me to have had this association with Homer.

In 1925 the Federal Reserve Board collected data on the number of member banks continuously indebted to their Reserve Banks for at least a year. As of August 31, 1925, 593 member banks had been borrowing for a year or more. Of this number, 239 had been borrowing since 1920 and 122 had begun borrowing before that. The Fed guessed that at least 80 percent of the 259 national member banks that had failed since 1920 had been "habitual borrowers" prior to their failure. Of 457 continuous borrowers in 1926, 41 banks suspended operations in 1927, while 24 liquidated voluntarily or merged (Shull 1971, 34-35).

The reason for citing these statistics for the 1920s is to call attention to an early episode in Federal Reserve history that contravened the ancient injunction to central banks to lend only to illiquid banks, not to insolvent ones, and that is eerily similar to a current episode. The Fed apparently learned little from the earlier episode, since there is even less justification for the use made of the discount window in the current than in the earlier episode.

The current episode came to light after the House Banking Committee requested data on all insured depository institutions that borrowed funds from the discount window from January 1, 1985, through May 10, 1991. Regulators grade banks on their performance, according to

a scale of 1 to 5. The grades are based on five measures known by the acronym of CAMEL, for Capital adequacy, Asset quality, Management, Earnings, Liquidity.[1] The Federal Reserve reported that of 530 borrowers from 1985 on that failed within three years of the onset of their borrowings, 437 were classified as most problem-ridden with a CAMEL rating of 5, the poorest rating; 51 borrowers had the next lowest rating, CAMEL 4. One borrower with a CAMEL rating of 5 remained open for as long as 56 months. The whole class of CAMEL 5-rated institutions were allowed to continue operations for a mean period of about one year.

At the time of failure, 60 percent of the borrowers had outstanding discount window loans. These loans were granted almost daily to institutions with a high probability of insolvency in the near term, new borrowings rolling over balances due. In aggregate, the loans of this group at the time of failure amounted to $8.3 billion, of which $7.9 billion was extended when the institutions were operating with a CAMEL 5 rating. Three months prior to failure, borrowings of all 530 institutions peaked at $18.1 billion. Rather than encouraging banks to pursue strategies to preserve their size, regulators often encourage institutions that are about to fail to shrink drastically first, so as to diminish the pool of assets that have to be liquidated after closing.

Some observers of bank performance have asserted that it is impossible to know whether an institution that applies for discount window assistance faces a liquidity or solvency problem. That assertion may be defensible for discount window lending in the 1920s even though an estimated 80 percent of long-time borrowers

failed, since CAMEL ratings did not then exist. Currently, CAMEL ratings 4 and 5 are known promptly. Why should it be impossible or even difficult to distinguish between an illiquid and an insolvent bank?

Support by the Fed for banks with a high probability of insolvency in the near term is not the only recurrent problem with the discount window. Equally troublesome is the history of actual or proposed Fed capital loans to nonbanks. Such use of the discount window distracts the Fed's attention from its monetary control function. Recent experience reinforces earlier doubts about the need for the discount window. The time has come for a truly basic change: eliminate the discount window and retrict the Fed to open market operations.[2] This change would have the added value of obliterating the symbolic role of the discount rate and weakening the tendency to regard the Fed as determining interest rates.

In the rest of this paper, I document the erosion of the historic restriction, at least since the 1930s, of Federal Reserve discount window assistance to liquidity-strained banks on the security of sound assets. Section 1 deals with lending operations from the founding until the post-World War II period, during which loans to nonbanks first occur. I then discuss Federal Reserve actions in recent decades that have further blurred the distinction between liquidity and solvency, and also the emergence of various nonbanks as candidates for discount window assistance. I ask why these developments have occurred when there has been no change in official declarations of commitment to supply only liquidity, not solvency or capital, to individual banks, not nonbanks (section 2). In the next sec-

---

[1]Brief official descriptions of composite CAMEL 4 and 5 ratings follow:

CAMEL 4 "Institutions in this group have an immoderate volume of serious financial weaknesses or a combination of other conditions that are unsatisfactory. Major and serious problems or unsafe and unsound conditions may exist which are not being satisfactorily addressed or resolved."

CAMEL 5 "This category is reserved for institutions with an extremely high immediate or near-term probability of failure."

CAMEL ratings of banks are not uniform from one district to another. Some New York CAMEL 4 banks may be rated 5 elsewhere.

[2]This recommendation has been disputed on the ground that establishing access to the discount window as a right — not a privilege — administered at a penalty rate would solve the problems that face the current administration of the window. It is not clear to me, however, that these

changes would eliminate the problem of political pressures on the Fed to lend to nonbanks. As for the problem of loans to insolvent banks, access to the window as a right at a penalty rate might only result in worsening adverse selection.

See also Kaufman (1991), who argues that the discount window is not needed to protect the money supply — the basic justification for a lender of last resort — and that liquidity strains can be mitigated by open market operations without involving the Fed in discount window assistance. Credit-worthy banks can borrow at market rates, large ones in the Fed Funds market, small ones from their correspondent banks.

tion I examine the costs of Federal Reserve support for problem institutions that regulatory authorities eventually close and for nonbanks that would otherwise have to meet a market test (section 3). Finally, I consider whether reforms of discount window practices that have been proposed could remedy the inherent problems of the mechanism. I comment on provisions in the FDIC Improvement Act of 1991 that may be worthy reform proposals but do not address these problems (section 4). I offer my conclusions in section 5.

1. HISTORIC ROLE OF DISCOUNT WINDOW LENDING

A. Before the New Deal

Regulation A—the first one adopted by the Federal Reserve Board at its creation, in recognition of the expectation that the discount window at Federal Reserve Banks would serve as the main purveyor of member bank reserves—establishes the rules under which the Banks may extend credit. Periodically revised, the regulation has consistently set out the procedures that banks had to follow to gain access to the discount window. The initial regulation provided for rediscounting short-term agricultural, industrial, and commercial paper, defined as eligible paper. Later, to accommodate Treasury financing needs in World War I, the Reserve Banks were empowered to extend direct collateral loans to member banks, occasionally secured by pledge of eligible paper but usually by obligations of the U.S. government. Until the 1930s, even though the Federal Reserve had become familiar with open market purchases as a source of member bank reserves, bills discounted usually exceeded U.S. government securities in Reserve Bank portfolios.

The discount window provided accommodation for periods up to 90 days for a non-agricultural discount or advance collateralized by eligible paper or government obligations, but of up to nine months for agricultural paper. As noted earlier, continuous borrowing year in and year out in the 1920s was not uncommon. A later (1954) Federal Reserve document, deploring continuous borrowing, noted that "it was possible by the mid-Thirties to speak of an established tradition against member bank reliance on the discount facility as a supplement to its resources" (Shull 1971, 41). A similar statement

appears in an internal Federal Reserve history of the discount mechanism: "extended borrowings by a member bank from its Reserve Bank would in effect constitute a use of Federal Reserve credit as a substitute for the member's capital" (Hackley 1973, 194). The 1973 version of Regulation A states, as a general principle, that "Federal Reserve credit is not a substitute for capital and ordinarily is not available for extended periods." Both the 1980 and 1990 versions of Regulation A state, as a general requirement, that "Federal Reserve credit is not a substitute for capital."

A broader statement of the foregoing principle, covering banks and nonbanks, appeared in 1932 in a conference report by representatives of the Federal Reserve Bank of New York, who had met with South American central banks: "Central banks must not in any way supply capital on a permanent basis either to member banks or to the public, which may lack it for the conduct of their business" (*Federal Reserve Bulletin* 1932, 43).

Legislative changes in the administration of the discount window, made in response to bank failures in the Great Depression, sometimes observed this advice. The Glass-Steagall Act of February 27, 1932, authorized Federal Reserve Banks (in a new section 10(b) added to the Federal Reserve Act) to make advances to member banks on their promissory notes secured by otherwise ineligible collateral, if acceptable to the Reserve Banks, for periods up to four months at rates not less than one-half percent per annum above the prevailing highest discount rate. No provision was made for solvency loans of capital or loans to receivers of closed banks.

Although no provision was made for solvency loans of capital, the Emergency Relief and Construction Act of July 21, 1932 (in a new section 13(3) added to the Federal Reserve Act), opened the discount window to nonbanks. It permitted the Reserve Banks to discount for individuals, partnerships and corporations, with no other sources of funds, notes, drafts and bills of exchange eligible for discount for member banks. The New Deal confirmed this type of authority (in section 403 of Title III of the Emergency Banking Act of March 9, 1933, that added section 13(13) to the Federal Reserve Act). It allowed 90-day advances to individuals, partnerships and corporations on the security of direct obligations of the U.S. government, at interest rates fixed by the Reserve Banks.

B. From the New Deal to Post-World II

Before turning to the New Deal change that involved Reserve Banks in extending capital loans to nonbanks, let me review the other changes in the Emergency Banking Act of 1933. Title II created conservatorships for national banks. Section 304 of Title III authorized the Reconstruction Finance Corporation, not the Federal Reserve, to subscribe to preferred stock "of any national banking association or any State bank or trust company in need of funds for capital purposes." It is significant that Title IV, referring to the Federal Reserve, conferred on it no authority to make solvency loans of capital. Section 402 authorized Federal Reserve Banks, until March 3, 1934, to make advances in exceptional and exigent circumstances to member banks on their own notes on the security of any acceptable assets, superseding the provision regarding advances to member banks in the February 1932 Glass-Steagall Act. By Presidential proclamation, this provision, the forerunner of the present section 10(b), was extended until March 3, 1935, when it expired (Board *Annual Report* 1933, 261-265). The new form of section 10(b) became permanent as part of the Banking Act of 1935.

I now turn to the New Deal change that authorized the Federal Reserve to make solvency/capital loans to nonbanks. The Act of June 19, 1934, added a new section 13(b) to the Federal Reserve Act, which authorized Reserve Banks directly or in participation with a member or nonmember bank to make advances to established commercial or industrial enterprises for the purpose of supplying working capital if the borrower were unable to obtain assistance from usual sources. A participating member or nonmember bank had to obligate itself for at least 20 percent of any loss. The loans were for periods up to five years (Board *Annual Report* 1934, 50-51). Through 1939, the Reserve Banks had approved 2,800 applications and commitments amounting to $188 million at rates from 2.5 to 6 percent (Smead 1941, 259). Although the Reconstruction Finance Corporation then assumed a more important role in providing working capital for established businesses, the Reserve Banks continued to participate, approving an additional 742 applications amounting to $382 million through 1946 (Board *Annual Report* 1946, 10).

The aggregate amount of such loans was limited by law to the surplus of the Reserve Banks as of July 1, 1934, plus $139 million that the Treasury was to repay the Banks for their required subscription to the Federal Deposit Insurance Corporation in an amount equal to one-half of their surplus on January 1, 1933. The required subscription was stipulated in the Banking Act of 1933 that created the FDIC (Board *Annual Report* 1933, 276-277). A commentator has noted that this "is the only instance in United States history in which Congress required the central bank to expend its own funds to subscribe for more than a *de minimis* amount of the capital of another unrelated enterprise, other than obligations of the Treasury itself" (Todd 1988, 60). In any event, the Reserve Banks charged off the value of FDIC stock on their books in the same calendar year in which they paid for the subscription. Capital loans to nonbanks under the authority of section 13(b), unlike the FDIC subscription, were voluntary decisions of the Reserve Banks. In section 2 below, I note a more recent attempt by the Treasury that was foiled to require the Fed to expend its own funds in support of the FDIC.

Congressional authorization and Federal Reserve implementation of loans to nonbanks for use as capital was, in my judgment, a sorry reflection on both Congress's and the Fed's understanding of the System's essential monetary control function. In 1946 the Federal Reserve Board sought to eliminate its authority under section 13(b) to make loans directly to business enterprises and replace it with a mandate restricted to guaranteeing such loans. A bill introduced in Congress early in 1947 on the Board's behalf would have authorized Reserve Banks to guarantee, up to a maximum of 90 percent, loans, extended by banks for as long as 10 years to small- and medium-size business enterprises, subject to a fee charge that increased with the guarantee percentage. The bill also provided for the return of the amounts advanced by the Treasury (up to $139 million) for the Banks' industrial loan operations, and pledged that no further Treasury appropriations for this purpose would be required (Board *Annual Report* 1946, 8-10; 1947, 11-12).

Since the bill was not enacted, Reserve Banks for the next decade continued to make and cofinance working capital industrial loans. Section 13(b) was finally repealed by the Small Business Investment Act of 1958, under the terms of which the Reserve Banks restored to the Treasury the amounts it had advanced under section

13(b) (Board *Annual Report* 1958, 95). Chairman William McChesney Martin, in testimony before the Subcommittee on Small Business of the Senate Banking and Currency Committee, when it was considering this bill, stated well the Federal Reserve's considered judgment on its venture into capital industrial loans: Its primary objective was "guiding monetary and credit policy," and "it is undesirable for the Federal Reserve to provide the capital and participate in management functions" of the proposed small business financing institutions (*Federal Reserve Bulletin* 1957, 769).

I conclude this survey of Federal Reserve lending activities since its founding by noting its support for solvency/capital lending programs under wartime V-loan authority that it adopted on April 6, 1942, and revised on September 26, 1950 (Board *Annual Report* 1942, 89-91; 1950, 72-74). Federal Reserve Banks were authorized to act on behalf of the guaranteeing agencies (War Department, Navy Department, etc.) as fiscal agents of the United States, meaning, the Treasury was required to reimburse them for their outlays. In acting as a guarantor of defense production loans, the Federal Reserve provided a model of guaranteeing authority that was later invoked when bailouts of peacetime enterprises were proposed in the 1970s. Those developments and Federal Reserve lending since the 1980s to institutions with a high probability of insolvency in the near term represent a major departure from its historic mandate to provide loans to illiquid but not insolvent depository institutions. In the section that follows I discuss the apparent change in how the Federal Reserve regards its mandate.

2. ASSISTANCE TO INSOLVEMENT NONBANKS AND BANKS SINCE THE 1970s

A. Assistance to Insolvent Nonbanks

An appropriate starting point is the official response to financial problems of the Penn Central Railroad that surfaced in 1970. Though it did not lead to discount window assistance, nonetheless it reveals clearly the pressures that were to lead to such a practice. The Nixon Administration proposed to give the company a V-loan as a bailout. However, for the loan to be of any use, legislative approval was required, since guarantees of loans for defense production were

to expire on June 30. When legislation stalled in Congress, the Administration requested the Federal Reserve Board to authorize the Federal Reserve Bank of New York to give the company discount window assistance. The request was rejected and, as a result, the company filed for bankruptcy on June 21, 1970. On June 30, Congress approved a Joint Resolution, extending the Defense Production Act but limiting the maximum obligation of any guaranteeing agency (for example, the Federal Reserve) to $20 million, and stipulating that "The authority conferred by this section shall not be used primarily to prevent the financial insolvency or bankruptcy of any person, unless" the President certifies "a direct and substantially adverse effect upon defense production" (*Federal Reserve Bulletin* 1970, 720).

If the incident had ended at this point, it would have been remembered primarily as a political dispute between the Administration and the Congress. Penn Central's bankruptcy would have been regarded simply as the key to the restructuring of its operations. Instead, the Federal Reserve reacted as though the company's default on $82 million of commercial paper it had outstanding would generate a financial crisis. The Federal Reserve assumed that lenders would not discriminate between a troubled issuer and other perfectly sound issuers of commercial paper, so that the latter would be unable to roll over their issues. "It was made clear that the Federal Reserve discount window would be available to assist banks in meeting the needs of businesses unable to roll over maturing commercial paper" (Board *Annual Report* 1970, 18).

However, commercial paper issuers that faced difficulties did so not because of the condition of the market as such, but because of conditions peculiar to themselves (for example, Chrysler Financial and Commercial Credit) (Carron 1982, 398). The fully justified verdict of the 1971 *Economic Report of the President* (69) was that no "genuine liquidity crisis existed in mid-1970."

The Penn Central episode fostered the view that bankruptcy proceedings by a large firm created a financial crisis, and that, if possible, bankruptcy should be prevented by loans and loan guarantees: the "too big to fail" doctrine in embryo.

The financial difficulties faced by New York City in 1975 led the Federal Reserve to inform

Congress of its response to suggestions that it might serve as a source of emergency credit. Governor George W. Mitchell cautioned "against any proposals that would provide direct access to central bank credit by hard-pressed governmental units" (*Federal Reserve Bulletin* 1975, 410). Chairman Arthur F. Burns reiterated that caution and reported on contingency plans to increase temporary discount window lending to commercial banks in the event of a major municipal default. However, he added that if a default ultimately required "write-downs that could seriously impair the capital of some banks," the Federal Deposit Insurance Corporation, not the Federal Reserve, had statutory powers to assist federally insured banks that found their capital impaired (*Federal Reserve Bulletin* 1975, 635-636).

In the end, Congress passed a law guaranteeing New York City loans of up to $2.3 billion in 1975, reduced to $1.65 billion in 1978, but the Federal Reserve's involvement in the rescue arrangement was only limited fiscal agency services.[3] The Fed also acted as fiscal agent for Treasury loan guarantees issued during the bailouts of the Lockheed (1971) and Chrysler (1979) corporations.[4]

I have been unable to find any reference to the Fed's involvement in these three loan guarantees in any of its publications. By consulting the U.S. Code—a source for lawyers, not economists—however, I have been able to ferret out the details of that involvement. My guess is that the Fed was unwilling to publicize its role because it was not voluntary.

The most recent attempt to use the discount window to assist a nonbank involved the FDIC. In March 1991, the FDIC chairman, William Seidman, requested Congressional authorization for a direct loan of $25 billion by the Federal Reserve to the Bank Insurance Fund. Chairman Alan Greenspan, testifying before the Senate Banking Committee the next month, opposed such a loan. That did not deter Treasury Under Secretary for Domestic Finance Robert Glauber from renewing the request in testimony on May 29, 1991, before a subcommittee of the House Ways and Means Committee. The Federal Reserve's required subscription to the FDIC on its establishment in his view served as a precedent. The FDIC recapitalization and banking reform bill that the Treasury prepared included

provisions authorizing the FDIC to borrow $25 billion from the Federal Reserve Banks and amending section 13 of the Federal Reserve Act, authorizing any Federal Reserve Bank "to make advances to the Federal Deposit Insurance Corporation, upon its request" (Treasury bill 1991, sec. 402, 245). The July 23, 1991, bill prepared by the House Committee on Banking, Finance and Urban Affairs did not include those provisions (H.R.6, 102nd Cong., 1st sess.). The final legislation, the FDIC Improvement Act of 1991, follows the House bill in increasing from $5 billion to $30 billion the FDIC's authority to borrow directly from the Treasury, not from the Fed.

Despite the restraint in the Act with respect to recapitalizing the FDIC, restraint is absent from another provision. The Act amended Section 13 of the Federal Reserve Act that deals with the Federal Reserve's authority to discount for nonbanks. The amendment eliminated the requirement that the notes, drafts or bills tendered by nonbanks be eligible for discount by member banks. As interpreted by Sullivan & Cromwell, a New York law firm, for its clients in a memorandum of December 2, 1991, this provision enables the Fed to lend directly to security firms in emergency situations. Traditionally, commercial banks, knowing they had access to the discount window, have lent to brokerage firms and others short of cash in a stock market crash. It is not clear why the traditional practice was deemed unsatisfactory. In my view, the provision in the FDIC Improvement Act of 1991 portends expanded misuse of the discount window.

To this date, the Fed has apparently been a reluctant participant in loans and loan guarantees to nonbanks. The question must be asked whether it will be firm in the future in resisting pressures to fund insolvent firms that are politically well-connected.

B. Assistance to CAMEL 4- and 5-Rated Banks

In 1974 the Federal Reserve behaved contrary to traditional principles when uninsured depositors started a run on the Franklin National Bank after news surfaced that it had large foreign exchange losses.The Comptroller of the Currency did not close it promptly. The decision of the

---

[3]See U.S. Code (1975, 1978).

[4]See U.S. Code (1971) for Lockheed and U.S. Code (1980) for Chrysler.

regulators was that the Federal Reserve discount window, starting in May, would provide loans until the FDIC found a purchaser of the failed institution. Over the next five months, the Federal Reserve Bank of New York lent continuously to Franklin; the maximum amount lent, on October 7, 1974, was $1.75 billion, representing nearly one-half of Franklin's assets. On October 8, the bank was declared insolvent and taken over by a foreign consortium.

Among the precedents established by discount window lending to Franklin National was that its London branch assets were accepted as collateral, and that, for the first time, the borrowings covered withdrawals from the London branch as well as Franklin's domestic branches. Although, on one hand, Franklin National simply borrowed at the discount rate, what was unusual in this episode was that in September 1974 the Federal Reserve assumed responsibility to execute Franklin's existing foreign exchange contracts, since bidders for the bank were unwilling to honor them. It also agreed to extend discount window assistance, if needed, to the purchasing bank. The FDIC, moreover, did not immediately repay the discount window loan—its normal practice—but signed a three-year note obligating itself to do so as the collateral Franklin supplied was liquidated. In effect, the Federal Reserve lent capital funds that the insurance agency contributed to the purchasing bank. The interest cost to the Fed of subsidizing loans to Franklin has been estimated at $20 million (Garcia and Plautz 1988, 228). In executing Franklin's foreign exchange contracts, the Fed also incurred opportunity costs of staff time and lost interest on part of its portfolio.

The rescue of Franklin National Bank shifted discount window use from short-term liquidity assistance to long-term support of an insolvent institution pending final resolution of its problems. The bank was insolvent when its borrowing began and insolvent when its borrowing ended. The loans merely replaced funds that depositors withdrew, the inflow from the Reserve Bank matching withdrawals.

The undeclared insolvency of Continental Illinois in 1984 was also papered over by extensive discount window lending from May 1984 to February 1985, albeit with smaller subsidies than in the case of Franklin National—an amendment to Regulation A as of September 25, 1974, permitted application of a special rate on emergency credit after eight weeks that was closer to a market rate. The borrowing covering Continental's holding company as well as the bank at some dates amounted to as much as $8 billion. Again the FDIC assumed the bank's loan, which it eventually repaid from the proceeds of liquidating the bank's assets, concluding with one large $2.1 billion payment in September 1989—an enormous cash drain.

The discount window has been valued by the Federal Reserve as a mechanism for directing funds to an individual bank with liquidity problems. It regarded its loans to Franklin National and Continental Illinois as exceptional occurrences. However, when hundreds of CAMEL 4- and 5-rated banks, as noted at the beginning of this paper, were receiving extended accommodation even though they faced a high probability of near-term failure, discounting can no longer be regarded simply as a means of providing temporary liquidity. What explains this transformation in practice?

C. Why the Departure From the Historic Norm of Discount Window Use?

In the United States, with federal spending budgeted at an all-time high, policy makers see the discount window as a mechanism for providing funds off budget. Legislation is necessary to authorize the Fed to provide assistance to favored nonbanks, so the use of the window isn't kept secret, but it may seem a cost-free way of funding them because repayment can be rolled forward indefinitely. This may explain the recent spate of efforts to use discount window assistance for nonbanks.

With respect to loans to banks with a high probability of insolvency in the near term, it is noteworthy that in 1974 only four banks failed. By the late 1980s, failures were numbered in triple digits, and the FDIC's problem banks, in four digits. Having once set the precedent of lending to such problem institutions, at least two explanations may account for this practice by the Federal Reserve: (1) As in the case of Franklin and Continental Illinois Banks, its actions may have been taken at the bidding of the FDIC, or perhaps on its own initiative, in order to mitigate the FDIC's plight. (2) The Federal Reserve may regard failure of a large institution as potentially triggering a financial collapse or a run on the dollar. Fear of contagion may have become the Federal Reserve's overriding concern.

Even if these explanations account for the change in Federal Reserve discount window practice, neither may justify the change. Before dealing with this issue, it is important to assess the costs of wholesale discount window lending to insolvent institutions, to which I now turn.

3. COSTS OF LENDING TO INSOLVENT INSTITUTIONS

For the Fed to lend directly to the Treasury or to government agencies that the Treasury would otherwise fund through regular appropriations is a slippery slope. The costs are politicization of the money supply process. The Fed's charter wisely prohibits such lending.

Discount window accommodation to insolvent institutions, whether banks or nonbanks, misallocates resources. Political decisions substitute for market decisions. Institutions that have failed the market test of viability should not be supported by the Fed's money issues.

A depository institution traditionally was said to be eligible for discount window assistance when it was illiquid but solvent. In recent years, it has been given assistance when it was liquid but its insolvency was undeclared. On a market value basis, an institution is insolvent when assets are less than liabilities. Since book values are the usual measure of assets and liabilities, the divergence between assets and liabilities may only be revealed long after the market value of its assets has fallen below the market value of its liabilities. In addition, an institution may be liquid but insolvent, so long as its cash flow is positive.

A decision to declare an institution insolvent is the prerogative of the chartering agency: the Comptroller of the Currency for national banks, state authorities for state-chartered banks supervised by the Fed and the FDIC. The FDIC, since 1989, however, can both remove deposit insurance and substitute itself as receiver of state banks. It can force a state to close banks and, as the thrift insurer, close insolvent insured thrifts. (It does not have authority to close credit unions.) If the chartering agency has delayed closure, the Federal Reserve has acted as if a troubled institution is entitled to discount window assistance provided it can fur-

nish acceptable collateral. The question I raise is whether the Federal Reserve's position is defensible when inferior CAMEL ratings provide independent evidence on the likelihood of insolvency in the near or immediate future.

Since the Federal Reserve routinely sterilizes discount window reserve infusions, does it make any difference whether the reserves are provided to solvent or insolvent banks or whether the period for which the reserves are provided is limited or extended? After all, if the reserves were not made available through the discount window, open market purchases would add an equivalent reserve contribution to the banking system.

I believe that it does make a difference whether reserves are injected by open market purchases or by discount window lending, especially if insolvent banks are permitted to borrow for extended periods. Discount window lending may not affect proposed monetary growth, but it has other effects that make it an undesirable source of reserves.

Open market operations are anonymous. The market allocates reserve injections or withdrawals among participants according to their relative size and current opportunities. Much greater discretion is exercised by the Federal Reserve in the allocation of reserves through discounting, since the Fed knows the institutions that request discount accommodation. The public learns about the magnitude of both open market operations and discount window credit from the data the Federal Reserve publishes, but it does not learn the names of the banks that received loans. The data made available to the House Banking Committee in 1991 revealed the names of the institutions that had failed despite extended discount window loans, but not the names of the few banks that had received such loans but had not failed.[5] The secrecy may be good public policy, but it leaves open the question whether provision of loans on a case-by-case basis assures equal treatment for all. This is an argument against discount window lending in general, not specifically to insolvent banks, an argument that has often been made in the past without reference to the specific problem of insolvent banks (for example, Friedman 1960, 38).

[5] It was only extended credit borrowers that failed in the data the House Banking Committee obtained from the Fed in 1991. Banks that obtained seasonal credit at the window did not fail. As footnote 2 on page 59 contends, the discount window is not essential for this use.

Since the 1970s, the Federal Reserve has extended long-term discount window assistance to depository institutions that by objective standards were likely to fail. It has done so in the belief that, in the absence of such assistance, contagious effects would spread from the troubled institutions to sound ones. The belief is particularly entrenched for large troubled intermediaries, reflecting an apprehension that halting the operation of such institutions would have dire unsettling effects on financial markets. Before 1985, the goal of such discount window assistance was a restructuring of the problem institution as a viable entity with both insured and uninsured deposits made whole.

That was the situation in 1984 when Continental Illinois was rescued. The implication was that any other response would have brought on contagion. Yet if the bank had been closed before its net worth turned negative, the institutional depositors, foreign bank depositors and creditors who ran on it might well have redeposited their withdrawals elsewhere or bought financial assets to replace the certificates of deposit Continental issued and that they were no longer willing to buy. Even if some interbank depositors that held as much as half their equity in uninsured deposits at Continental had obtained only a fraction of their claims immediately upon the closing, they would ultimately have recovered the full nominal value of their claims after liquidation of the bank.[6] The market would have known that the claimants on Continental were not in jeopardy. Even if closing Continental had led to runs on the foreign interbank depositors — ostensibly the reason for keeping Continental in operation — the lenders of last resort in the nations concerned could have provided adequate liquidity in their markets to tide the banks over if the Continental deposits were their only problem. Fear of contagion should not determine a regulator's decision to keep an insolvent bank open. It should lead the Fed to lend to the market to prevent the contagion.

If fear of contagion is a lesson the Federal Reserve has learned from the banking panics of 1930-33, it is the wrong lesson. Contagion then occurred in an environment in which the Fed permitted the money supply to decline drastically, rendering banks insolvent not because of their own actions but simply because of the collapsing economy. The right lesson is that contagion need not arise if open market purchases are made adequate both to reassure the market and to prevent a collapse in the quantity of money. Examples are the Fed's provision of liquidity to cushion the economy from the effects of the 1987 stock market crash and the collapse of Drexel Burnham.

Since 1985, prolonged discount window assistance has generally terminated not with restructuring but with closure of the insolvent banks. When banks are known to be insolvent, postponement of recognition of losses that have occurred might well have increased current losses. Uninsured depositors have more time to withdraw their funds. The insurance agency, which is to say the taxpayer, ultimately bears any added costs of delayed closure. By lending to the banks in question, the Federal Reserve encourages this practice. Absent regulatory restraints or incentives to the contrary, the policy clearly encourages risk-taking and invites moral hazard problems. If a bank with the least desirable CAMEL rating can obtain subsidized discount window assistance, that institution obtains a competitive advantage over solvent ones for as long as the Federal Reserve supports it.[7]

The Federal Reserve Banks decide whether they will extend a loan on the basis of collateral that the would-be borrowers offer. This decision is undoubtedly influenced by the condition of the insurance agency: whether it has the funds to pay off depositors and take over the failing bank's assets when it cannot arrange a merger of an insolvent bank with a solvent one or arrange removal of existing management by placing an institution in receivership. The condition of the insurance agency in turn is also affected by a chartering agency's forbearance or prompt action in declaring the insolvency of one of its constituents. The Federal Reserve has cooperated by extending long-term support to an institution with a high probability of failure until a resolution of its problems was arrived at.

Discount window lending to insolvent banks might cease if, under the terms of the FDIC Im-

---

[6]This assumes that the bank's value would have been realized in a forced sale of assets.

[7]The subsidized rate may have risen to take into account rates on market sources of funds but it is still a subsidy compared to what deposit brokers and other non-Fed sources of funds would charge if the dying banks were left to fend for themselves in the market.

provement Act of 1991, the Fed no longer advanced funds to keep critically undercapitalized institutions in operation, and the insurance agency took prompt corrective action to appoint a receiver for those institutions.[8]

4. COULD REFORM REMEDY WHAT'S WRONG WITH THE DISCOUNT WINDOW?

Would discontinuation of Fed lending to insolvent banks establish the inviolability of the discount window? For many reasons that is not the case.

Regardless of one's attitude to discount window lending for seasonal and adjustment purposes—the private sector could accommodate those needs—many economists customarily assign one indispensable function to the discount window, namely, as lender of last resort: The Federal Reserve should use discounting in "exceptional circumstances" (in the words of Regulation A) to provide extended credit to solvent institutions with liquidity problems.

However, the Federal Reserve does not need the window to serve as a lender of last resort. The case against operation of the discount window has rested on grounds that open market operations are sufficient for the execution of monetary policy in both ordinary and exceptional circumstances, that individual banks that need and can justify special assistance can receive such assistance through the federal funds market, and that discounting invites discretionary subsidies to banks favored by the Fed (Goodfriend and King 1988, 216-53).

There is still another reason that the discount window serves no useful purpose. A review of the use of the discount mechanism by the Federal Reserve since its founding demonstrates a series of misconceptions on its part about what it can achieve by affording banks the opportunity to acquire borrowed reserves. The misconceptions have varied over time, depending on the objectives the System pursued.

Let me note some of the misconceptions:

(1) The Federal Reserve can determine whether borrowed reserves serve "productive" rather than "speculative" use of credit.

(2) Banks borrow only for "need" and not for profit.

(3) The absolute level of free reserves or borrowings indicates whether banks choose to acquire or liquidate assets.

(4) The spread between discount rates and market rates does not affect bank borrowing.

(5) The tradition against continuous borrowing is a satisfactory substitute for a penalty discount rate.

(6) Borrowing by banks signals bank weakness.

(7) Little bank borrowing signifies money market ease.

(8) "Technical" adjustments may explain discount rate changes, but their announcement conveys useful information to financial markets.

(9) Banks can be dissuaded from excessive use of the discount window by Reserve Bank appeals and exhortations instead of discount rate increases.

Since the 1970s, the Federal Reserve has acted on the belief that discount window assistance to banks with a high probability of insolvency in the near term, especially large ones, will, by delaying closure, eliminate contagious effects on financial markets.

In practice, the Federal Reserve's discount window activities have created perverse incentives, shifting risk from depositors to taxpayers. If a threat of systemic bank failures did arise, the Fed should counter it by open market operations, rather than by assistance to individual institutions. However, if the Federal Reserve prevents serious declines in the money supply, it is highly unlikely that the failure of an individual bank, however large, would trigger systemic bank failures.

Closing the discount window would free the Fed not only from likely future misconceptions but also from the recurrent pressures, to which it has been subject since the New Deal, to accommodate nonbanks, not to mention pressures from the Treasury to fund government agencies off budget.

5. CONCLUDING COMMENTS

I've surveyed the changes in Federal Reserve discount window practices since the System's founding. Early on the System emphasized that

---

[8]For an alternative view of the possible consequences of a prompt corrective action strategy, see R. Alton Gilbert (1991).

borrowing was supposed to be limited to short-term reserve needs. By the 1980s, hundreds of banks rated by regulators as having a high probability of failure in the near term and which ultimately failed were receiving extended accommodation at the discount window.

My reading of this recent experience is that the change in discount window practices, by delaying closure of failed institutions, increased the losses the FDIC and ultimately taxpayers bore.[9] Recent legislation limits use of the discount window for long-term loans to troubled banks. More important, it also provides that a supervisory agency is to appoint a receiver for an institution that falls below a critical capital ratio, curtailing the regulator's discretion regarding when to intervene in the case of an undercapitalized bank.

These changes, even if implemented—it is not yet known whether they will be—do not sanctify the continued operation of the discount window. Individual banks that need assistance, if credit-worthy, can obtain loans without subsidies from the federal funds market. The Fed can be an effective lender of last resort if restricted to open market operations. In administering the discount window, the Fed has been prone to mistake the effects of its actions. These mistakes have marred its execution of monetary policy.

Without a discount window, the Fed will avoid pressures to lend off budget to nonbanks and to government agencies, which should be funded through regular appropriations. Without the distraction of monitoring collateral and deciding which bank applicants qualified for assistance, it can concentrate its energies on open market operations, the single instrument it needs to control the quantity of money. Without a discount window, there will be no announcement effects since the Fed will not have to set the discount rate. That should dispel the impression that it controls market interest rates.

A Federal Reserve System without the discount window would be a better functioning institution.

REFERENCES

Board of Governors of the Federal Reserve System. *Annual Report* (Washington, D.C., selected years).

Economic Report of the President (Washington, D.C., 1971).

Federal Deposit Insurance Corporation Improvement Act of 1991.

Federal Reserve Bulletin (Washington, D.C., selected years and months).

---

[9]My conclusion from reading the 1991 Senate and House hearings on the long-delayed closings of the Bank of New England and Madison National Bank is that delay increased resolution costs. The condition of the institutions deteriorated as the Fed continued to lend to them — they were already rated CAMEL 5 when the continuous loans began. With the value of liabilities greater than of earning assets in these institutions, the gap grew as interest incurred on liabilities exceeded earnings on assets.

Delay, moreover, allowed outflows of uninsured deposits. Had the institutions been closed promptly, the earnings deficiency could have been offset at least somewhat by reducing the principal paid to uninsured depositors. Even so, at the closing, Bank of New England held $2.3 billion in uninsured deposits, of which $1.25 billion were brokered. At their peak, Fed advances amounted to $2.26 billion. The FDIC's loss was $2.5 billion. As L. William Seidman testified, the FDIC decided to protect all depositors of the Bank of New England, at "the additional cost [of] somewhere in the $200 to $300 million range up front" (see his statement in Senate Hearings on the failure of the Bank of New England, pp. 25-26).

In the absence of Fed lending to a bankrupt institution, early closing would have prevented a flight of uninsured deposits. In effect, Fed lending merely replaced withdrawals of uninsured deposits. Before it was declared insolvent, the Fed had lent the Madison National Bank $125 million, which kept the bank open to permit withdrawals of $108 million in uninsured deposits. The FDIC's final loss was $156 million.

The FDIC suffers losses not only in cases in which it liquidates failed banks but also in cases in which it arranges some form of purchase and assumption. When it liquidates a failed bank, the FDIC pays depositors the face value of their claims and suffers a reduction in its reserves measured by the full difference between book and current values of the assets that support the deposits. When it arranges a purchase and assumption, it transfers the deposits to the buyer and, depending on the purchase price it has accepted, the FDIC is responsible for the remaining difference between book and current values of assets that support the deposits transferred.

In my opinion, delay in recognizing losses that already exist cannot be justified by the claim that the FDIC uses the time to improve the behavior of the bankrupt institution, even if it were true that supervisors would be successful at this late date in the institution's history in reforming it. Delay may, however, be helpful to the FDIC's public image by postponing a publicly disclosed decline in the stated value of the reserves in the fund. That may be the real reason the FDIC welcomes delay.

The FDIC has stated that the value of a bank to potential bidders goes down when it is already involved in resolving a failed bank case, as if that would validate delay. In fact, it is the difference between book and market value of a bank's assets which a potential bidder learns that accounts for any decline in the bank's value, not the fact that an effort at resolution is under way.

What needs to be explained is why the Fed is the lender and not the FDIC, which has had the authority since 1982 to lend to open bankrupt banks and since 1989 to conservators. Some buyers might argue that assets are worth more if FDIC can step in after it has fixed a price in a purchase and assumption transaction and abrogates contracts the institution has with third parties, e.g., for rent, utilities, etc. That would be harder to do legally if the FDIC already controlled the bank or had an outstanding loan to it. That is not an argument that should encourage the Fed to lend instead of the FDIC.

Friedman, Milton. *A Program for Monetary Stability.* The Millar Lectures. Number Three (Fordham University Press, 1960).

Garcia, Gillian and Elizabeth Plautz. *The Federal Reserve: Lender of Last Resort* (Ballinger, 1988).

Gilbert, R. Alton. "Supervision of Undercapitalized Banks: Is There a Case for Change?" this *Review* (May/June 1991), pp.16-30.

Goodfriend, Marvin and Robert G. King. "Financial Deregulation, Monetary Policy and Central Banking," in W.S. Haraf and R.M. Kushmeider (eds.) *Restructuring Banking & Financial Services in America* (American Enterprise Institute, 1988).

Hackley, Howard. *Lending Functions of the Federal Reserve Banks: A History* (Board of Governors of the Federal Reserve System, 1973).

Kaufman, George. "Lender of Last Resort: A Contemporary Perspective," *Journal of Financial Services Research*, Vol. 5 (1991), pp. 95-110.

Shull, Bernard. "Report on Research Undertaken in Connection With a System Study," in *Reappraisal of the Federal Reserve Discount Mechanism*, Vol. 1 (Board of Governors of the Federal Reserve System, 1971), pp. 27-76.

Smead, Edward L. "Operations of the Reserve Banks," in *Banking Studies* (Board of Governors of the Federal Reserve System, 1941), pp. 249-70.

Todd, Walker F. "Lessons of the Past and Prospects for the Future in Lender of Last Resort Theory," Federal Reserve Bank of Cleveland Working Paper 8805 (1988).

U.S. Code. Emergency Loan Guarantee Act. 15 United States Code Sections 1841-1852. Public Law 92-70; 85 Stat. 178-182 (August 9, 1971), pp. 193-98; Legislative History, pp. 1270-78.

_____. New York City Seasonal Financing Act of 1975. Formerly enacted (now omitted from Code) as 31 United States Code Sections 1501-1510. Public Law 94-143; 89 Stat. 797 (Dec. 9, 1975), pp. 797-99; Legislative History, p. 1509.

_____. New York City Loan Guarantee Act of August 8, 1978. Formerly enacted as 31 United States Code Sections 1521-1531; see note under 31 U.S.C. Section 6702. Public Law No. 95-339.

_____. Chrysler Corporation Loan Guarantee Act of 1979. 15 United States Code Sections 1861-1875. Public Law 96-185;85 Stat. 1324 (January 7, 1980), pp. 1324-37.

U.S. House of Representatives, Committee on Banking, Finance and Urban Affairs, Staff. "An Analysis of Federal Reserve Discount Window Loans to Failed Institutions" (Processed, 1991).

U.S. House of Representatives, Committee on Banking, Finance and Urban Affairs. *A Bill to Reform the Deposit Insurance System ... and For Other Purposes* (1991).

_____. *Failure of Madison National Bank*, Hearing, 102nd Cong. 1st sess., Serial No. 102-38 (Government Printing Office, May 31, 1991).

_____. *The Failure of the Bank of New England Corporation and Its Affiliate Banks*, Hearing, 102nd Cong. 1st sess., Serial No. 102-49 (Government Printing Office, June 13, 1991).

U.S. Senate, Committee on Banking, Housing, and Urban Affairs. *The Failure of the Bank of New England*, Hearings, 102nd Cong. 1st sess., Serial No. 102-354 (Government Printing Office, January 9, May 6, and September 19, 1991).

U.S. Treasury Department, Capital Markets Legislation Office. "A Bill to Reform the Federal Deposit Insurance System ... and for Other Purposes: (Processed, 1991).