Derivatives Debacles

Case Studies of Large Losses in Derivatives Markets

Anatoli Kuprianov

To avoid all mistakes in the conduct of great enterprises is beyond man's powers. Plutarch, *Lives: Fabius.*

Recent years have witnessed numerous accounts of derivatives-related losses on the part of established and reputable firms. These episodes have precipitated concern, and even alarm, over the recent rapid growth of derivatives markets and the dangers posed by the widespread use of such instruments.

What lessons do these events hold for policymakers? Do they indicate the need for stricter government supervision of derivatives markets, or for new laws and regulations to limit the use of these instruments? A better understanding of the events surrounding recent derivatives debacles can help to answer such questions.

This article presents accounts of two of the costliest and most highly publicized derivatives-related losses to date. The episodes examined involve the firms of Metallgesellschaft AG and Barings PLC. Each account begins with a review of the events leading to the derivatives-related loss in question, followed by an analysis of the factors responsible for the debacle. Both incidents raise a number of public policy questions: Can government intervention stop such incidents from happening again? Is it appropriate for the government even to try? And if so, what reforms are indicated? These issues are addressed at the end of each case study, where the lessons and public policy concerns highlighted by each episode are discussed.

Alex Mendoza assisted in the preparation of this article. Ned Prescott, John Walter, and John Weinberg provided valuable comments on earlier drafts. Any remaining errors or omissions are the responsibility of the author. The views expressed are those of the author and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1. RISK AND REGULATION IN DERIVATIVES MARKETS

Perhaps the most widely cited report on the risks associated with derivatives was published in 1993 by the Group of Thirty—a group consisting of prominent members of the international financial community and noted academics. The report identified four basic kinds of risks associated with the use of derivatives.[1] *Market risk* is defined as the risk to earnings from adverse movements in market prices. Press accounts of derivatives-related losses have tended to emphasize market risk; but the incidents examined in this article illustrate the importance of *operational risk*—the risk of losses occurring as a result of inadequate systems and control, human error, or management failure.

Counterparty credit risk is the risk that a party to a derivative contract will fail to perform on its obligation. Exposure to counterparty credit risk is determined by the cost of replacing a contract if a counterparty (as a party to a derivatives contract is known) were to default.

Legal risk is the risk of loss because a contract is found not to be legally enforceable. Derivatives are legal contracts. Like any other contract, they require a legal infrastructure to provide for the resolution of conflicts and the enforcement of contract provisions. Legal risk is a prime public policy concern, since it can interfere with the orderly functioning of markets.

These risks are not unique to derivative instruments. They are the same types of risks involved in more traditional types of financial intermediation, such as banking and securities underwriting. Legal risk does pose special problems for derivatives markets, however. The novelty of many derivatives makes them susceptible to legal risk because of the uncertainty that exists over the applicability of existing laws and regulations to such contracts.

Although the risks associated with derivatives are much the same as those in other areas of finance, there nonetheless seems to be a popular perception that the rapid growth of derivatives trading in recent years poses special problems for financial markets. Most of these concerns have centered on the growth of the over-the-counter (OTC) derivatives market. As Stoll (1995) notes, concern about the growth of OTC derivatives markets has arisen because these instruments are nonstandard contracts, without secondary trading and with limited public price information. Moreover, OTC markets lack some of the financial safeguards used by futures and options exchanges, such as margining systems and the daily marking to market of contracts, designed to ensure that all market participants settle any losses promptly. The absence of such safeguards, along with the complexity of many of the new generation of financial derivatives and the sheer size of the market, has given rise to concerns that the growth of derivatives trading might somehow contribute to financial instability. Finally,

[1] See Global Derivatives Study Group (1993).

there is some concern among policymakers that the federal financial regulatory agencies have failed to keep pace with the rapid innovation in OTC derivatives markets.[2] Such concerns have only been reinforced by frequent reports of derivatives-related losses in recent years.

The traditional rationale for regulating financial markets stems from concerns that events in these markets can have a significant impact on the economy. Much of the present-day financial regulatory system in the United States evolved as a response to financial panics that accompanied widespread economic recessions and depressions. For example, the creation of the Federal Reserve System was prompted in large part by the Panic of 1907; the advent of federal deposit insurance was a response to the thousands of bank failures that accompanied the Great Depression.

The present-day financial regulatory system has several goals. The most important is to maintain smoothly functioning financial markets. A prime responsibility of institutions like the Federal Reserve is to keep isolated events, such as the failure of a single bank, from disrupting the operation of financial markets generally. During the twentieth century, U.S. financial market regulation expanded to encompass at least two more goals. The creation of a system of federal deposit insurance in 1933 gave the federal government a stake in the financial condition of individual commercial banks, since a federal agency was now responsible for meeting a bank's obligations to its insured depositors in the event of insolvency. In addition, Congress enacted the Securities Exchange Act to help protect investors by requiring firms issuing publicly traded securities to provide accurate financial reports. The act created the Securities and Exchange Commission (SEC) to regulate the sales and trading practices of securities brokers, as well as to enforce the provisions of the law more generally.

Although financial market regulation deals largely with the problem of managing risk, it cannot eliminate all risk. Risk is inherent in all economic activity. Financial intermediaries such as commercial and investment banks specialize in managing financial risks. Regulation can seek to encourage such institutions to manage risks prudently, but it cannot eliminate the risks inherent in financial intermediation. There is a tension here. Regulators seek to reduce the risks taken on by the firms they regulate. At the same time, however, firms cannot earn profits without taking risks. Thus, an overzealous attempt to reduce risk could prove counterproductive—a firm will not survive if it cannot earn profits.

Conventional wisdom views derivatives markets as markets for risk transfer. According to this view, derivatives markets exist to facilitate the transfer of market risk from firms that wish to avoid such risks to others more willing

[2] See U.S. General Accounting Office (1994).

or better suited to manage those risks. The important thing to note in this regard is that derivatives markets do not create new risks—they just facilitate risk management. Viewed from this perspective, the rapid growth of derivatives markets in recent years simply reflects advances in the technology of risk management. Used properly, derivatives can help organizations reduce financial risk. Although incidents involving large losses receive the most public attention, such incidents are the exception rather than the rule in derivatives markets.

Most public policy concerns center around the speculative use of derivatives. Speculation involves the voluntary assumption of market risk in the hope of realizing a financial gain. The existence of speculation need not concern policymakers as long as all speculative losses are borne privately—that is, only by those individuals or organizations that choose to engage in such activities. But many policymakers fear that large losses on the part of one firm may lead to a widespread disruption of financial markets. The collapse of Barings illustrates some of the foundations for such concerns. In the case of an insured bank, regulators discourage speculation because it can lead to losses that may ultimately become the burden of the government.[3]

A view implicit in many recent calls for more comprehensive regulation of derivatives markets is that these markets are subject to only minimal regulation at present. But exchange-traded derivatives, such as futures contracts, have long been subject to comprehensive government regulation. In the United States, the SEC regulates securities and options exchanges while the Commodity Futures Trading Commission (CFTC) regulates futures exchanges and futures brokers. Although OTC derivatives markets are not regulated by any single federal agency, most OTC dealers, such as commercial banks and brokerage firms, are subject to federal regulation.[4] As it happens, both incidents examined in this article involve instruments traded on regulated exchanges. Any judgment as to whether these incidents indicate a need for more comprehensive regulation of these markets requires some understanding of just what happened in each case.

2. METALLGESELLSCHAFT

Metallgesellschaft AG (hereafter, MG) is a large industrial conglomerate engaged in a wide range of activities, from mining and engineering to trade and financial services. In December 1993, the firm reported huge derivatives-related

[3] Recent losses by firms such as Gibson Greetings and Procter & Gamble have also raised concerns about sales practices and the disclosure of risks associated with complex financial derivatives. Neither of the cases examined in this study involves such concerns, however.

[4] Many securities companies book their OTC derivatives through unregulated subsidiaries. Although these subsidiaries are not subject to formal SEC regulation, the largest brokerage firms have agreed to abide by certain regulatory guidelines and to make regular disclosures to both the SEC and CFTC about their management of derivatives-related risks. See Taylor (1995a).

losses at its U.S. oil subsidiary, Metallgesellschaft Refining and Marketing (MGRM). These losses were later estimated at over $1 billion, the largest derivatives-related losses ever reported by any firm at the time. The incident helped bring MG—then Germany's fourteenth largest industrial corporation—to the brink of bankruptcy. After dismissing the firm's executive chairman, Heinz Schimmelbusch, and several other senior managers, MG's board of supervisors was forced to negotiate a $1.9 billion rescue package with the firm's 120 creditor banks (Roth 1994a, b).

MG's board blamed the firm's problems on lax operational control by senior management, charging that "speculative oil deals . . . had plunged Metallgesellschaft into the crisis."[5] Early press reports on the incident echoed this interpretation of events, but subsequent studies report that MGRM's use of energy derivatives was an integral part of a combined marketing and hedging program under which the firm offered customers long-term price guarantees on deliveries of petroleum products such as gasoline and heating oil. Reports that MG's losses were attributable to a hedging program have raised a host of new questions. Many analysts remain puzzled by the question of how a firm could lose over $1 billion by hedging.

The Metallgesellschaft debacle has sparked a lively debate on the shortcomings of the firm's hedging strategy and the lessons to be learned from the incident. The ensuing account draws from a number of recent articles, notably Culp and Hanke (1994); Culp and Miller (1994a, b, 1995a, b, c, d); Edwards and Canter (1995a, b); and Mello and Parsons (1995a, b).

MGRM's Marketing Program

In 1992, MGRM began implementing an aggressive marketing program in which it offered long-term price guarantees on deliveries of gasoline, heating oil, and diesel fuels for up to five or ten years. This program included several novel contracts, two of which are relevant to this study. The first was a "firm-fixed" program, under which a customer agreed to fixed monthly deliveries at fixed prices. The second, known as the "firm-flexible" contract, specified a fixed price and total volume of future deliveries but gave the customer some flexibility to set the delivery schedule. Under the second program, a customer could request 20 percent of its contracted volume for any one year with 45 days' notice. By September 1993, MGRM had committed to sell forward the equivalent of over 150 million barrels of oil for delivery at fixed prices, with most contracts for terms of ten years.

Both types of contracts included options for early termination. These "cash-out provisions" permitted customers to call for cash settlement on the full volume of outstanding deliveries if market prices for oil rose above the

[5] As cited in Edwards and Canter (1995b), p. 86.

contracted price. The firm-fixed contract permitted a customer to receive one-half the difference between the current nearby futures price (that is, the price of the futures contract closest to expiration) and the contracted delivery price, multiplied by the entire remaining quantity of scheduled deliveries. The firm-flexible contract permitted a customer to receive the full difference between the second-nearest futures price and the contract price, multiplied by all remaining deliverable quantities.[6]

MGRM negotiated most of its contracts in the summer of 1993. Its contracted delivery prices reflected a premium of $3 to $5 per barrel over the prevailing spot price of oil. As is evident in Figure 1, energy prices were relatively low by recent historical standards during this period and were continuing to fall. As long as oil prices kept falling, or at least did not rise appreciably, MGRM stood to make a handsome profit from this marketing arrangement. But a significant increase in energy prices could have exposed the firm to massive losses unless it hedged its exposure.

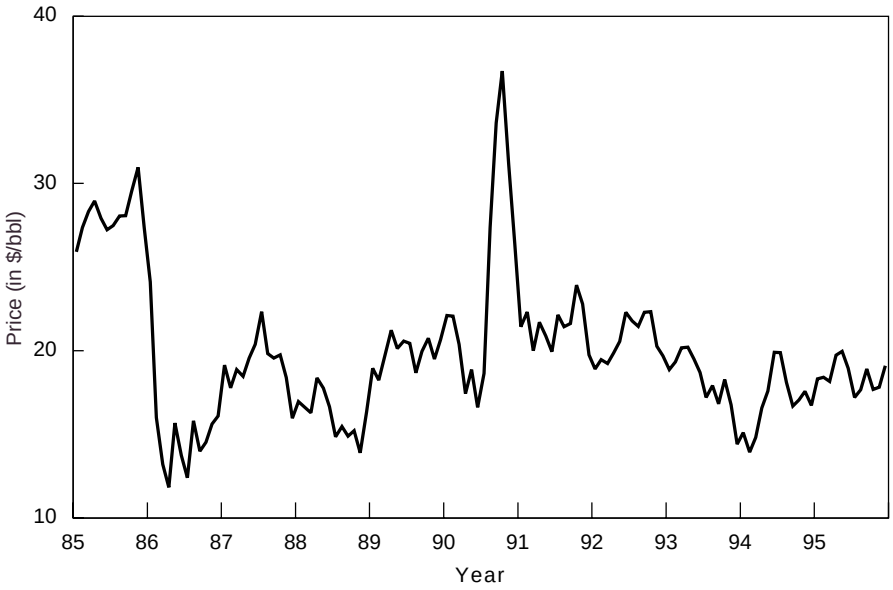
MGRM sought to offset the exposure resulting from its delivery commitments by buying a combination of short-dated oil swaps and futures contracts as part of a strategy known as a "stack-and-roll" hedge. In its simplest form, a stack-and-roll hedge involves repeatedly buying a bundle, or "stack," of short-dated futures or forward contracts to hedge a longer-term exposure. Each stack is rolled over just before expiration by selling the existing contracts while buying another stack of contracts for a more distant delivery date; hence the term stack-and-roll. MGRM implemented its hedging strategy by maintaining long positions in a wide variety of contract months, which it shifted between contracts for different oil products (crude oil, gasoline, and heating oil) in a manner intended to minimize the costs of rolling over its positions.

Had oil prices risen, the accompanying gain in the value of MGRM's hedge would have produced positive cash flows that would have offset losses stemming from its commitments to deliver oil at below-market prices. As it happened, however, oil prices fell even further in late 1993. Moreover, declines in spot and near-term oil futures and forward prices significantly exceeded declines in long-term forward prices. As a result, contemporaneous realized losses from the hedge appeared to exceed any potential offsetting gains accruing to MGRM's long-term forward commitments.

This precipitous decline in oil prices caused funding problems for MGRM. The practice in futures markets of marking futures contracts to market at the end of each trading session forced the firm to recognize its futures trading losses immediately, triggering huge margin calls. Normally, forward contracts have the advantage of permitting hedgers to defer recognition of losses on long-term commitments. But MGRM's stack-and-roll hedge substituted

[6] Mello and Parsons (1995b) provide a detailed description of these contracts.

Figure 1 Crude Oil Prices: 1985–1995



short-term forward contracts (in the form of short-term energy swaps maturing in late 1993) for long-term forward contracts. As these contracts matured, MGRM was forced to make large payments to its counterparties, putting further pressure on its cash flows. At the same time, most offsetting gains on its forward delivery commitments were deferred.

Rumors of MGRM's problems began to surface in early December. In response to these developments, the New York Mercantile Exchange (NYMEX), the exchange on which MGRM had been trading energy futures, raised its margin requirements for the firm. This action, which was intended to protect the exchange in case of a default, further exacerbated MGRM's funding problems. Rumors of the firm's financial difficulties led many of its OTC counterparties to begin terminating their contracts. Others began demanding that it post collateral to secure contract performance.

Upon learning of these circumstances, MG's board of supervisors fired the firm's chief executive and installed new management. The board instructed MG's new managers to begin liquidating MGRM's hedge and to enter into negotiations to cancel its long-term contracts with its customers. This action further complicated matters, however. NYMEX withdrew its hedging exemption once MGRM announced the end of its hedging program. Hedging exemptions permit firms to take on much larger positions in exchange-traded futures than

those allowed for unhedged, speculative positions. The loss of its hedging exemption forced MGRM to reduce its positions in energy futures still further (Culp and Miller 1994b).

The actions of MG's board of supervisors in this incident have spurred widespread debate and criticism, as well as several lawsuits. Some analysts argue that MGRM's hedging program was seriously flawed and that MG's board was right to terminate it. Others, including Nobel Prize-winning economist Merton Miller, argue that the hedging program was sound and that MG's board exacerbated any hedging-related losses by terminating the program prematurely. The discussion that follows reviews the hedging alternatives that were open to the firm, the risks associated with the strategy it chose, and critiques of that strategy offered by a number of economists.

Hedging Alternatives

In common usage, the term "hedging" refers to an attempt to avoid the risk of loss by matching a given risk exposure with a counterbalancing risk, as in hedging a bet. Elementary finance textbooks are replete with examples of perfect hedges, wherein a firm uses futures or forward contracts to offset perfectly some given exposure. Hedging strategies employed by firms tend to be somewhat more complex, however. In practice, a perfect hedge can be difficult to arrange. And even when feasible, such a strategy often leaves little room for profit.

Edwards and Canter (1995a, b) note that MGRM had at least three hedging options open to it: physical storage, long-dated forward contracts, and some variant of a stack-and-roll strategy. Physical storage would have required MGRM to purchase the oil products it had committed itself to deliver in the future and then store those products until the promised delivery dates. Physical storage would have been expensive, however. First, it would have required MGRM to finance the cost of the required inventories. Second, it would have entailed the cost of the requisite storage facilities. Together, these two costs comprise what is known as the cost of carry. Available evidence suggests that the costs associated with physical storage would have rendered MGRM's marketing program unprofitable.[7]

Alternatively, MGRM could have chosen among a number of derivatives-based hedging strategies involving either futures or forward contracts, or some combination of both. Putting together a perfect hedge using such instruments would have required the purchase of a bundle of oil futures or forward contracts with expiration dates just matching MGRM's promised delivery dates. But oil futures typically trade only for maturities of three years or less. Moreover, liquidity tends to be poor for contracts with maturities over 18 months. Thus, MGRM would have had to buy a bundle of long-dated forward contracts from

[7] See Edwards and Canter (1995a, b), and Mello and Parsons (1995b).

an OTC derivatives dealer to put together a hedge that just offset its exposure to long-term energy prices.

Like physical storage, however, the cost associated with buying a bundle of long-dated forward contracts probably would have been prohibitive. To understand why, note that buying a futures or forward contract is equivalent to physical storage in the sense that both strategies ensure the future availability of an item at some predetermined cost. For this reason, the strategy of buying futures or forward contracts to lock in the cost of future delivery is sometimes termed "synthetic storage." Accordingly, Arbitrage Pricing Theory predicts that the forward price for a commodity should reflect its cost of carry. Based on the factors considered to this point, then, the theoretical no-arbitrage or benchmark forward price should be

$$\text{THEORETICAL FORWARD PRICE} = \text{SPOT PRICE} + \text{COST OF CARRY}.$$

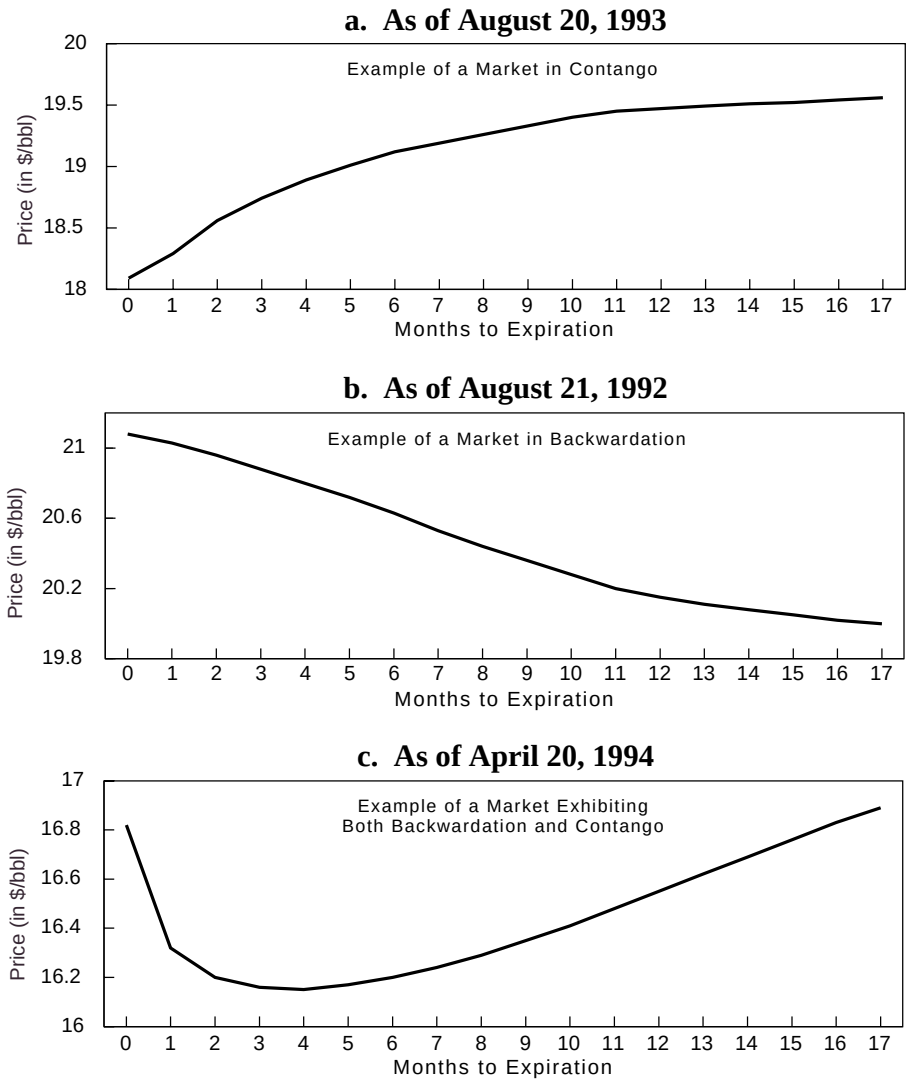
Notice that this relationship implies that buyers of futures and forward contracts should pay a premium for deferred delivery. This premium is known as contango in the parlance of futures markets. Figure 2a, which shows the term structure of crude oil futures prices as of August 20, 1993, provides an example of a contango market.

These theoretical considerations suggest that futures prices should always exhibit contango. As is evident in Figure 2b, however, they do not always do so. Arbitrage ensures that the forward price of a commodity can never exceed the theoretical benchmark price, but it evidently does not prevent futures and forward prices from falling below this theoretical benchmark. To understand why, consider the opportunities for arbitrage that would arise if futures prices exceeded the benchmark forward price derived above. In this case, an arbitrageur could earn riskless profits by buying and storing the commodity in question while selling it forward at a price exceeding the purchase price plus the cost of carry. Futures prices can fail to reflect a commodity's full cost of carry if firms place a premium on current availability, however, as they sometimes do when available supplies of the commodity are scarce. In such cases there is said to be a convenience yield associated with physical storage. The simple cost-of-carry price relation presented above fails to take account of convenience yields, but it does suggest a way to measure them. The convenience yield for an item can be measured by computing the difference between the benchmark forward price (the sum of the current spot price and the cost of carry) and the prevailing market-determined forward price.

$$\begin{gathered}\text{CONVENIENCE YIELD} = \\ \text{SPOT PRICE} + \text{COST OF CARRY} - \text{ACTUAL FORWARD PRICE}\end{gathered}$$

Sometimes the convenience yield is high enough to offset the cost of carry, causing forward prices to be lower than spot prices, as in Figure 2b. This latter phenomenon is known as backwardation.

Figure 2 Term Structure of Crude Oil Futures Prices



When it exists, backwardation in commodity markets tends to be evident only in short-term futures and forward prices. Carrying costs increase with time to delivery so that longer-term futures or forward contracts typically sell at a premium even when prices for short-dated contracts exhibit backwardation. Figure 2b shows that the pattern of backwardation extended out to at least 18

months as of August 21, 1992. At other times, however, futures prices begin increasing at shorter horizons. Figure 2c shows the term structure of crude oil futures prices as of April 20, 1994. On this latter date, futures prices exhibited backwardation only for the first four delivery months and then began rising.

The foregoing discussion shows that a hedging strategy based on long-term forward contracts can be almost as expensive as physical storage, even when short-term futures and forward prices exhibit backwardation. So although MGRM could have hedged its exposure by buying long-term forward contracts from an OTC derivatives dealer, doing so would have reduced, if not eliminated, any profits from its marketing program. Moreover, any dealer selling such contracts would have faced similar hedging problems.

A stack-and-roll strategy appeared to offer a means of avoiding such carrying costs because short-dated futures markets for oil products historically have tended to exhibit backwardation. In markets that exhibit persistent backwardation, a strategy of rolling over a stack of expiring contracts every month can generate profits. Thus, MGRM's management apparently thought that a stack-and-roll hedging strategy offered a cost-effective means of locking in a spread between current spot prices and the long-term price guarantees it had sold to its customers. As noted earlier, however, this strategy was not without risks. These risks are examined in more detail below.

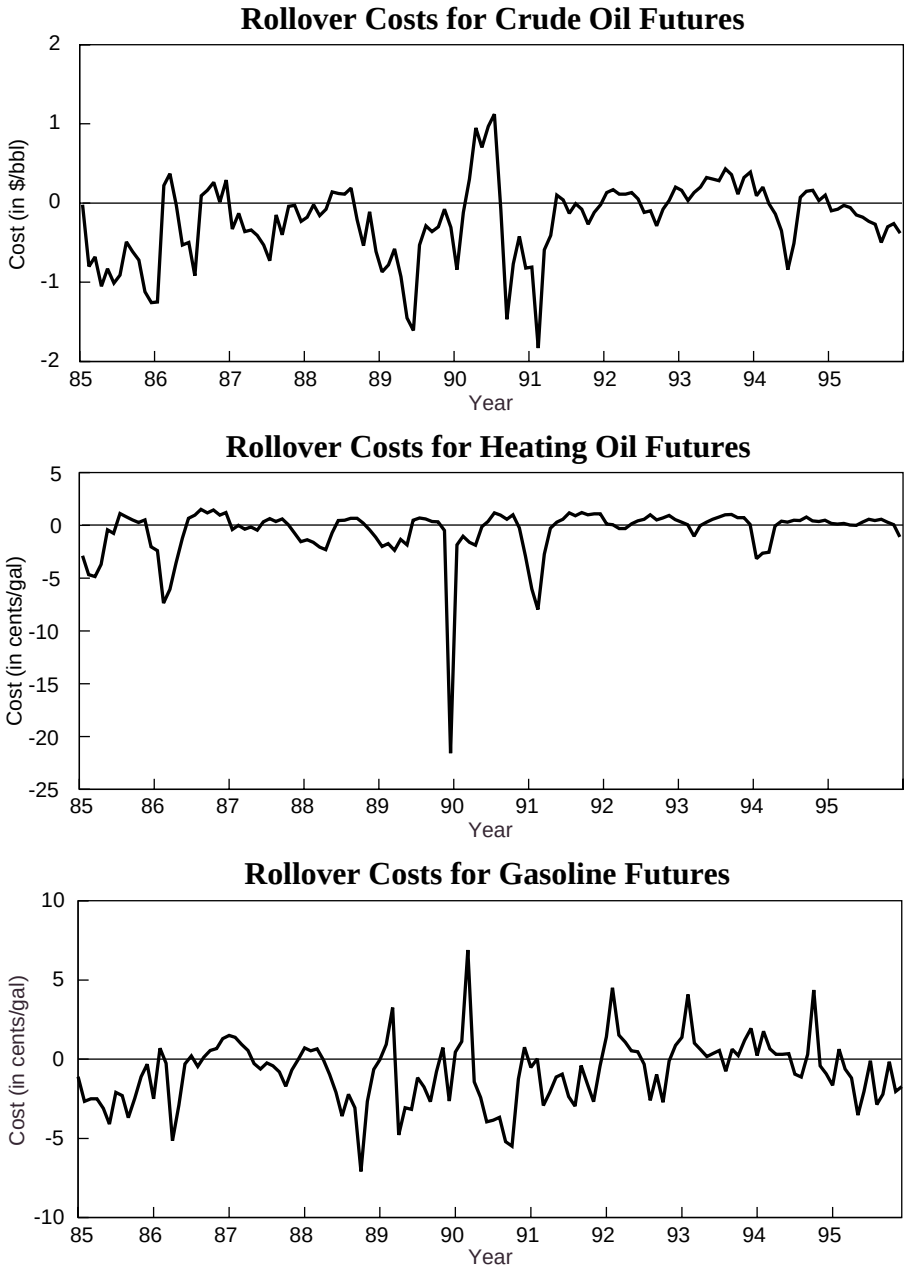
Basis Risk

The term "basis" refers to the difference between the spot price of an item and its futures price. Basis changes over the life of a contract, usually for fundamental economic reasons but sometimes for reasons that are not well understood. MGRM's stack-and-roll hedging strategy exposed it to basis risk—the risk that the price behavior of its stack of short-dated oil contracts might diverge from that of its long-term forward commitments. As it happened, the behavior of energy futures prices became most unusual in 1993—in that short-term energy futures exhibited a pattern of contango rather than backwardation for most of the year. Once near-dated energy futures and forward markets began to exhibit contango, MGRM was forced to pay a premium to roll over each stack of short-term contracts as they expired. These rollover costs reflected the cost of carry normally associated with physical storage.

This shift is evident in Figure 3, which shows the behavior of rollover costs for three different energy futures contracts (crude oil, heating oil, and gasoline) from 1985 through the end of 1995.[8] As is evident from these figures, rollover costs were positive for most of 1993. The expected profitability of MGRM's

[8] Here, rollover costs are measured by the difference between the closing price of the nearby futures contract three days before the contract expiration date and the price of the next more distant futures contract, as in Edwards and Canter (1995a).

Figure 3



Note: Rollover costs are measured as of three days before contract expiration.

combined marketing and hedging program was predicated on the assumption that energy futures markets would continue to exhibit a pattern of backwardation, however. MG's board of supervisors apparently feared that the need to pay these rollover costs could add further to MGRM's losses and chose to liquidate the subsidiary's hedge and terminate its long-term delivery contracts with its customers.

Critiques of MGRM's Hedging Program

As Figure 1 shows, oil prices began rising in 1994, soon after MGRM's new management lifted the firm's hedge. It thus appears that MGRM could have recouped most if not all of its losses simply by sticking to its hedging program. Whether management should have been able to anticipate this outcome is the topic of an active debate, however.

Criticisms of MGRM's hedging program have focused on two issues. The first deals with the assumptions the architects of MGRM's hedging strategy made regarding the likely future behavior of basis in oil futures and forward markets. The second concerns the steps MGRM could have taken to reduce the variability of its cash flows.

Both Edwards and Canter (1995a) and Mello and Parsons (1995b) show that MGRM's hedging program would have generated huge losses if contango energy markets had persisted throughout 1994. A key question, then, is whether MG's board of supervisors should have viewed the behavior of energy futures prices during 1993 as a temporary aberration, or whether it had reasonable grounds to believe that this price behavior could persist indefinitely.

Edwards and Canter conclude that permanent changes in the behavior of basis are possible and have occurred in other futures markets. As evidence, they cite experience with two other commodity futures contracts: soybeans and copper. Both markets were characterized by backwardation from 1965 to 1975, but then began exhibiting persistent contango. Thus, while a stack-and-roll hedging strategy for either commodity would have produced positive cash flows on average before 1975, such a strategy would have lost money consistently over the ensuing ten-year period—meaning that a hedger employing a stack-and-roll strategy of the type used by MGRM in either soybean or copper futures markets would have experienced large and persistent losses after 1975.

Along with Mello and Parsons (1995a, b), Edwards and Canter (1995a, b) argue that MGRM was overhedged because short-term oil futures prices tend to be much more volatile than prices on long-term forward contracts. According to these authors, MGRM's managers could have—and should have—designed a hedge that would have reduced the variability of the firm's short-term cash flows. Edwards and Canter find that the correlation of short-term energy futures and forward prices with long-term prices is approximately 50 percent. Thus, they argue that MGRM could have minimized the variance of

its cash flows with a hedge approximately 50 percent smaller than the total of its future delivery commitments.[9] Mello and Parsons observe that the exact size of a minimum-variance hedge is difficult to calculate because MGRM's contracts gave its customers options to terminate their contracts after three years. They find that the minimum-variance hedge ratio could be as high as 75 percent if one assumes that all such options would be exercised at the end of three years.

While critical of certain aspects of MGRM's hedging strategy, Edwards and Canter are agnostic as to whether MG's board was correct to terminate its U.S. subsidiary's oil-hedging program.[10] Mello and Parsons (1995a, b) are more critical of MGRM's hedging strategy, arguing that it was speculative in its design and intent. They base their views on a written strategic plan prepared by MGRM's management, which outlined a plan to exploit backwardation in futures markets as part of its hedging program. Where the plan went wrong, according to Mello and Parsons, was in assuming that the firm could take advantage of backwardation to price its long-term customer contracts below the full cost of carry. They conclude that viewing MGRM's stack-and-roll strategy as a hedge reverses the order of cause and effect, arguing that it should be viewed as a misguided speculative attempt to profit from the backwardation normally present in futures markets for petroleum products while using forward delivery contracts as a partial hedge.

A Defense of MGRM's Hedging Strategy

Culp and Miller (1994a, b, 1995a, b, c, d) and Culp and Hanke (1994) are critical of MG's board of supervisors for terminating MGRM's marketing and hedging program. These authors argue that MGRM's hedging strategy was sound and that the firm's losses are attributable primarily to the way the board terminated the program.

While acknowledging that the volatility of short-term oil prices did make MGRM's cash flows volatile, Culp and Miller argue that short-term cash flow volatility is irrelevant to judgments about the efficacy of MGRM's hedging program. They base this argument on two considerations. The first stems from a theoretical analysis of the properties of a stack-and-roll hedge, the second from a practical analysis of MG's ability to continue funding the program.

First, Culp and Miller (1994b, 1995d) demonstrate that a stack-and-roll hedge of the type employed by MGRM will offset perfectly any changes in the

[9] A 50 percent hedge ratio does not take into account that changes in the value of a long-term forward contract will not be realized for many years, however. The procedure for doing so is known as "tailing the hedge" (see Kawaller [1986] for a description). Tailing the hedge lowers the recommended hedge ratio even further. Edwards and Canter estimate that MGRM could have minimized the variance of its cash flows by buying short-term futures contracts for 61 million barrels of oil to hedge a 160 million barrel long-term exposure.

[10] In a more recent article, however, Edwards (1995) is somewhat more critical of the decision to liquidate MGRM's forward delivery contracts.

value of a long-term forward commitment so long as the factors determining basis—interest rates, storage costs, and the implicit convenience yield associated with physical storage—do not change. Thus, according to Culp and Miller, it is misleading to blame MG's losses on changes in the term structure of oil prices. While short-term price volatility can make cash flows volatile, it does not affect the net present value of the hedged exposure as long as basis remains unchanged.

As noted earlier, however, the behavior of basis did change in the summer of 1993. Culp and Miller acknowledge that MGRM's hedging strategy exposed the firm to basis risk, but they argue that this risk was relatively small considering the historical behavior of energy futures prices. Their analysis shows that changes in basis affect only the portion of carrying costs borne by the hedger. The hedger bears no carrying costs as long as the convenience yield is greater than or equal to the cost of carry—that is, when the market exhibits backwardation—but must bear at least some portion of carrying costs in a contango market. These carrying costs appear as rollover costs.

No one has attempted to refute Culp and Miller's theoretical results. Rather, other authors question the presumption that oil markets would always tend to exhibit backwardation, whereas Culp and Miller argue that any long-run expected losses due to basis risk were minimal considering historical patterns of backwardation in energy markets.

At first glance, the results of Culp and Miller's analysis appear difficult to reconcile with the $1.3 billion loss auditors later attributed to MGRM's marketing and hedging program. Culp and Miller (1995c) take issue with this estimate, however, arguing that MG's auditors underestimated the value of MGRM's contracts with its customers. They argue that taking proper account of unrealized gains in the value of such contracts results in a net loss of $170 million rather than $1.3 billion. According to Culp and Miller, most of MG's reported losses were attributable to the manner in which its new management chose to terminate its subsidiary's marketing program, not to defects in its hedging strategy. It is not unusual for the parties to such agreements to negotiate termination of a contract before it expires. The normal practice in such circumstances involves payment by one party to the other to compensate for any changes in the value of the contract. In contrast, it appears that MGRM's new management simply agreed to terminate its contracts with its customers without asking for any payment to reflect changes in the value of those contracts. The hedge—however imperfect—effectively was transformed by this action into a huge speculative transaction after the fact.

Although Culp and Miller do find that MGRM's hedging program had suffered losses (albeit much smaller losses than those calculated by MG's auditors), they argue that those losses did not justify terminating MGRM's hedging program. First, they emphasize that any past losses were sunk costs. At the same time, they find that the program had a positive expected net present

value at the end of 1993.[11] Thus, they argue that the firm had good reason to continue the program. Culp and Miller reject the board's argument that terminating MGRM's hedge was the only way of dealing with the subsidiary's massive cash outflows. They note that the firm could have bought options to remain hedged while it sought solutions to its longer-term funding problems. Moreover, they argue that short-term cash flow constraints should not have presented any insurmountable problems in view of MG's long-standing and close relations with Deutsche Bank, Germany's largest commercial bank. They emphasize that Deutsche Bank was not only a creditor to MG but also one of its largest shareholders. In addition, a Deutsche Bank executive, Ronaldo Schmitz, was chairman of MG's board of supervisors at the time. Accordingly, Culp and Miller conclude that the Deutsche Bank should have been willing to continue financing MGRM's hedge in view of its close relations with MG and its expertise in finance.

At the very least, Culp and Miller suggest, MG's management could have bought options to hedge its oil exposure while seeking a longer-term solution to its funding problems, as suggested by MGRM's management. As a longer-term solution, they argue that the firm could have spun off the combined marketing and hedging program into a separate subsidiary, which could have been sold to another firm. This argument is supported by Edwards (1995), who reports that at least one major U.S. bank had offered to provide secured financing to MGRM based on a plan to securitize its forward delivery contracts.

Besides taking issue with the actions of MG's board of supervisors, Culp and Hanke (1994) fault NYMEX for the actions the exchange took against MGRM. They argue that these actions needlessly exacerbated MGRM's temporary cash flow problems and thereby helped to precipitate a funding crisis for the firm.

Reconciling Opposing Views

Disagreements over the efficacy of MGRM's hedging program stem from differing assumptions about (1) the goal of the hedging program (or, perhaps more accurately, what the goal should have been), and (2) the feasibility of continuing the program in light of the large negative cash flows MGRM experienced in late 1993. Both Edwards and Canter (1995a, b) and Mello and Parsons (1995a, b) emphasize the difficulties that the large negative cash flows produced by the hedging program caused the parent company. These authors argue that MGRM's management should have sought to avoid such difficulties by designing a hedge that would have minimized the volatility of its cash flows.

Although they are critical of MGRM's hedging strategy, Edwards and Canter offer no opinion as to whether MG's board was right to terminate

[11] Note, however, that this estimate is based on the assumption that expected carrying costs would be zero over the long run. See Culp and Miller (1995a, b, d).

the program. Like Culp and Miller, they are puzzled about the decision to terminate existing contracts with customers without negotiating some payment to compensate for the increase in the value of those contracts.

Mello and Parsons's criticisms of MGRM's hedging strategy are unequivocal. They argue that MGRM's strategy was fatally flawed, and they defend the decision to terminate the hedging program as the only means of limiting even greater potential future losses. They also emphasize the difficulty that MG's new management would have had in securing the financing necessary to maintain MGRM's hedging program and argue that funding considerations should have led the subsidiary's managers to synthesize a hedge using long-dated forward contracts. In this context, Mello and Parsons note that the parent firm already had accumulated a cash flow deficit of DM 5.65 billion between 1988 and 1993. This deficit had been financed largely by bank loans. Considering these circumstances, they find the reluctance of MG's creditor banks to fund the continued operation of the oil marketing program understandable.

Culp and Miller accept that MGRM's hedge was intended to exploit the backwardation normally present in energy futures markets, but they reject the argument that its hedging program represented reckless speculation. They emphasize that few, if any, commodity dealers always hedge away all risks, citing the results of previous studies on the behavior of commodity dealers to support their assertions (Culp and Miller 1995a, b). Thus, they conclude that short-term cash flow constraints should not have presented any insurmountable problems in view of MG's long-standing and close relations with Deutsche Bank, which they feel should have been willing to continue financing MGRM's hedging program.

These disagreements over the efficacy of MGRM's hedging strategy seem unlikely ever to be resolved, based as they are on different assumptions about the goals management should have had for its strategy. The main issue, then, is whether MG's senior management and board of supervisors fully appreciated the risks the firm's U.S. oil subsidiary had assumed. If they did, the firm should have arranged for a line of credit to fund its short-term cash flows. Indeed, Culp and Miller (1995a, b, d) claim that MGRM had secured lines of credit with its banks just to prepare for such contingencies. Yet the subsequent behavior of MG's board suggests that its members had very little prior knowledge of MGRM's marketing program and were uncomfortable with its hedging strategy, despite the existence of a written strategic plan.

It is difficult for an outside observer to assign responsibility for any misunderstandings between MG's managers and its board of supervisors. MG's board ultimately held Heinz Schimmelbusch, the firm's executive chairman, responsible for the firm's losses, claiming that he and other senior managers had lost control over the activities of the firm and concealed evidence of losses.[12] In

[12] See *The Wall Street Journal* (1993) and *The Economist* (1993).

response, Schimmelbusch has filed suit against Ronaldo Schmitz and Deutsche Bank, seeking $10 million in general and punitive damages (Taylor 1995b). Arthur Benson, former head of MGRM and architect of the firm's ill-fated hedging program, is suing MG's board for $1 billion on charges of defamation (Taylor 1994). Thus, the issue of blame appears destined to be settled by the U.S. courts.

Response of the CFTC

The Metallgesellschaft debacle did not escape the attention of U.S. regulators. In July 1995, the U.S. Commodity Futures Trading Commission (CFTC) instituted administrative proceedings against MGRM and MG Futures, Inc. (MGFI), an affiliated Futures Commission Merchant that processed trades for MGRM and other MG subsidiaries.[13] The CFTC order charged both MGRM and MGFI with "material inadequacies in internal control systems" associated with MGRM's activity in energy and futures markets. In addition, MGFI was charged with failing to inform the CFTC of these material inadequacies, while MGRM was charged with selling illegal, off-exchange futures contracts. The two MG subsidiaries settled the CFTC action without admitting or denying the charges and agreed to pay the CFTC a $2.5 million settlement. They also agreed to implement a series of CFTC recommendations to reform their internal controls and to refrain from violating CFTC regulations. The CFTC's action rendered MGRM's firm-fixed agreements "illegal and void."[14] Thus, the CFTC's action would have created legal risk for Metallgesellschaft and its customers except that the firm had already canceled most of the contracts in question.

The CFTC's actions in this case have proven somewhat controversial. Under the Commodity Exchange Act, the CFTC is charged with regulating exchange-traded futures contracts. At the same time, the act explicitly excludes ordinary commercial forward contracts from the jurisdiction of the CFTC. The legal definition of a futures contract is open to differing interpretations, however, leading to some uncertainty over the legal status of OTC derivatives under the Commodity Exchange Act. Most market participants felt that this uncertainty was resolved in 1993 when, at the behest of Congress, the CFTC agreed to exempt off-exchange forward and swaps contracts from regulations governing exchange-traded contracts. CFTC chairman Mary Schapiro maintains that the agency's action against MGRM does not represent a reversal of its policy on OTC contracts. According to Schapiro, the CFTC's order is worded narrowly so as to apply only to contracts such as the firm-fixed (45-day) agreements sold

[13] A Futures Commission Merchant is a broker that accepts and executes orders for transactions on futures exchanges for customers. Futures Commission Merchants are regulated by the CFTC.

[14] See U.S. Commodity Futures Trading Commission (1995a, b).

by MGRM in this case.[15] Nonetheless, this action has prompted some critics to charge the agency with creating uncertainty about the legal status of commercial forward contracts. Critics of the action include Miller and Culp (1995) and Wendy Gramm, a former chairman of the CFTC.[16] The CFTC's action has also been criticized by at least two prominent members of Congress—Rep. Thomas J. Bliley, Jr., Chairman of the House Commerce Committee; and Rep. Pat Roberts, Chairman of the House Agricultural Committee.[17]

Since the CFTC's action against Metallgesellschaft is narrowly directed and involves somewhat esoteric legal arguments, it is too soon to know what its effect will be on OTC derivatives markets generally. Still, commodity dealers must now take extra care in designing long-term delivery contracts to avoid potential legal problems.[18]

An Overview of Policy Concerns

Considering the debate over the merits of MGRM's hedging strategy, it would seem naive simply to blame the firm's problems on its speculative use of derivatives. It is true that MGRM's hedging program was not without risks. But the firm's losses are attributable more to operational risk—the risk of loss caused by inadequate systems and control or management failure—than to market risk. If MG's supervisory board is to be believed, the firm's previous management lost control of the firm and then acted to conceal its losses from board members. If one sides with the firm's previous managers (as well as with Culp, Hanke, and Miller), then the supervisory board and its bankers misjudged the risks associated with MGRM's hedging program and panicked when faced with large, short-term funding demands. Either way, the loss was attributable to poor management.

Does this episode indicate the need for new government policies or more comprehensive regulation of derivatives markets? The answer appears to be no. MGRM's losses do not appear ever to have threatened the stability of financial markets. Moreover, those losses were due in large part to the firm's use of futures contracts, which trade in a market that is already subject to comprehensive regulation. The actions taken by the CFTC in this instance demonstrate clearly that U.S. regulators already have the authority to intervene when they deem it necessary. Unfortunately, the nature of those actions in this case may create added legal risk for other market participants.

To view the entire incident in its proper perspective, it must be remembered that MG's losses were incurred in connection with a marketing program

[15] See *BNA's Banking Report* (1995a).

[16] For a summary of Gramm's comments see *The Wall Street Journal* (1995).

[17] See Fox (1995).

[18] See Rance (1995) for a legal analysis of these issues.

aimed at providing long-term, fixed-price delivery contracts to customers—a type of arrangement common to many types of commercial activity. Systematic attempts to discourage such arrangements would seem to be poor public policy.

Finally, MG's financial difficulties were not attributable solely to its use of derivatives. As noted earlier, the firm's troubles stemmed in part from the heavy debt load it had accumulated in previous years. Moreover, MGRM's oil marketing program was not the only source of its parent company's losses during 1993. MG reported losses of DM 1.8 billion on its operations for the fiscal year ended September 30, 1993, in addition to the DM 1.5 billion loss auditors attributed to its hedging program as of the same date (Roth 1994a). Simply stated, the MG debacle resulted from poor management. As a practical matter, government policy cannot prevent firms such as Metallgesellschaft from making mistakes. Nor should it attempt to do so.

3. BARINGS

At the time of its demise in February 1995, Barings PLC was the oldest merchant bank in Great Britain. Founded in 1762 by the sons of German immigrants, the bank had a long and distinguished history. Barings had helped a fledgling United States of America arrange the financing of the Louisiana Purchase in 1803. It had also helped Britain finance the Napoleonic Wars, a feat that prompted the British government to bestow five noble titles on the Baring family.

Although it was once the largest merchant bank in Britain, Barings was no longer the powerhouse it had been in the nineteenth century. With total shareholder equity of £440 million, it was far from the largest or most important banking organization in Great Britain. Nonetheless, it continued to rank among the nation's most prestigious institutions. Its clients included the Queen of England and other members of the royal family.

Barings had long enjoyed a reputation as a conservatively run institution. But that reputation was shattered on February 24, 1995, when Peter Baring, the bank's chairman, contacted the Bank of England to explain that a trader in the firm's Singapore futures subsidiary had lost huge sums of money speculating on Nikkei-225 stock index futures and options. In the days that followed, investigators found that the bank's total losses exceeded US$1 billion, a sum large enough to bankrupt the institution.

Barings had almost failed once before in 1890 after losing millions in loans to Argentina, but it was rescued on that occasion by a consortium led by the Bank of England. A similar effort was mounted in February 1995, but the attempt failed when no immediate buyer could be found and the Bank of England refused to assume liability for Barings's losses. On the evening of Sunday, February 26, the Bank of England took action to place Barings into

administration, a legal proceeding resembling Chapter 11 bankruptcy-court proceedings in the United States. The crisis brought about by Barings's insolvency ended just over one week later when a large Dutch financial conglomerate, the Internationale Nederlanden Groep (ING), assumed the assets and liabilities of the failed merchant bank.

What has shocked most observers is that such a highly regarded institution could fall victim to such a fate. The ensuing account examines the events leading up to the failure of Barings, the factors responsible for the debacle, and the repercussions of that event on world financial markets.[19] This account is followed by an examination of the policy concerns arising from the episode and the lessons these events hold for market participants and policymakers.

Unauthorized Trading Activities

In 1992, Barings sent Nicholas Leeson, a clerk from its London office, to manage the back-office accounting and settlement operations at its Singapore futures subsidiary. Baring Futures (Singapore), hereafter BFS, was established to enable Barings to execute trades on the Singapore International Monetary Exchange (SIMEX). The subsidiary's profits were expected to come primarily from brokerage commissions for trades executed on behalf of customers and other Barings subsidiaries.[20]

Soon after arriving in Singapore, Leeson asked permission to take the SIMEX examinations that would permit him to trade on the floor of the exchange. He passed the examinations and began trading later that year. Some time during late 1992 or early 1993, Leeson was named general manager and head trader of BFS. Normally the functions of trading and settlements are kept separate within an organization, as the head of settlements is expected to provide independent verification of records of trading activity. But Leeson was never relieved of his authority over the subsidiary's back-office operations when his responsibilities were expanded to include trading.

[19] This account is based on the findings of a report by the Board of Banking Supervision of the Bank of England (1995) and on a number of press accounts dealing with the episode. Except where otherwise noted, all information on this episode was taken from the Board of Banking Supervision's published inquiry.

[20] Most of BFS's business was concentrated in executing trades for a limited number of financial futures and options contracts. These were the Nikkei-225 contract, the 10-year Japanese Government Bond (JGB) contract, the three-month Euroyen contract, and options on those contracts (known as futures options). The Nikkei-225 contract is a futures contract whose value is based on the Nikkei-225 stock index, an index of the aggregate value of the stocks of 225 of the largest corporations in Japan. The JGB contract is for the future delivery of ten-year Japanese government bonds. The Euroyen contract is a futures contract whose value is determined by changes in the three-month Euroyen deposit rate. A futures option is a contract that gives the buyer the right, but not the obligation, to buy or sell a futures contract at a stipulated price on or before some specified expiration date.

Leeson soon began to engage in proprietary trading—that is, trading for the firm's own account. Barings's management understood that such trading involved arbitrage in Nikkei-225 stock index futures and 10-year Japanese Government Bond (JGB) futures. Both contracts trade on SIMEX and the Osaka Securities Exchange (OSE). At times price discrepancies can develop between the same contract on different exchanges, leaving room for an arbitrageur to earn profits by buying the lower-priced contract on one exchange while selling the higher-priced contract on the other. In theory this type of arbitrage involves only perfectly hedged positions, and so it is commonly regarded as a low-risk activity. Unbeknownst to the bank's management, however, Leeson soon embarked upon a much riskier trading strategy. Rather than engaging in arbitrage, as Barings management believed, he began placing bets on the direction of price movements on the Tokyo stock exchange.

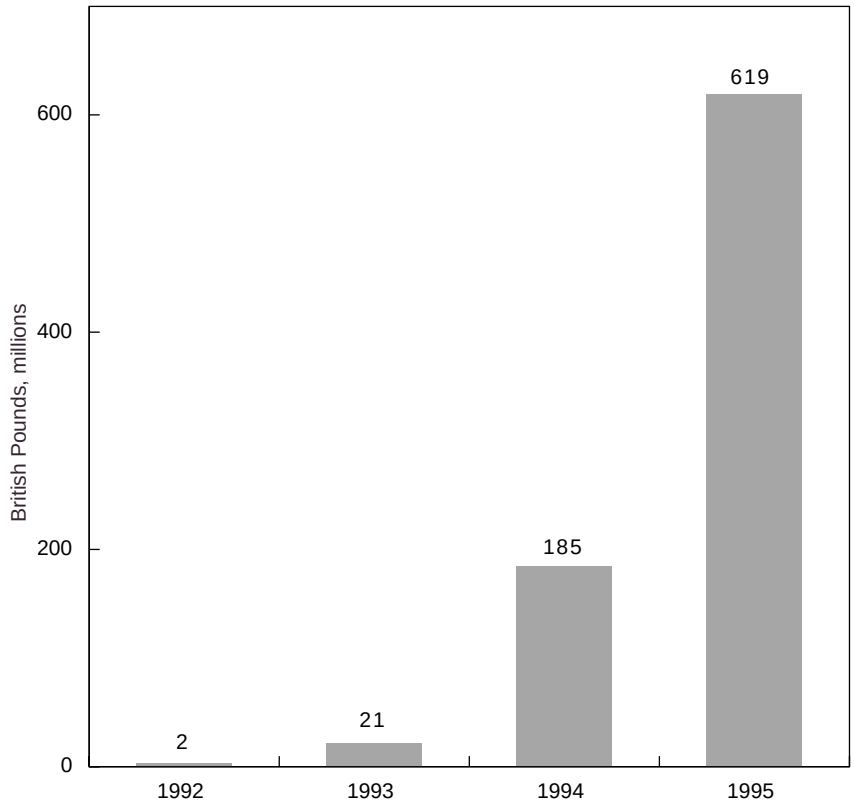
Leeson's reported trading profits were spectacular. His earnings soon came to account for a significant share of Barings total profits; the bank's senior management regarded him as a star performer. After Barings failed, however, investigators found that Leeson's reported profits had been fictitious from the start. Because his duties included supervision of both trading and settlements for the Singapore subsidiary, Leeson was able to manufacture fictitious reports concerning his trading activities. He had set up a special account—account number 88888—in July 1992, and instructed his clerks to omit information on that account from their reports to the London head office. By manipulating information on his trading activity, Leeson was able to conceal his trading losses and report large profits instead.

Figure 4 shows Leeson's trading losses from 1992 through the end of February 1995. By the end of 1992—just a few months after he had begun trading—Leeson had accumulated a hidden loss of £2 million. That figure remained unchanged until October 1993, when his losses began to rise sharply. He lost another £21 million in 1993 and £185 million in 1994. Total cumulative losses at the end of 1994 stood at £208 million. That amount was slightly larger than the £205 million profit reported by the Barings Group as a whole, before accounting for taxes and for £102 million in scheduled bonuses.

A major part of Leeson's trading strategy involved the sale of options on Nikkei-225 futures contracts. Figures 5a and 5b show the payoff at expiration accruing to the seller of a call or put option, respectively. The seller of an option earns a premium in return for accepting the obligation to buy or sell the underlying item at a stipulated strike price. If the option expires "out-of-the-money," the option premium becomes the seller's profit. If prices turn out to be more volatile than expected, however, an option seller's potential losses are virtually unlimited.

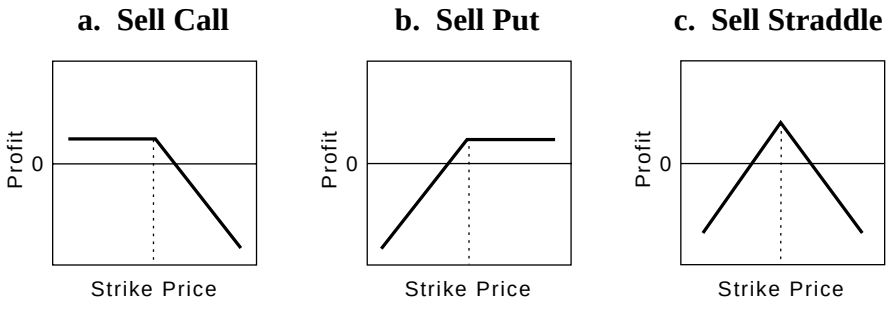
Some time in 1994, Leeson began selling large numbers of option straddles, a strategy that involved the simultaneous sale of both calls and puts on Nikkei-225 futures. Figure 5c shows the payoff at expiration to a sold option

Figure 4 Concealed Trading Losses



Source: Bank of England, Board of Banking Supervision

Figure 5 Payoffs to Selected Options Trading Strategies



straddle. Option prices reflect the market's expectation of the price volatility of the underlying item. The seller of an option straddle earns a profit only if the market proves less volatile than predicted by option prices. As is evident in Figure 5c, Leeson's strategy amounted to a bet that the Japanese stock market would neither fall nor increase by a great deal—any large movement in Japanese stock prices would result in losses. By January 1, 1995, Leeson was short 37,925 Nikkei calls and 32,967 Nikkei puts. He also held a long position of just over 1,000 contracts in Nikkei stock index futures, which would gain in value if the stock market were to rise.

Disaster struck on January 17 when news of a violent earthquake in Kobe, Japan, sent the Japanese stock market into a tailspin. Over the next five days, the Nikkei index fell over 1,500 points—Leeson's options positions sustained a loss of £68 million. As stock prices fell, he began buying massive amounts of Nikkei stock index futures. He also placed a side bet on Japanese interest rates, selling Japanese government bond futures by the thousands in the expectation of rising interest rates.

This strategy seemed to work for a short time. By February 6, the Japanese stock market had recovered by over 1,000 points, making it possible for Leeson to recoup most of the losses resulting from the market's reaction to the earthquake. His cumulative losses on that date totaled £253 million, about 20 percent higher than they had been at the start of the year. But within days the market began falling again—Leeson's losses began to multiply. He continued to increase his exposure as the market kept falling. By February 23, Leeson had bought over 61,000 Nikkei futures contracts, representing 49 percent of total open interest in the March 1995 Nikkei futures contract and 24 percent of the open interest in the June contract. His position in Japanese government bond futures totaled just over 26,000 contracts sold, representing 88 percent of the open interest in the June 1995 contract. Leeson also took on positions in Euroyen futures. He began 1995 with long positions in Euroyen contracts (a bet that Japanese interest rates would fall) but then switched to selling the contracts. By February 23 he had accumulated a short position in Euroyen futures equivalent to 5 percent of the open interest in the June 1995 contract and 1 percent of the open interest in both the September and December contracts.

Barings faced massive margin calls as Leeson's losses mounted. While these margin calls raised eyebrows at the bank's London and Tokyo offices, they did not prompt an immediate inquiry into Leeson's activities. It was not until February 6 that Barings's group treasurer, Tony Hawes, flew to Singapore to investigate irregularities with the accounts at BFS. Accompanying Hawes was Tony Railton, a settlements clerk from the London office.

While in Singapore, Hawes met with SIMEX officials, who had expressed concern over Barings's extraordinarily large positions. Hawes assured them that his firm was aware of these positions and stood ready to meet its obligations

to the exchange. His assurances were predicated on the belief that the firm's exposure on the Singapore exchange had been hedged with offsetting positions on the Osaka exchange. He was soon to learn that this belief was incorrect.

Leeson's requests for additional funding continued during February, and Barings's London office continued to meet those requests—in all, Barings had committed a total of £742 million to finance margin calls for BFS. Meanwhile, Tony Railton, the clerk Hawes had dispatched to Singapore, found that he could not reconcile the accounts of BFS. Particularly disturbing was a US$190 million discrepancy in one of BFS's accounts. For over a week, Railton attempted to meet with Leeson to resolve these discrepancies. Leeson had become hard to find, however. Railton finally tracked him down on the floor of the Singapore exchange on Thursday, February 23, and persuaded Leeson to meet with him that evening. When the meeting began, Railton began asking a series of difficult questions. At that point Leeson excused himself, stating that he would return shortly. But he never did return. Instead, he and his wife left Singapore that evening. The next day, Leeson faxed his resignation to Barings's London office from a hotel in Kuala Lumpur, stating in part, "My sincere apologies for the predicament I have left you in. It was neither my intention nor aim for this to happen."[21]

After Leeson failed to return, Railton and others at Barings's Singapore office began investigating his private records and quickly discovered evidence that he had lost astronomical sums of money. Peter Baring, the bank's chairman, did not learn of the bank's difficulties until the next day, when he was forced to call the Bank of England to ask for assistance. Ironically, this was the same day that Barings was to inform its staff of their bonuses. Leeson was to receive a £450,000 bonus, up from £130,000 the previous year, on the strength of his reported profits. Baring himself expected to receive £1 million.

The Bank of England's Board of Banking Supervision (1995) subsequently conducted an inquiry into the collapse of Barings. According to the Board's report, total losses attributable to Leeson's actions came to £927 million (approximately US$1.4 billion), including liquidation costs; an amount far in excess of Barings total equity of £440 million. Most of the cost of the Barings debacle was borne by its shareholders and by ING, the firm that bought Barings. Barings was a privately held firm; most of its equity was held by the Baring Foundation, a charity registered in the United Kingdom. Barings's executive committee held the firm's voting shares, which constituted a small fraction of the firm's total equity. Although ING was able to buy the failed merchant bank for a token amount of £1, it had to pay £660 million to recapitalize the firm. SIMEX subsequently reported that the funds Barings had on deposit with the exchange were sufficient to meet the costs incurred in liquidating its positions

[21] The full text of Leeson's letter of resignation can be found in Springett (1995).

(Szala, Nusbaum, and Reerink 1995). It is not known whether the OSE suffered any losses as a result of Barings's collapse.

Leeson was later detained by authorities at the airport in Frankfort, Germany, and was extradited to Singapore the following November. In Singapore, Leeson pleaded guilty to charges of fraud and was sentenced to a 6½-year prison term (Mark 1995).

Certain material facts regarding the entire incident are not yet known, as Leeson refused to cooperate with British authorities unless extradited to Great Britain. He later contested the findings of the Banking Board's inquiry, however. A letter to the board from his solicitors states,

> These conclusions are inaccurate in various respects. Indeed, in relation to certain of the matters they betray a fundamental misunderstanding of the actual events. Unfortunately, given the uncertainty regarding Mr. Leeson's position we are not able to provide you with a detailed response to your letter.[22]

Leeson has promised to write a book describing his own version of events while serving out his prison term in Singapore.

Market Aftershocks

Once the Singapore and Osaka exchanges learned that Barings would not be able to meet its margin calls, they took control of all the bank's open positions. The Nikkei index fell precipitously when market participants learned that the exchanges would be liquidating such large positions. Thus, in the days immediately following the announcement of Barings's collapse, it was not known whether the margin money the bank had deposited with the exchanges would cover the losses stemming from the liquidation of its positions.

Matters were further complicated when SIMEX announced it would double margin requirements on its Nikkei stock index futures contract effective Tuesday, February 28. Fearing that their margin money might be used to pay for Barings's losses, several of the exchange's U.S. clearing members threatened to withhold payment of the additional margin SIMEX was demanding of them unless given assurances that such margin payments would be used solely to collateralize their own accounts. A refusal to pay would have caused the affected dealers to forfeit their positions. If that had happened, SIMEX would have been faced with a series of defaults. According to CFTC chairman Schapiro, such an event could have "destroyed the ability of SIMEX to manage the situation."[23] Indeed, there are reports that many market participants feared that the very solvency of the SIMEX clearinghouse was in question. To complicate matters further, Japanese and Singaporean regulators were slow to

[22] Board of Banking Supervision (1995), para. 1.77.

[23] As cited in McGee (1995b).

inform market participants of the steps they were taking to insure the financial integrity of the exchange clearinghouses. This lack of communication served only to exacerbate the fears of market participants (Falloon 1995; Irving 1995; McGee 1995b, c; Szala, Nusbaum, and Reerink 1995).

Upon learning of the situation, Chairman Schapiro contacted the Monetary Authority of Singapore (MAS) to persuade the agency to assure SIMEX's clearing members that their margin deposits would not be used to offset Barings's proprietary losses. The MAS subsequently acceded to these requests and provided its assurance in a short statement released before the start of trading on Tuesday. SIMEX's margin calls were met and a potential crisis was avoided.

This was not the end of headaches for Barings's customers, however. BFS was one of the largest clearing member firms on SIMEX. As such, it handled clearing and settlement for 16 U.S. firms and held approximately $480 million in margin funds on their behalf when it went bankrupt.

U.S. futures exchanges typically arrange the immediate transfer to other firms of all customer accounts of a financially troubled clearing member. Laws in the United States facilitate such transfers because they provide for strict segregation of customer accounts, which prevents the creditors of a broker or clearing member firm from attaching the assets of customers. That Japanese law contains no such provisions was not well known before the collapse of Barings. Although laws in Singapore do recognize the segregation of accounts, SIMEX had never before dealt with the insolvency of a clearing member firm. To complicate matters further, most of BFS's customer accounts had been booked through Baring Securities in London. Consequently, SIMEX did not have detailed information on individual customer positions. It had records only on a single commingled account for Baring Securities. Finally, much of the information that Leeson had provided to the exchange, as well as to Barings's other offices, was false. These circumstances made the task of sorting out the positions of individual customers extremely difficult.

During the next week, Barings's U.S. customers scrambled to reproduce documentation of their transactions with the bank and supplied this information to SIMEX and the OSE. But while this information made it possible for the exchanges to identify customer positions, Barings's bankruptcy administrator in London had asked the exchanges to block access to all Barings's margin deposits. The bankruptcy administrator had raised questions about whether U.K. laws on the segregation of customer accounts were applicable in an insolvency of this kind (Szala, Nusbaum, and Reerink 1995).

It was not until ING took over Barings on March 9 that the bank's customers were assured of access to their funds. Even then, access was delayed in many cases. By one account, several major clients waited more than three weeks before their funds were returned (Irving 1995).

Policy Concerns Highlighted by Barings's Default

All futures exchanges maintain systems to prevent the accumulation of large speculative losses. But events surrounding the collapse of Barings have served to highlight weaknesses in risk management on the part of SIMEX and other futures exchanges. They also suggest a need for closer international cooperation among futures exchanges and their regulators, and for clearer laws on the status of customer accounts when a clearing member firm becomes insolvent.

Futures exchanges maintain stringent speculative position limits for individual firms and traders to prevent large losses and to limit their exposure. It appears that SIMEX relaxed some of these restrictions for BFS, however. It is not unusual for futures exchanges to grant exemptions to established position limits for hedged positions, such as those Leeson claimed to maintain. But it is normal for the exchange clearinghouse to monitor closely the activities of firms receiving such exemptions and to take steps to verify the existence of offsetting exposures. It now appears that SIMEX failed to pursue such precautions in its dealings with Barings.

The exchange's attitude toward Barings was influenced in part by the bank's strong international reputation, but its willingness to relax normal risk management guidelines also may have been attributable to its desire to attract business. Although the OSE was first to list Japanese government bond and Nikkei-225 stock index futures, SIMEX soon began listing similar contracts in direct competition with the Osaka exchange. Thereafter, the two exchanges battled each other for market share. Barings was one of the most active firms on SIMEX—and Leeson was responsible for much of the exchange's trading volume in Nikkei stock index futures and options. Thus, some observers believe that SIMEX may have been too willing to accommodate BFS (McGee 1995c). Critics include representatives of U.S. futures exchanges, who maintain that their risk management standards are more stringent.[24] A report on the incident commissioned by the government of Singapore came to a similar conclusion, finding that the exchange may have been too liberal in granting increases in position limits.[25]

Communication between exchanges can be important for identifying and resolving potential problems. Communication between SIMEX and the OSE was minimal, however. This lack of communication not only helped make it possible for Leeson to accumulate large losses but also hampered efforts to contain the damage once Barings collapsed. Although the OSE routinely published a list of the positions of its most active traders, SIMEX did not make such disclosures. It now seems apparent that SIMEX officials never consulted

[24] See *BNA's Banking Report* (1995f) and Falloon (1995).

[25] See *The Economist* (1995).

the OSE's list to verify Leeson's claim that he was hedging his large positions in Singapore with offsetting exposures on the Osaka exchange.

Some observers blame this lack of communication on the rivalry between the two exchanges. Arrangements existing between U.S. exchanges suggest that competition need not preclude information sharing, however. In the United States, futures exchanges attempt to coordinate their activities with the CFTC and other futures exchanges. Each exchange maintains strict speculative position limits established under CFTC oversight. The CFTC monitors compliance through a comprehensive surveillance policy that includes a large-trader reporting system. Market participants are required to justify unusually large positions. This system enabled the CFTC to ascertain quickly that Barings had no significant positions on any U.S. futures exchange at the time of its collapse.[26]

While competitive concerns may sometimes give exchanges incentives to relax prudential standards, as many observers seem to think that SIMEX did, it does not follow that regulators should seek to discourage such competition. Competition among exchanges serves an important economic function by encouraging innovation. Securities and futures exchanges constantly compete with one another to provide new products to their customers. Thus, whereas futures exchanges once listed contracts only for agricultural and other commodities, a significant fraction of all futures trading today involves contracts for financial instruments. The growth of trading in such instruments has provided important benefits to international financial markets, helping to make them more efficient while facilitating risk management by financial intermediaries and commercial firms alike. Moreover, competition gives futures exchanges an incentive to maintain strong financial controls and risk management systems, as most market participants seek to avoid risks like those faced by SIMEX customers after the collapse of Barings. Finally, policymakers need not restrict competition to address the problems highlighted by the Barings debacle.

The events surrounding the collapse of Barings led futures industry regulators from 16 nations to meet in Windsor, England, in May 1995 to discuss the need for legal and regulatory reform. At that meeting, officials agreed on a plan of action now known as the Windsor Declaration. The declaration calls for regulators to promote, as appropriate, "national provisions and market procedures that facilitate the prompt liquidation and/or transfer of positions, funds and assets, from failing members of futures exchanges," and to support measures "to enhance emergency procedures at financial intermediaries, market members and markets and to improve existing mechanisms for international co-operation and communication among market authorities and regulators."[27] The International Organization of Securities Commissions (IOSCO) later endorsed the Windsor

[26] See the summary of Chairman Schapiro's testimony before Congress in *BNA's Banking Report* (1995e, f).

[27] As cited in *BNA's Banking Report* (1995d).

Declaration and pledged to study the issues it raised. IOSCO also asked its members to promote declaration measures in cross-border transactions.[28]

The Barings debacle has also spurred efforts by market participants to strengthen financial safeguards at futures and options exchanges. In March 1995, the Futures Industry Association (FIA) organized a task force to investigate measures to improve the financial integrity of futures and options exchanges. The association's Global Task Force on Financial Integrity (1995) subsequently published a report containing 60 recommendations, ranging from risk management practices to customer protection issues. The FIA report encourages all nations to review their bankruptcy laws to clarify the status of customer funds and to modify provisions that might conflict with the laws of other nations. It recommends that exchanges and their regulators establish procedures for the transfer of a troubled clearing member firm's customer assets *before* it is declared insolvent, as is now typically done in the United States. In addition, the report encourages exchange clearinghouses to monitor their clearing member firms closely and to perform periodic audits. Thus, the FIA's recommendations are broadly consistent with the principles espoused by the Windsor Declaration, especially in their emphasis on customer protection and the need for improved information sharing among exchanges and government authorities.

Subsequently, the clearing organizations for 19 U.S. stock, stock option, and futures exchanges announced their intent to begin pooling data on transactions of member firms (McGee 1995a). In addition, CFTC Chairman Schapiro has announced that her staff will work with the futures industry to develop concrete customer protection proposals.[29]

The Barings debacle has served to galvanize an international effort—one that has been joined by government officials and market participants alike—to reevaluate risk management systems, customer protection laws, and procedures for dealing with the failure of a large clearinghouse member. It also has prompted increased communication and pledges of greater cooperation among regulators from different nations. It is still too early to pass judgment on the ultimate success of such initiatives, however. While regulators have pledged increased international cooperation, recent press accounts have noted that officials in Britain, Japan, and Singapore have not always cooperated with one another in conducting their investigations of the Barings case.[30]

Lessons from the Barings Debacle

The losses suffered by Barings provide a good example of the market risk associated with derivatives. But, as with the case of Metallgesellschaft, the

[28] See *BNA's Banking Report* (1995b).

[29] See *BNA's Banking Report* (1995c).

[30] See *The Economist* (1995).

Barings debacle best illustrates operational risk and legal risk. In this regard, the Bank of England's Board of Banking Supervision inquiry concluded,

> Barings' collapse was due to the unauthorized and ultimately catastrophic activities of, it appears, one individual (Leeson) that went undetected as a consequence of a failure of management and other internal controls of the most basic kind. Management failed at various levels and in a variety of ways . . . to institute a proper system of internal controls, to enforce accountability for all profits, risks and operations, and adequately to follow up on a number of warning signals over a prolonged period.[31]

The board's inquiry found nine separate warning signs that should have alerted Barings management to problems with its Singapore futures subsidiary. A partial list of those warning signs includes the following:

- *The lack of segregation of duties between front and back offices*. This lack was identified as a weakness and potential problem area in an internal audit report following a review of BFS's operations in the summer of 1994. Barings's management failed to act on the report's recommendations to remedy this situation.
- *The high level of funding requested by Leeson.* Between December 31, 1994, and February 24, 1995, Barings provided Leeson with £521 million to meet margin calls. Total funding of BFS stood at £742 million, more than twice the reported capital of the Barings Group, when Leeson's activities were finally discovered on February 24.[32]
- *The unreconciled balance of funds transferred to BFS to meet margin calls.* In his requests for additional funding, Leeson often claimed the money was needed for client accounts but never provided detailed information about these accounts as was the usual practice. Nonetheless, the bank's head office in London paid those funds without any independent check on the validity of Leeson's requests and with no attempt to reconcile those requests with known trading positions. Perhaps the most troubling aspect of Barings's behavior in this regard is that SIMEX rules prohibit its members from financing the margin accounts of customers. Barings's management apparently ignored evidence that the firm might be doing so in violation of SIMEX rules.
- *The apparent high profitability of Leeson's trading activities relative to the low level of risk as perceived and authorized by Barings's management in London.* High returns typically entail high risk. Yet no one in senior management seriously questioned how Leeson's strong reported profits

[31] Board of Banking Supervision (1995), para. 14.1.
[32] Board of Banking Supervision (1995), para. 6.21.

could result from what was supposed to have been a low-risk activity. To be sure, at least one executive observed that "This guy must be busting his intraday limits or something."[33] But Leeson's reports were never challenged until too late, and management did little to restrain his trading activities. According to interviews with Barings's staff, Leeson was regarded as "almost a miracle worker," and there was "a concern not to do anything which might upset him."[34]

- *The discovery of discrepancies in Leeson's accounts by outside auditors.* Barings's auditors, the firm of Coopers & Lybrand, informed the bank's management of a £50 million discrepancy in BFS's accounts on or before February 1, 1995. Although this discrepancy ultimately did prompt Barings's treasurer to investigate Leeson's accounts, the Board of Banking Supervision concluded that management was too slow in responding to this warning sign.

- *Communications from SIMEX.* The rapid buildup of Leeson's positions during January 1995 prompted SIMEX to seek assurances from Barings's management in London regarding the ability of BFS to fund its margin calls. In retrospect, it appears that Barings's management was too hasty in providing such assurances.

- *Market rumors and concerns made known to Barings's management in January and February.* By late January, rumors were circulating on the OSE regarding Barings's large positions in Nikkei futures. On January 27, the Bank for International Settlements in Basle, Switzerland, raised a high-level inquiry with Barings executives in London regarding a rumor that the bank had experienced losses and could not meet its margin calls on the OSE. On the same day, another Barings executive received a call from the Bloomberg information service inquiring into the bank's large positions on the OSE.

Taken together, these warning signs suggest that Barings's management had ample cause to be concerned about Leeson's activities. But management was too slow to act on these warning signs. An on-site examination of Leeson's accounts came too late to save the bank.

The Board of Banking Supervision's report outlined a number of lessons to be learned from the failure of Barings. They emphasize five lessons for the management of financial institutions:

- Management teams have a duty to understand fully the businesses they manage;

[33] Board of Banking Supervision (1995), para. 3.57.

[34] Board of Banking Supervision (1995), para. 7.12.

- Responsibility for each business activity has to be clearly established and communicated;
- Clear segregation of duties is fundamental to any effective control system;
- Relevant internal controls, including independent risk management, have to be established for all business activities;
- Top management and the Audit Committee have to ensure that significant weaknesses, identified to them by internal audit or otherwise, are resolved quickly.[35]

The report also had some criticisms for the Bank of England's supervision of Barings. U.K. banking regulations require all banks to notify the Bank of England before entering into a transaction that would expose more than 25 percent of the organization's capital to the risk of loss. A Bank of England manager granted Barings an informal concession permitting it to exceed this limit in its exposure to SIMEX and the OSE without first referring the matter to the Bank's senior management. But while the report is somewhat critical of the Bank of England on this matter, it concludes,

> The events leading up to the collapse of Barings do not, in our view, of themselves point to the need for any fundamental change in the framework of regulation in the UK. There is, however, a need for improvements in the existing arrangements.[36]

The report goes on to suggest a number of ways to improve the Bank of England's supervision of banks. According to the report,

- the Bank should explore ways of increasing its understanding of the non-banking businesses . . . undertaken by those banks for which it is responsible;[37]
- it should prepare explicit internal guidelines to assist its supervisory staff in identifying activities that could pose material risks to banks and ensure that adequate safeguards are in place;
- it should work more closely with the Securities and Futures Authority, the agency responsible for regulating the domestic operations of British-based securities firms, as well as with regulators from other nations; and
- it should address deficiencies in the implementation of rules dealing with large exposures.

---

[35] Board of Banking Supervision (1995), para. 14.2.
[36] Board of Banking Supervision (1995), para. 14.5.
[37] Board of Banking Supervision (1995), para. 14.35.

The report also recommended an independent quality assurance review of the Bank of England's supervisory function.

The Board of Banking Supervision's report did not blame the collapse of Barings on its use of derivatives. Instead, it placed responsibility for the debacle on poor operational controls at Barings:

> The failings at Barings were not a consequence of the complexity of the business, but were primarily a failure on the part of a number of individuals to do their jobs properly....While the use of futures and options contracts did enable Leeson to take much greater levels of risk (through their leverage) than might have been the case in some other markets, it was his ability to act without authority and without detection that brought Barings down.[38]

This point has been reinforced recently by news of a similar debacle at the New York office of Daiwa Bank, where a trader concealed large trading losses for over ten years before finally confessing to his activities.[39] Parallels between the Daiwa and Barings debacles are striking, as both incidents resulted from the unauthorized activities of a single trader. Daiwa's losses were in no way related to derivatives, however. The bank incurred over $1 billion in losses as a result of unauthorized trading in U.S. government bonds, widely regarded as the safest of financial instruments.

Some Final Observations on the Barings Debacle

The events surrounding the collapse of Barings have highlighted certain weaknesses in international financial markets that represent legitimate concerns for policymakers. The most notable of these weaknesses relate to (1) the lack of communication between securities and futures exchanges and regulators in different countries, and (2) conflicting laws on the legal status of customer accounts at futures brokers and clearing agents in the event of insolvency. These weaknesses can be addressed only by increased international cooperation among futures exchanges, regulators, and lawmakers.

At the same time, it does not appear that more stringent government regulation of futures markets could have prevented the Barings debacle. Leeson acted outside existing regulatory guidelines and outside the law in concealing the true nature of his trading activities and the losses resulting therefrom. Existing laws and regulations should have been able to prevent, or at least to detect, Leeson's activities before he could incur such astronomical losses. But Barings, SIMEX, and the Bank of England were all lax in enforcing those rules. Barings was lax in enforcing basic operational controls. In doing so, it violated not only official regulations but also commonly accepted market standards for managing risk. Similarly, it appears that SIMEX may have been

[38] Board of Banking Supervision (1995), para. 14.35.

[39] See Sapsford, Sesit, and O'Brien (1995) for early details of the Daiwa debacle.

too liberal in granting increases in position limits to BFS. Finally, the Bank of England granted Barings an exemption that helped make it possible for Leeson to continue his illicit activities undetected.

4. CONCLUDING COMMENTS

The cases of Metallgesellschaft and Barings provide an interesting study in contrasts. Both cases involve exchange-traded derivatives contracts. In both cases, senior management has been criticized for making an insufficient effort to understand fully the activities of their firms' subsidiaries and for failing to monitor and supervise the activities of those subsidiaries adequately. But while critics have faulted MG's management for overreacting to the large margin calls faced by one of its subsidiaries, Barings's management has been faulted for being overly complacent in the face of a large number of warning signs.

If these two disparate incidents offer any single lesson, it is the need for senior management to understand the nature of the firm's activities and the risks that those activities involve. In the case of Metallgesellschaft, the sheer scale of its U.S. oil subsidiary's marketing program exposed the firm to large risks. Although there is a great deal of disagreement over the efficacy of the hedging strategy employed by MGRM, it would seem difficult to argue that members of MG's board of supervisors fully appreciated the nature or magnitude of the risks assumed by the firm's U.S. oil subsidiary. If they had, they would not have been so shocked to find the firm facing large margin calls. In the case of Barings, senior management seemed content to accept that a single trader could earn huge profits without exposing the firm to large risks. With the benefit of hindsight, it seems clear that senior executives of both firms should have taken more effort to understand the activities of subordinates.

News of derivatives-related losses often prompts calls for more comprehensive regulation of derivatives markets. But the cases of Metallgesellschaft and Barings—which rank among the largest derivatives-related losses to date—involve instruments traded in markets already subject to comprehensive regulation. In the case of Barings, the debacle involved a regulated merchant bank trading in regulated futures markets. If anything, the Barings debacle illustrates the limits of regulation. Established rules and regulations should have been able to prevent a single trader from accumulating catastrophic losses. But both SIMEX and the Bank of England granted exemptions that helped make it possible for Leeson to continue his activities for years without being detected. It appears that regulatory organizations can also be subject to operational weaknesses.

Moreover, the instruments traded by these two firms—oil futures, stock index futures, and stock index options—are not the kinds of complex and exotic instruments responsible for concerns often expressed in connection with the

growth of derivatives markets. In the case of Barings, the Bank of England's Board of Banking Supervision concluded that it was not the complexity of the business but the failure of a large number of individuals to do their jobs properly that made the bank susceptible to catastrophic losses by a single trader. As the recent misfortune of Daiwa Bank shows, weaknesses in operational controls can lead to losses in many areas of a firm's operations, not just those involved with derivatives. The losses suffered by Daiwa resulted from trading in U.S. Treasury bonds, widely regarded as the safest of all securities.

Unfortunately, no amount of regulation can remove all risk from financial markets. Risk is inherent in all economic activity, and financial markets exist to help market participants diversify such risks. At the same time, regulation can impose costs on market participants. The Metallgesellschaft case shows that attempts at stringent regulation can sometimes have undesirable side effects. According to critics, the CFTC's action against MG's U.S. subsidiaries has introduced uncertainty about the legal status of commercial forward contracts. As a general rule, government policy should attempt to minimize legal risk rather than create it.

To be sure, the Barings debacle did highlight the need for certain legal and regulatory reforms and for more international cooperation among exchanges and their regulators. But market discipline is also a powerful form of regulation. Highly publicized accounts of derivatives-related losses have led many firms to scrutinize their risk management practices—not only in the area of derivatives, but in other areas of their operations as well. Thus, while it is true that derivatives debacles often reveal the existence of disturbing operational weaknesses among the firms involved, such incidents can also teach lessons that help to make financial markets safer in the long run. As the foregoing accounts show, regulation cannot substitute for sound management practices. At the same time, government policymakers can act to minimize the potential for disruption to financial markets by promoting laws and policies that minimize legal risk.

REFERENCES

BNA's Banking Report. "Schapiro Defends CFTC Action against Two MG Subsidiaries," vol. 65 (September 25, 1995a), p. 503.

_____. "IOSCO Endorses Value-at-Risk Models for Capital Adequacy Calculations," vol. 65 (July 17, 1995b), pp. 120–21.

_____. "Schapiro Says CFTC, Industry to Craft Customer Protection Proposals," vol. 65 (July 3, 1995c), pp. 15–16.

_____. "World Futures Regulators Adopt Plan of Action in View of Barings," vol. 64 (May 22, 1995d), pp. 1017–18.

_____. "CFTC Undertaking Regulatory Review: Schapiro Address[es] Failure of Barings PLC," vol. 64 (March 6, 1995e), pp. 469–70.

_____. "CFTC Chairman Schapiro Tells Congress Barings-Type Disaster Unlikely in U.S.," vol. 64 (March 6, 1995f), pp. 468–69.

Board of Banking Supervision. *Report of the Board of Banking Supervision Inquiry into the Circumstances of the Collapse of Barings*. London: HMSO, 1995.

Culp, Christopher L., and Steve H. Hanke. "Derivative Dingbats," *The International Economy* (July/August 1994).

Culp, Christopher L., and Merton H. Miller. "Basis Risk and Hedging Strategies: Reply to Mello and Parsons," *Derivatives Quarterly*, vol. 1 (Summer 1995a), pp. 20–26.

_____. "Hedging in the Theory of Corporate Finance: A Reply to Our Critics," *Journal of Applied Corporate Finance*, vol. 8 (Spring 1995b), pp. 121–27.

_____. "Auditing the Auditors," *Risk*, vol. 8 (April 1995c), pp. 36–39.

_____. "Metallgesellschaft and the Economics of Synthetic Storage," *Journal of Applied Corporate Finance*, vol. 7 (Winter 1995d), pp. 62–76.

_____. "Letter to the Editor," *Risk*, vol. 11 (November 1994a), p. 18.

_____. "Hedging a Flow of Commodity Deliveries with Futures: Lessons from Metallgesellschaft," *Derivatives Quarterly*, vol. 1 (Fall 1994b), pp. 7–15.

The Economist. "The Barings Collapse: Spot the Smoking Receivable," October 21, 1995, p. 79.

_____. "Revolution at Metallgesellschaft," December 25, 1993, p. 90.

Edwards, Franklin R. "Derivatives Can Be Hazardous to Your Health: The Case of Metallgesellschaft," *Derivatives Quarterly*, vol. 1 (Spring 1995), pp. 8–17.

_____, and Michael S. Canter. "The Collapse of Metallgesellschaft: Unhedgeable Risks, Poor Hedging Strategy, or Just Bad Luck?" *The Journal of Futures Markets*, vol. 15 (May 1995a), pp. 211–64.

_____. "The Collapse of Metallgesellschaft: Unhedgeable Risks, Poor Hedging Strategy, or Just Bad Luck?" *Journal of Applied Corporate Finance*, vol. 8 (Spring 1995b), pp. 86–105.

Falloon, William. "Who's Missing from the Picture?" *Risk*, vol. 8 (April 1995), pp. 19–22.

Fox, Justin. "2 in House Fault Futures Trading Panel on Over-the-Counter Derivatives Ruling," *American Banker*, December 26, 1995.

Futures Industry Association Global Task Force on Financial Integrity. *Financial Integrity Recommendations*. Washington: Futures Industry Association, 1995.

Global Derivatives Study Group. *Derivatives: Practices and Principles.* Washington: Group of Thirty, 1993.

Irving, Richard. "Beyond Barings," *Risk*, vol. 8 (April 1995), p. 6.

Kawaller, Ira. "Hedging with Futures Contracts: Going the Extra Mile," *Journal of Cash Management*, vol. 6 (July/August 1986), pp. 34–36.

Mark, Jeremy. "With Leeson in Singapore Prison, Focus Shifts to Barings Executives," *The Wall Street Journal*, December 4, 1995.

McGee, Suzanne. "Nineteen U.S. Exchanges To Share Data on Dealings," *The Wall Street Journal*, September 6, 1995a.

_____. "New U.S. Futures Chief Is Crafting a Global Role," *The Wall Street Journal*, May 17, 1995b.

_____. "Reform Sought to Prevent Future Barings," *The Wall Street Journal*, March 17, 1995c.

Mello, Antonio S., and John E. Parsons. "Hedging a Flow of Commodity Deliveries with Futures: Problems with a Rolling Stack," *Derivatives Quarterly*, vol. 1 (Summer 1995a), pp. 16–19.

_____. "Maturity Structure of a Hedge Matters: Lessons from the Metallgesellschaft Debacle," *Journal of Applied Corporate Finance*, vol. 8 (Spring 1995b), pp. 106–20.

Miller, Merton H., and Christopher L. Culp. "Rein in the CFTC," *The Wall Street Journal*, August 17, 1995.

Rance, Brian D. "The Commodity Futures Trading Commission Order of Settlement with MG Refining and Marketing, Inc. and MF Futures, Inc.," *Derivatives Quarterly*, vol. 2 (Winter 1995), pp. 13–17.

Roth, Terrence. "German Firm's Bailout Package Gets Approved," *The Wall Street Journal*, January 17, 1994a.

_____. "Metallgesellschaft Sets Shareholder Vote Despite Worries over Financial Package," *The Wall Street Journal*, January 14, 1994b.

Sapsford, Jathon, Michael R. Sesit, and Timothy L. O'Brien. "How Daiwa Bond Man in New York Cost Bank $1.1 Billion in Losses," *The Wall Street Journal*, September 27, 1995.

Springett, Pauline. "The Barings Rescue: 'Apologies, but the pressures have become too much to bear,' " *The Guardian* (London), March 7, 1995.

Stoll, Hans R. "Lost Barings: A Tale in Three Parts Concluding with a Lesson," *The Journal of Derivatives*, vol. 3 (Fall 1995), pp. 109–15.

Szala, Ginger, David Nusbaum, and Jack Reerink. "Barings Abyss," *Futures*, vol. 24 (May 1995), pp. 68–74.

Taylor, Jeffrey. "Securities Firms Agree to Set Controls on Derivatives," *The Wall Street Journal*, March 9, 1995a.

_____. "Documents at MG Appear to Contradict Fired Chairman's Stance on Derivatives," *The Wall Street Journal*, January 27, 1995b.

_____. "MG Corp. Aims to Stay Active in Oil Business," *The Wall Street Journal*, March 14, 1994.

U.S. Commodity Futures Trading Commission. "Order Instituting Proceedings Pursuant to Sections 6(c) and 8(a) of the Commodity Exchange Act and Findings and Order Imposing Remedial Sanctions," CFTC Docket No. 95-14 (1995a).

_____. "CFTC Order Imposes $2.25 Million Civil Penalty against MG Refining and Marketing, Inc. and MG Futures, Inc.; Other Remedial Sanctions Include Review of Internal Control Systems for Risk Management," Release No. 3859-95 (July 27, 1995b).

U.S. General Accounting Office. *Financial Derivatives: Actions Needed to Protect the Financial System.* GAO/GCD-94-133, May 1994.

The Wall Street Journal. "Former CFTC Chief Scores Agency's Move in Derivatives Case," December 1, 1995.

_____. "Metallgesellschaft A.G. Dismisses Chairman and Finance Chief," December 20, 1993.

Was the Disinflation of the Early 1980s Anticipated?

Michael Dotsey and Jed L. DeVaro

The United States experienced a rise in inflation from 5.5 percent in the first quarter of 1976 to 9.4 percent in the first quarter of 1980. This steep increase was followed by the desire to stem inflation and to reverse its course. Therefore, soon after Paul Volcker was appointed chairman of the Federal Reserve Board in August 1979, the Federal Reserve (the Fed) announced new operating procedures and put in place a disinflationary policy.

During the implementation of that policy, two recessions came in rapid succession. The first began in January 1980 and lasted through the middle of that year. Although short in duration, the recession was deep, with real GDP falling at an annual rate of 10.4 percent in the second quarter of 1980, the largest postwar decline on record. The second recession followed almost immediately, beginning in July 1981 and ending in November 1982. At approximately the same time, the Fed's aggressive disinflationary policy ended. Some economists have calculated the cost in terms of cumulative output lost to be roughly 24 percent of total output over the period 1980:1 to 1983:4.[1] But during that period, inflation was brought down from 10.1 percent in 1980 to 4.4 percent in 1982.

This episode was dramatic because of both the significant loss of output and the equally significant decline in inflation. Also, the policy of disinflation was announced and was carried out over a period of three years. Thus, the episode has the potential to shed light on a number of competing macroeconomic theories that attempt to explain the linkage between real economic activity and

We wish to thank Alan Stockman for suggesting this topic. We are also indebted to James Hamilton for many helpful discussions and for providing the programs that with slight modifications were used in the statistical analysis. Mary Finn, Tom Humphrey, Peter Ireland, and Roy Webb contributed a number of useful suggestions. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Using somewhat different procedures, both Fischer (1986) and Ball (1993) arrive at this conclusion.

nominal disturbances. The usefulness of the 1980's experience in this purpose depends critically on whether the decline in inflation was anticipated. If the disinflation was largely unexpected, even after the policy had been in place for some time, then it will be difficult to discriminate between the theories.

The first set of theories can broadly be classified as Keynesian and are characterized by elements of nominal rigidities in prices and wages. Specifically, we have in mind models like those of Fischer (1977), Taylor (1980), or Ball, Mankiw, and Romer (1988). In these models, staggered price setting slows the adjustment of the price level to changes in the money stock. Thus, inflation has inertia, and anticipated monetary contractions can precipitate a recession by causing aggregate demand to fall. The fall in aggregate demand will of course be greater if the disinflationary policy is unanticipated, since even those firms in the process of resetting their prices will make mistakes.

Other models in which unanticipated changes in monetary policy have an effect on real economic activity are the neoclassical monetary business-cycle models of Lucas (1973), McCallum (1980), and Sargent and Wallace (1975). In these models, agents partially infer nominally induced movements in prices as being driven by real disturbances. Thus, an unanticipated fall in money causes prices to be lower than expected, resulting in a decline in the supply of output. If policy is anticipated, however, and there is no confusion about the impulses driving prices, agents are not induced to supply less output. The entire effect of policy is nominal.

Economists have varying opinions concerning the extent to which disinflation was anticipated during the 1980-83 period, and these views partially determine their preferred theory. For example, in discussing a number of policy changes, including the one in which we are interested, Akerlof, Rose, and Yellen (1988) argue that

> Mr. Volcker's similar policy produced changes in equilibrium output long after the policies were announced (and seem to be credible). The changes persisted sufficiently long after their announcement that it is extraordinarily difficult to believe that the changes in employment and output they caused were due to the slow propagation of unanticipated shocks. (P. 68)

Their view is that the Fed possessed sufficient credibility, implying that the policy change was believed either immediately or soon thereafter and hence that the disinflation was largely anticipated. If that was indeed the case, then their conclusion that the episode was more in line with Keynesian-style theories than neoclassical ones is well taken.

The degree of Fed credibility, and hence the extent to which the disinflation was anticipated, is open to question. Goodfriend (1992), in his description of the Fed's fight against inflation, divides the period into three distinct parts: the aborted fight against inflation from October 1979 to April 1980, when short-term interest rates rose by over 600 basis points; an easing of policy during

the first recession, when rates fell by 800 basis points; and a second aggressive disinflationary policy from August 1980 to October 1982, when rates were eventually pushed up by 1000 basis points over their July 1980 levels. This last sustained tightening broke the inflationary environment. Thus Federal Reserve policy was far from uniform and this lack of uniformity may have impaired the Fed's credibility, implying that the disinflation could very well have been unanticipated. Thus, how well the public anticipated the disinflation can help measure the Federal Reserve's credibility during that period.

1. DESCRIPTION OF THE 1976–1983 PERIOD

We begin with a brief overview of the period in question, confining ourselves to a description of the behavior of inflation, real GDP growth, M1 growth, and the federal funds rate (see Figures 1a to 1c). Inflation is measured by quarterly changes in the GDP deflator and all quarterly growth rates are annualized.

As Figure 1a illustrates, the period was characterized by a run-up in inflation from 5.5 percent in 1976:1 to 11.1 percent in 1981:1 and then by a rapid decline to 4.09 percent in 1983:4. Associated with this rapid disinflation were two recessions (Figure 1b). The first one began in January 1980 and was accompanied by a 10.4 percent annualized decline in real GDP during the second quarter of that year, and the second one began in July 1981.

Figure 1a GDP Deflator Inflation: 1970 to 1986

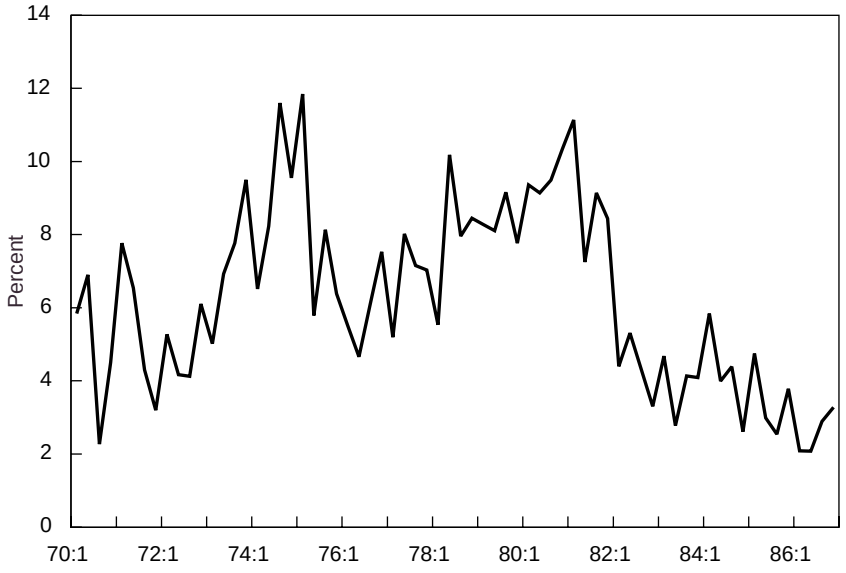
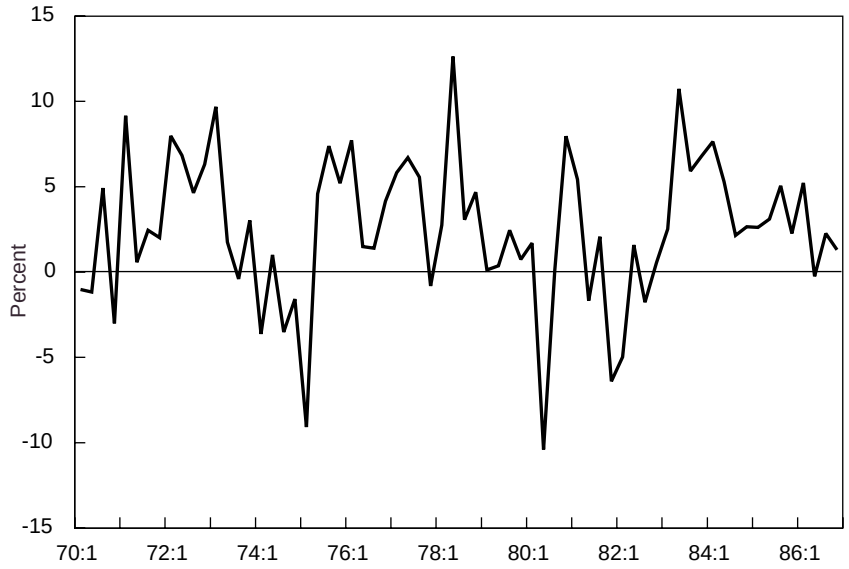


Figure 1b Real GDP Growth: 1970 to 1986



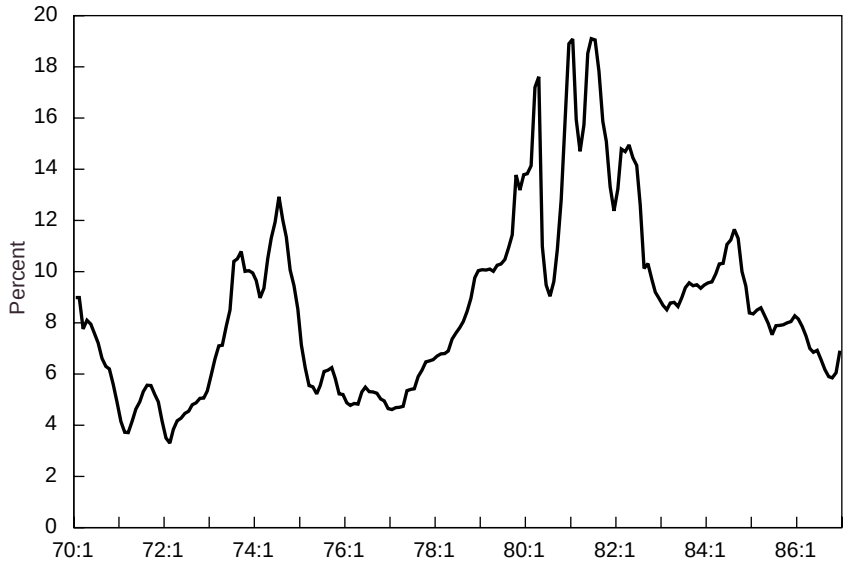
In response to the increase in inflation, the Federal Reserve announced a change in operating procedures in October 1979 and raised the funds rate from 11.4 percent in September to 17.6 percent in April 1980 (Figure 1c). The steep decline in economic activity caused the Fed to temporarily back off from its disinflationary policy, and the funds rate fell to 9.03 percent in July 1980. As a result, not much headway was made in curbing inflation.

As the economy recovered in the fourth quarter of 1980, the Fed resumed its policy of disinflation. The funds rate was raised from 10.8 percent in September 1980 to as high as 19.10 percent in June 1981. This second attempt at reversing the inflationary trend in the economy was successful. The economy experienced its second recession, but inflation also fell from 11.1 percent in 1981:1 to 4.4 percent in 1982:1.[2]

The go-stop-go nature of the Fed's fight against inflation is also depicted in the growth of effective M1 against its targets. When gauging the tightness of monetary policy by the discrepancy between money growth and monetary targets, effective M1 is the appropriate aggregate to look at. Effective M1 is a measure of actual M1 adjusted for the effects of deregulation and was the

[2] For a much more detailed description of this period, see Goodfriend (1992).

Figure 1c Federal Funds Rate: 1970 to 1986



monetary variable of primary concern to the Fed over this period.[3] In 1980, when the Fed relaxed its disinflationary policy, effective M1 grew at a rate of 6.9 percent, which exceeded the top of the target range by 0.4 percent. Also, much of this growth took place in the second half of the year after policy had eased. In 1981, however, effective M1 grew by only 2.4 percent, a full 1.1 percent below the bottom of the target range. This decline in M1 growth was a reflection of the reinstitution of tight monetary policy and served to further signal the Fed's renewed anti-inflationary stance.

Thus, an examination of the two most relevant economic series for depicting monetary policy during the early 1980s—the funds rate and effective M1—indicates that the fight against inflation was somewhat discontinuous. It is, therefore, highly possible that the credibility of the Fed's policy was achieved only gradually.

The discontinuous nature of policy led to renewed inflation in 1980 as well as to a sharp increase in long-term bond rates relative to the funds rate—what Goodfriend calls an "inflation scare"—earlier in the year. From Goodfriend's account, one may conclude that the Fed did not achieve credibility until the

[3] For more detail on effective M1 behavior and its construction, see Broaddus and Goodfriend (1984).

summer of 1982. It is thus entirely possible that although the Fed announced its resolve to fight inflation as early as October 1979, the policy was not fully credible until mid-1982. Thus, it may be that much of the disinflation was unexpected and that the 1980-83 period is perfectly consistent with the neoclassical model.

To shed further light on the degree to which the disinflation was anticipated, we adopt the innovative empirical methodology of Hamilton (1992). By combining information in commodity futures markets with the macroeconomic information readily available at the time, we decompose actual inflation into its anticipated and unanticipated components. Our results imply that a substantial portion of the disinflation was unanticipated and that the Fed suffered from a credibility problem.

2. METHODOLOGY[4]

To analyze the degree to which the disinflation of the early 1980s was anticipated, we must construct a series depicting the public's expectations of inflation. The methodology we use for constructing such a series is that of Hamilton (1992). His procedure incorporates publicly available aggregate data and financial market data on commodity futures contracts to estimate price-level expectations. The data on futures contracts are optimally weighted in an effort to uncover information possessed by the public but not by the econometrician. The use of future changes in commodity prices from their expected values represents a novel way to uncover what agents believed at any given moment. Because the change in policy was a major one, these data are potentially useful for uncovering beliefs about inflation since financial markets often react aggressively to changing inflationary expectations.[5]

Specifically, suppose that the public's expectation of next period's price level, p_{t+1}, is given by

$$p^e_{t+1} = x'_t\delta + \alpha_t, \tag{1}$$

where the superscript e denotes an expectation, x_t is a vector of all relevant aggregate information, and α_t represents information that agents find valuable in forecasting prices but that is unavailable to the econometrician. For example, x_t could include economic time series published by the Commerce Department or the Federal Reserve, while α_t could involve disaggregated information that individuals observe but that is unpublished.

[4] The description in this section draws heavily on Hamilton (1992), who provides a more detailed and technically rigorous description.

[5] The model does not explicitly incorporate regime changes and to some extent suffers from the same problems as a standard VAR when regimes actually do change.

Let the public's true expectational error be defined by $a_{t+1} \equiv p_{t+1} - p^e_{t+1}$. Note that under rational expectations a_{t+1} is white noise and uncorrelated with time t information. Next, consider the forecasting error, u_{t+1}, that occurs from the prediction of p_{t+1} using only the information in x_t. Specifically, u_{t+1} is the error term in

$$p_{t+1} = x_t'\delta + u_{t+1}, \tag{2}$$

which is a typical forecasting equation. From equation (1) and the definition of a_{t+1}, u_{t+1} is equal to $\alpha_t + a_{t+1}$. By including enough own lags of p_t in x_t, u_{t+1} can be made white noise. A simple and consistent assumption is to treat α_t as white noise as well.

Because the variance of u_{t+1}, σ_u^2 is equal to $\sigma_\alpha^2 + \sigma_a^2$, we observe that if the econometrician uses only the information in x_t to forecast p_{t+1}, the variance of the forecast errors will exceed that of the true forecast errors. Improving upon these forecasts requires inferences of α_t. The better the inference of α_t the closer the econometrician's forecast will coincide with the public's. Hamilton proposes that data from commodities markets be used to help form an optimal prediction of α_t.

For ease of exposition, we will analyze the case in which there is one commodity. Using data on the log of the commodity's forward price at date t, f_t, and the log of its expected future spot price next period, $E_t s_{t+1}$, a simple efficient markets model implies that

$$f_t = E_t s_{t+1} - k, \tag{3}$$

where k incorporates the variance of s_t and a constant risk premium. Equation (3) implies that the forecast error of the spot price is observable and given by

$$s_{t+1} - f_t = k + v_{t+1}, \tag{4}$$

where v_{t+1} is the forecast error. Under rational expectations, v_{t+1} is uncorrelated with time t information. In particular, it is uncorrelated with the elements of x_t and α_t. It is reasonable, however, to believe that the forecast error for a given commodity and the aggregate price level are correlated. For example, any demand shock (such as a monetary policy shock) could influence all prices, including commodity prices, in a similar way. This relationship is given by

$$v_{t+1} = q^a a_{t+1} + e_{t+1}. \tag{5}$$

Hence, observation of v_{t+1} implies some knowledge of a_{t+1}.

Next, consider a regression of the forward price on information available to the econometrician,

$$f_t = x_t'\beta + \omega_t. \tag{6}$$

In this regression, ω_t represents information that agents find relevant for pricing a commodity and that is unavailable to the econometrician, because f_t is

observed at time t, and if the econometrician had all the relevant information then ω_t would be zero. Thus, ω_t has an interpretation similar to α_t, and the two should be related. Let this relationship be described by the linear projection

$$\omega_t = q^{\alpha}\alpha_t + \varepsilon_t, \tag{7}$$

where ε_t denotes information that agents have about future commodity price movements that is uncorrelated with aggregate price movements.

The statistical problem is to form an optimal forecast of α_t given knowledge of u_{t+1}, v_{t+1}, and ω_t, because it is these three observed disturbances that contain information about α_t. This optimal forecast allows us to form the expectation of the aggregate price level that best represents the one formed by the public. Using data on aggregate prices, commodity spot and forward prices, and x_t, construct the error terms u_{t+1}, v_{t+1}, and ω_t according to equations (2), (4), and (6). The optimal predictor of α_t can then be formulated as

$$E_{t+1}\alpha_t = A_1 u_{t+1} + A_2\omega_t + A_3 v_{t+1}, \tag{8}$$

where

$$\begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} \sigma_a^2 + \sigma_\alpha^2 & q^{\alpha}\sigma_\alpha^2 & q^{a}\sigma_a^2 \\ q^{\alpha}\sigma_\alpha^2 & (q^{\alpha})^2\sigma_\alpha^2 + \sigma_\varepsilon^2 & 0 \\ q^{a}\sigma_a^2 & 0 & (q^{a})^2\sigma_a^2 + \sigma_e^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_\alpha^2 \\ q^{\alpha}\sigma_\alpha^2 \\ 0 \end{bmatrix}. \tag{9}$$

The coefficients in equation (8) are population regression coefficients. The matrix in (9) is the variance-covariance matrix of the disturbances $u_{t+1} = a_{t+1} + \alpha_t$. In deriving this matrix, various orthogonality conditions implied by rational expectations were used. The vector $[\sigma_\alpha^2, q^{\alpha}\sigma_\alpha^2, 0]'$ is the covariance between α_t and the three observed disturbance terms.

The optimal forecast of next period's price level is then depicted by

$$p_{t+1}^{e} = x_t'\delta + A_1(p_{t+1} - x_t'\delta) + A_2(f_t - x_t'\beta) + A_3(s_{t+1} - f_t - k). \tag{10}$$

In calculating the public's expectations of next period's price level, the econometrician uses time $t+1$ information from commodity spot markets. This information is needed to extract the optimal forecast of α_t, which contains relevant information available to the public but not to the econometrician.

Unfortunately, there is one technical difficulty. To estimate the coefficients A_1, A_2, and A_3, we must have estimates of q^a, q^{α}, σ_a^2, σ_α^2, σ_e^2, and σ_ε^2. The variance-covariance matrix in (9), however, has only five independent pieces of information, which means that the system is not identified. An additional restriction is needed, and we follow Hamilton by imposing $q^a = q^{\alpha}$.[6]

[6] The solution is $A_1 = (1/\Delta)[q^2\sigma_\alpha^2\sigma_\varepsilon^2(\sigma_a^2 + \sigma_e^2)]$
$A_2 = (1/\Delta)q\sigma_a^2\sigma_\alpha^2\sigma_e^2$
$A_3 = -(1/\Delta)q\sigma_a^2\sigma_\alpha^2\sigma_\varepsilon^2$ and $\Delta = (\sigma_a^2 + \sigma_\alpha^2)\sigma_e^2\sigma_\varepsilon^2 + q^2\sigma_a^2\sigma_\alpha^2(\sigma_e^2 + \sigma_\varepsilon^2)$.

This restriction has some intuitive appeal. Suppose, for example, that the relevant aggregate information that agents possess is a demand shock that affects commodity prices and the aggregate price level in similar ways. That is, anticipated and unanticipated movements affect s_{t+1} and p_{t+1} proportionately, although the absolute effects of an anticipated movement need not be the same as those of an unanticipated movement. Then q^a and q^α should be equal. Under the restriction $q^a = q^\alpha$, the ratio of the covariance of u_{t+1} and v_{t+1} to the covariance of u_{t+1} and ω_t is equal to $\sigma_a^2/\sigma_\alpha^2$, which reflects the extent that agents are actually surprised by movements in the aggregate price level.

However, this restriction may not be valid. If, for example, the unanticipated disturbances that affect commodity prices are largely idiosyncratic, while the information that agents actually possess affects both markets similarly, then q^a would not equal q^α. In any event, the restriction can be tested. And with multiple commodities, the restriction need only be placed on one commodity to achieve identification.

3. ESTIMATION

Before estimating the analogous multivariate system given by (1), (4), (6), and the variance-covariance matrix in (9), we describe our commodity price data and perform some necessary diagnostic tests. We used data on wheat, corn, oats, and soybean futures, since these were the only commodity data available for our sample period. In what follows, we use the price of a futures contract that is about to expire as the measure of the spot price. Because the futures contracts are four months in duration, the data set is three times per year (such intervals are denoted by roman numerals).

Commodity Price Behavior

Expected and actual inflation rates for the four commodities are given in Table 1. Actual inflation in a commodity market is calculated as the change in the log of the spot price during the period in question, while expected inflation is represented by three times the average log difference between the four-month futures price and the spot price at each four-month interval. From the table it is evident that expected commodity price increases greatly exceeded actual increases during the 1981-83 period. Further, graphs of expected commodity price inflation versus actual commodity price inflation (Figures 2a to 2d) show that, with the exception of oats, expected inflation was generally higher than actual inflation. These results foreshadow the main results of the statistical model: the overestimate of commodity price inflation will be mirrored in an overprediction of the aggregate inflation rate.

Commodity Market Efficiency

One of the maintained hypotheses of the model discussed in Section 2 is that the commodity market is efficient. This was depicted by equation (4) in which

Table 1 Expected and Actual Inflation Rates for Four Commodities

	1978 to 1980	
Commodity	Expected	Actual
Wheat	4.01	10.16
Corn	14.82	17.24
Oats	31.11	13.79
Soybeans	9.93	6.44

	1981 to 1983	
Commodity	Expected	Actual
Wheat	19.78	−0.39
Corn	11.08	6.05
Oats	3.34	−5.48
Soybeans	10.80	9.01

Sources: Commodity prices were obtained from the Chicago Board of Trade Statistical Annual. Actual inflation figures represent the change in the log of the spot price between September 1978 and September 1980 (upper panel) and September 1981 and September 1983 (lower panel) divided by three, while the expected inflation figures represent three times the average log difference between the four-month future price at time t and the spot price at time t, where t is indexed three times per year from September 1978 to May 1980 (upper panel) and from September 1981 to May 1983 (lower panel).

the expectational error, v_{t+1}, is normally distributed white noise. We therefore test to determine whether v_{t+1} is indeed normal and whether it is correlated with time t information. P-values for the skewness and kurtosis tests indicated that in the cases of wheat and soybeans, we can reject normality. Furthermore, the error terms were correlated with a menu of available time t information. We thus cannot proceed under the assumption that equation (4) provides an adequate description of commodity market behavior. Instead we allow for time varying risk premiums and replace equation (4) with

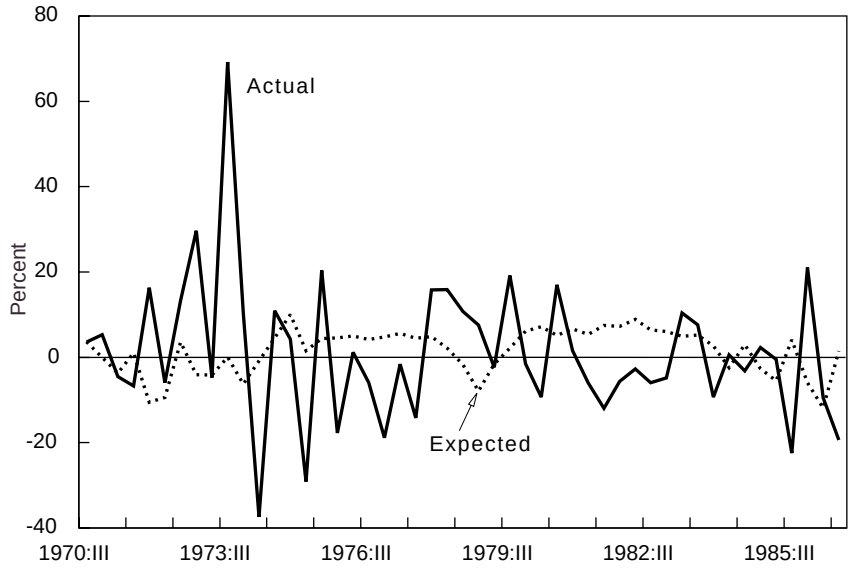
$$s_{t+1} - f_t = x_t'\kappa + v_{t+1} \tag{4'}$$

for each commodity. Upon doing so, we cannot reject the normality of v_{t+1} and we also find that v_{t+1} is uncorrelated with interest rates, forward commodity prices, and its own lagged values.

Estimating and Testing the Model

Before formally estimating the model, we performed augmented Dickey-Fuller tests on aggregate prices, spot prices, and forward commodity prices over

Figure 2a Expected and Actual Commodity Price Inflation for Wheat: 1970 to 1985



Note: Roman numerals denote four-month intervals.

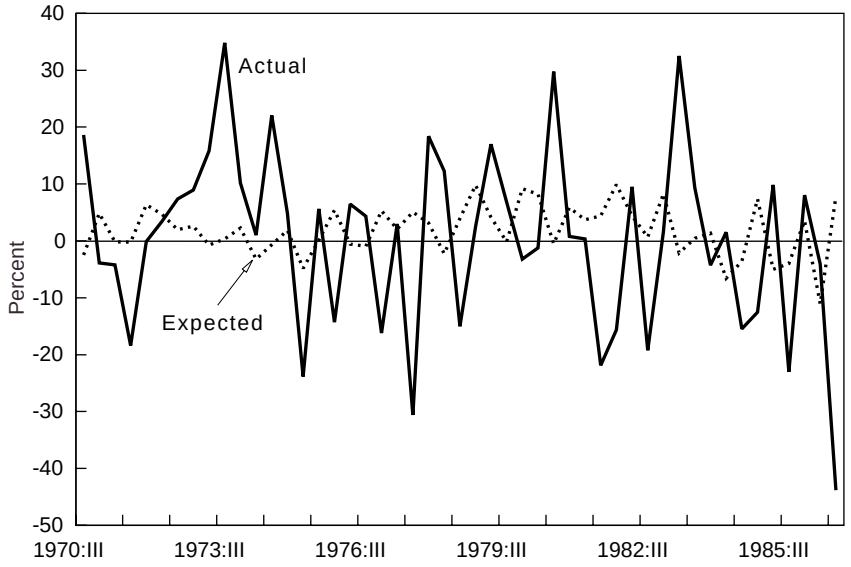
the period 1970:II to 1986:III. We could not reject nonstationarity of the price level, but we were able to reject nonstationarity of commodity prices. Also, for u_{t+1} to be white noise required the inclusion of two lags of the inflation rate. The system that we estimate is depicted by

$$\begin{bmatrix} \Delta p_{t+1} \\ f_t \\ s_{t+1} - f_t \end{bmatrix} = \begin{bmatrix} \delta' \\ \beta' \\ \kappa' \end{bmatrix} x_t + \begin{bmatrix} u_{t+1} \\ \omega_t \\ v_{t+1} \end{bmatrix}, \tag{11}$$

where f and s are 4 by 1 vectors, δ is a 9 by 1 coefficient vector, β and κ are 9 by 4 matrices of coefficients, and u, ω, and v are the disturbance terms. The vector of explanatory variables x_t includes a constant term, two seasonal dummies (d1 and d2), two lags of the inflation rate, and the four commodity spot prices.[7]

[7] This is a fairly parsimonious statistical model. In principle, other variables such as income or money could be included, but doing so would greatly increase the computational burden. Our representation of the inflation process follows Hamilton (1992).

Figure 2b Expected and Actual Commodity Price Inflation for Corn: 1970 to 1985



Note: Roman numerals denote four-month intervals.

The variance-covariance matrix in (9) is now 9 by 9 and is given by

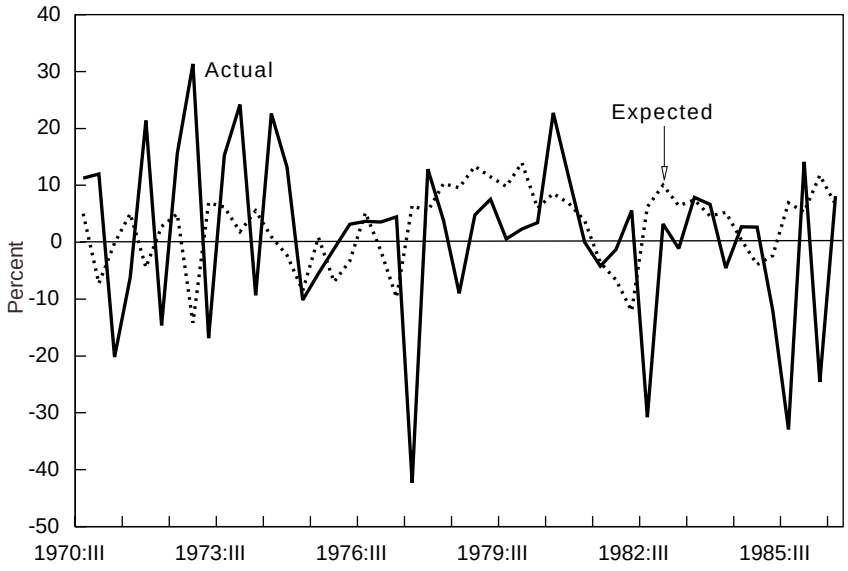
$$\Omega = \begin{bmatrix} \sigma_a^2 + \sigma_\alpha^2 & \sigma_\alpha^2 (q^\alpha)' & \sigma_a^2 (q^a)' \\ \sigma_\alpha^2 (q^\alpha)' & \sigma_\alpha^2 q^\alpha (q^\alpha)' + \Sigma & 0 \\ \sigma_a^2 (q^a)' & 0 & \sigma_a^2 q^a (q^a)' + S \end{bmatrix}, \tag{12}$$

where $\Sigma = E[\varepsilon_t \varepsilon_t']$ and $S = E[e_t e_t']$. We estimate equations (11) and (12) by full-information maximum-likelihood and jointly test the orthogonality conditions assumed under rational expectations and the identifying restriction $q^\alpha = q^a$. We use the Sims (1980) adjusted likelihood-ratio test, which is distributed $\chi^2(19)$.[8] The p-value for the test statistic was 0.45, implying that we cannot jointly reject the restrictions at standard confidence levels. However, for the system using all four commodities, there is not significant covariation between ω_t and u_{t+1}, and ω_t and v_{t+1}. As a result, we cannot reject the hypothesis that $q^a = q^\alpha = 0$.

This lack of rejection is due largely to the inclusion of wheat. Figure 2a shows that actual wheat inflation is volatile, especially early in the sample

[8] The test statistic is $2(\frac{T-k}{T})$ (likelihood [unrestricted]-likelihood [restricted]).

Figure 2c Expected and Actual Commodity Price Inflation for Oats: 1970 to 1985



Note: Roman numerals denote four-month intervals.

period. For a system containing only corn, oats, and soybeans, the covariation between the relevant error terms is more significant, and the hypothesis that $q^a = q^\alpha = 0$ can be rejected at the 9 percent significance level. Furthermore, the joint hypothesis involving $q^a = q^\alpha$ and the rational expectations orthogonality conditions has a p-value of 0.66.

The estimation results for the model are displayed in Table 2. The variance of the inflation forecast u_{t+1}, σ_u^2, is 0.538, most of which is due to $\sigma_a^2 = 0.52$. This result implies that most of the residual, u_{t+1}, took people by surprise.

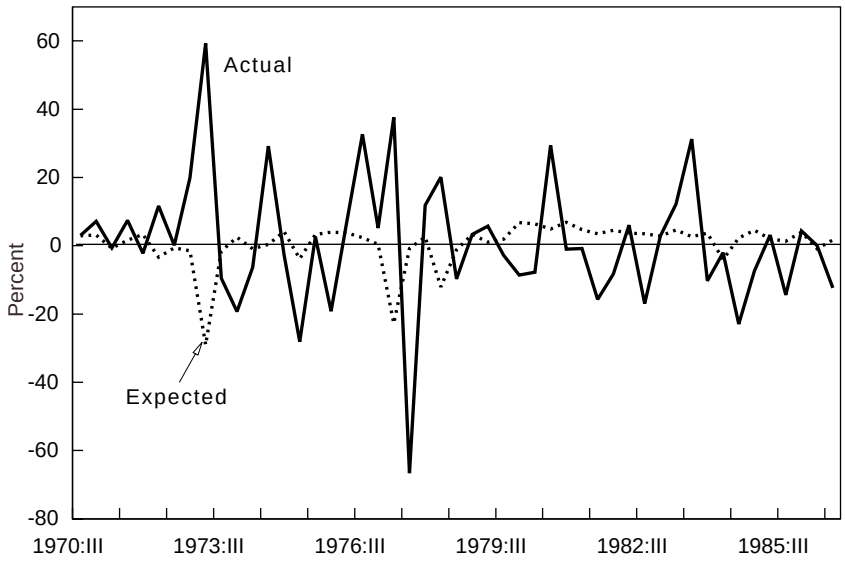
4. PREDICTION

The predictions of inflation over the period 1971:I to 1986:III are calculated using

$$\Delta p_{t+1}^e = x_t'\delta + E_{t+1}\alpha_t. \tag{13}$$

The last term in (13) is derived from the optimal prediction formula given in (10). Thus,

Figure 2d Expected and Actual Commodity Price Inflation for Soybeans: 1970 to 1985



Note: Roman numerals denote four-month intervals.

$$\begin{aligned}\Delta p^{e}_{t+1} = {} & -1.76 - 0.08d1_t - 0.08d2_t + 0.35\Delta p_t + 0.14\Delta p_{t-1} \\ & -0.006s_{c,t} - 0.012s_{o,t} + 0.018s_{s,t} \\ & +0.0045(p_{t+1} - x'_t\delta) \\ & -0.011(f_{c,t} - x'_t\beta_c) - 0.012(f_{o,t} - x'_t\beta_o) \\ & +0.041(f_{s,t} - x'_t\beta_s) \\ & -0.000041(s_{c,t+1} - x'_t\kappa_c) - 0.000027(s_{o,t+1} - x'_t\kappa_o) \\ & -0.000013(s_{s,t+1} - x'_t\kappa_s),\end{aligned} \tag{14}$$

where the subscripts *c*, *o*, and *s* refer to corn, oats, and soybeans.

Using (14) we can decompose the forecasts of inflation into simple forecasts, $x'_t\delta$, and the individual contributions from the price term ($p_{t+1} - x'_t\delta$), the three futures terms ($f_t - x'_t\beta$), and the three commodity market surprise terms ($s_{t+1} - f_t - x'_t\kappa$). These decompositions are given in Table 3. Examining the entire sample period, the contributions of the price term and the futures term can at times be meaningful, affecting the forecasts by an annual rate as high as 87 and 51 basis points, respectively. The commodity market surprise

Table 2 Estimation Results^{+}

unrestricted log likelihood : -760.43

restricted log likelihood : -765.55
($q^a = q^\alpha$ and orthogonality conditions imposed)

restricted log likelihood : -769.46
($q^a = q^\alpha = 0$ and orthogonality conditions imposed)

$$\sigma_\alpha^2 = 0.017 \quad \sigma_a^2 = 0.52 \quad q' = \begin{bmatrix} 4.60 & 2.59 & 23.41 \\ (0.65) & (0.60) & (0.39) \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 5.84 & 4.62 & 3.18 \\ 4.62 & 25.42 & 9.09 \\ 3.18 & 9.09 & 4.80 \end{bmatrix} \qquad S = \begin{bmatrix} 171.08 & 72.26 & 45.97 \\ 72.26 & 123.17 & -69.94 \\ 45.97 & -69.94 & 3712.3 \end{bmatrix}$$

$$\Omega = \begin{bmatrix} 0.54 & 0.08 & 0.04 & 0.40 & 2.40 & 1.35 & 12.22 \\ 0.08 & 6.21 & 4.82 & 5.03 & 0 & 0 & 0 \\ 0.04 & 4.82 & 25.53 & 10.13 & 0 & 0 & 0 \\ 0.40 & 5.03 & 10.13 & 14.19 & 0 & 0 & 0 \\ 2.40 & 0 & 0 & 0 & 182.1 & 78.48 & 102.2 \\ 1.35 & 0 & 0 & 0 & 78.48 & 126.7 & -38.3 \\ 12.22 & 0 & 0 & 0 & 102.2 & -38.3 & 39.98 \end{bmatrix}$$

$^{+}$ Standard errors for the elements of q' are in parentheses. The commodities are ordered corn, oats, and soybeans.

term, however, rarely has much effect on the forecasts, which is attributable to the small covariance between u_{t+1} and v_{t+1}. Expectations of inflation are displayed in Table 4. The first two columns give the four-month actual and expected inflation rates at annualized rates, the third and fourth columns give eight-month actual and expected inflation at annualized rates, and the fifth and sixth columns give the annual actual and expected inflation rates. The first expectation reported in each instance is the one that is conditional on 1979:III information, since this is the period in which the Fed announced its disinflationary objectives. For example, 13.78, the first number in column 6, is the inflation rate expected for 1980 given 1979:III information.

The eight-month-ahead and the one-year-ahead forecast errors strongly imply that a significant part of the disinflation in the early 1980s was unanticipated. Expected inflation exceeds actual inflation in all but three periods in the eight-month forecasts and in every period in the one-year forecast. On average, agents

Table 3 Decomposition of Expectation of Inflation

Period Ending	Actual	Simple Forecast[1]	Price Term[2]	Futures Term[3]	Surprise Term[4]
1970:III	1.77	1.43	0.049	0.0005	0.008
1971:I	1.00	1.38	−0.055	−0.028	0.015
1971:II	1.23	1.56	−0.047	−0.072	0.017
1971:III	0.98	1.54	−0.081	0.046	0.002
1972:I	0.97	1.25	−0.040	0.033	0.013
1972:II	1.19	1.58	−0.055	−0.11	−0.006
1972:III	1.42	1.35	0.010	0.002	−0.011
1973:I	2.77	1.44	0.19	0.17	−0.006
1973:II	2.92	3.28	−0.052	−0.009	−0.023
1973:III	3.48	3.29	0.027	−0.041	0.010
1974:I	3.77	7.84	0.13	0.046	0.004
1974:II	4.03	3.27	0.11	0.065	−0.026
1974:III	3.30	3.53	−0.033	−0.012	0.007
1975:I	1.52	3.15	−0.23	0.021	0.013
1975:II	2.79	2.04	0.11	0.071	−0.011
1975:III	2.17	1.70	0.068	−0.048	0.008
1976:I	1.07	1.80	−0.10	−0.060	−0.012
1976:II	2.11	1.27	0.12	−0.041	−0.014
1976:III	1.89	1.58	0.04	0.023	0.006
1977:I	2.52	2.11	0.060	0.049	−0.010
1977:II	1.81	2.92	−0.16	−0.014	0.032
1977:III	2.26	2.37	−0.016	−0.055	−0.002
1978:I	2.83	2.16	0.10	0.075	−0.008
1978:II	3.05	2.87	0.025	0.049	0.015
1978:III	2.96	3.16	−0.028	−0.007	0.012
1979:I	4.15	3.21	0.14	−0.047	−0.000
1979:II	4.12	3.57	0.078	−0.066	0.003
1979:III	4.73	3.91	0.12	0.072	0.004
1980:I	4.63	3.94	0.10	−0.026	0.001
1980:II	2.66	4.08	−0.20	−0.074	−0.021
1980:III	3.86	3.36	0.072	0.009	0.002
1981:I	2.83	2.73	0.013	0.005	−0.003
1981:II	3.72	2.97	0.118	−0.067	0.011
1981:III	1.39	2.64	−0.18	0.025	0.009
1982:I	1.58	2.18	−0.086	−0.039	−0.011
1982:II	1.86	1.27	0.084	−0.013	0.014
1982:III	0.20	1.57	−0.20	0.080	0.004
1983:I	1.32	1.11	0.030	−0.11	−0.010
1983:II	1.20	0.89	0.044	−0.05	0.021
1983:III	1.68	1.68	−0.001	−0.13	−0.001
1984:I	1.36	1.55	−0.027	0.003	−0.007
1984:II	1.15	1.74	−0.085	0.059	−0.005
1984:III	0.95	1.10	−0.021	0.069	0.001
1985:I	1.41	0.83	0.083	−0.066	−0.006
1985:II	.084	1.13	−0.042	0.099	0.019
1985:III	1.65	1.34	0.045	0.036	−0.007
1986:I	−0.82	1.22	−0.29	−0.031	0.016
1986:II	0.91	1.13	−0.031	0.11	0.015
1986:III	1.35	0.49	0.12	−0.071	−0.019

[1]The simple forecast is equal to $x_t'\hat{\delta}$.
[2]The price term is given by $A_1(p_{t+1} - x_t'\delta)$ in equation (10).
[3]The futures term is given by $A_2(f_t - x_t'\beta)$ in equation (10).
[4]The surprise term is given by $A_3(s_{t+1} - f_t - x_t'\kappa)$.

Table 4 Actual and Expected Inflation: 1980 to 1983

	One period ahead		Two periods ahead		One year ahead	
Period Ending	**Actual**	**Expected**	**Actual**	**Expected**	**Actual**	**Expected**
1980:I	14.19	11.88				
1980:II	13.89	11.88	14.04	12.87		
1980:III	7.98	12.66	10.94	12.80	12.02	13.78
1981:I	11.58	10.17	9.78	13.10	11.15	13.31
1981:II	8.49	8.19	10.02	10.17	9.34	13.91
1981:III	11.16	9.18	9.83	8.88	10.40	10.67
1982:I	4.17	7.83	7.67	8.81	7.93	8.90
1982:II	4.74	6.51	4.44	8.84	6.68	9.10
1982:III	5.58	4.47	5.16	6.03	4.82	9.29
1983:I	0.60	4.59	3.09	4.17	3.64	5.93
1983:II	3.96	2.76	2.28	5.03	3.38	4.23
1983:III	3.60	3.06	3.78	2.11	2.73	5.27

Note: Roman numerals denote four-month intervals.

expected 1.03 percent and 2.23 percent more inflation than actually occurred over the eight-month and one-year forecast horizons, respectively. The one-period-ahead forecast errors do not, however, give as clear a signal. Expected inflation exceeds actual inflation by only 27 basis points on average for the entire period. The combination of these two results implies that the four-month forecasts of inflation four months out and eight months out are drastically overpredicting inflation.[9] That is, even though inflation is declining, agents do not seem to believe that the disinflationary path will continue. Essentially, longer-run expectations appear to be much too static over the period. In this sense, the econometric model indicates that much of the disinflationary path was unanticipated.

5. CONCLUSION

The disinflation that occurred in the early 1980s was a dramatic event in post-World War II monetary policy. Inflation had reached unprecedented heights and was brought down fairly rapidly. It is likely that this disinflation was partially responsible for the two recessions that occurred in rapid succession. The disinflationary policy was announced and underscored by a change in Federal Reserve operating procedures. The policy was carried out over a prolonged period. These two factors could easily lead to the interpretation that

[9] We checked the unbiasedness of the forecasts by regressing Δp_{t+1} on a constant and $E_t \Delta p_{t+1}$. For the four-month and eight-month regressions, we could not reject a zero constant and a unitary slope coefficient. However, the hypothesis that $p_{t+1} = E_t p_{t+1}$ plus a white noise error was rejected at the annual frequency. Therefore, we conclude that greater weight should be placed on the eight-month forecast.

the disinflation was anticipated and, therefore, that this episode is consistent with theories emphasizing the impact of anticipated monetary policy on real economic activity.

Although the policy was announced, evidence documented by Goodfriend (1992) suggests that it may not have been entirely credible. Hence the actual disinflation could still have been unanticipated. Determining the extent to which the disinflation was unanticipated can help ascertain the degree to which Fed credibility was lacking.

In an attempt to resolve whether the disinflation was anticipated, we performed a statistical analysis of the public's expectations of inflation using the methodology developed in Hamilton (1992). Our conclusion is that much of the disinflation was unanticipated and that the Fed suffered from a credibility problem. The fact that much of the disinflation was unanticipated, however, does not allow us to discriminate among the competing models outlined in the introduction. Such discrimination would necessarily involve a more subtle hypothesis test.

REFERENCES

Akerlof, George A., Andrew Rose, and Janet Yellen. "Comments and Discussion," *Brookings Papers on Economic Activity*, 1:1988, pp. 66-75.

Ball, Lawrence. "What Determines the Sacrifice Ratio?" Working Paper No. 4306. Cambridge: National Bureau of Economic Research, March 1993.

_____, N. Gregory Mankiw, and David Romer. "The New Keynesian Economics and the Output-Inflation Trade Off," *Brookings Papers on Economic Activity*, 1:1988, pp. 1-65.

Broaddus, J. Alfred, and Marvin Goodfriend. "Base Drift and the Longer-Run Growth of M1: Experience from a Decade of Monetary Targeting," Federal Reserve Bank of Richmond *Economic Review*, vol. 70 (November/December 1984), pp. 3-14.

Fischer, Stanley. "Contracts, Credibility, and Disinflation," in John W. Neville and Victor E. Argy, eds., *Inflation and Unemployment: Theory, Experience, and Policymaking.* Boston: Allen and Unwin, 1985.

_____. "Wage Indexation and Disinflation," in *Stabilization of the Domestic and International Economy*. Carnegie-Rochester Conference Series on Public Policy, vol. 5 (1977), pp. 107-48.

Goodfriend, Marvin. "Interest Rate Policy and the Inflation Scare Problem: 1979-1992," Federal Reserve Bank of Richmond *Annual Report* (1992), pp. 7-19.

Hamilton, James D. "Was the Deflation during the Great Depression Anticipated? Evidence from the Commodity Futures Market," *American Economic Review*, vol. 82 (March 1992), pp. 157-78.

Lucas, Robert E. "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, vol. 63 (June 1973), pp. 326-34.

McCallum, Bennett T. "Rational Expectations and Macroeconomic Stabilization Policy: An Overview," *Journal of Money, Credit, and Banking*, vol. 12 (November 1980), pp. 716-46.

Sargent, Thomas J., and Neil Wallace. "Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *Journal of Political Economy*, vol. 83 (April 1975), pp. 241-54.

Sims, Christopher A. "Macroeconomics and Reality," *Econometrica*, vol. 48 (January 1980), pp. 1-48.

Taylor, John B. "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, vol. 88 (February 1980), pp. 1-23.

The Fair Lending Laws and Their Enforcement

John R. Walter

Journalists, businessmen, politicians, and regulators are paying increasing attention to the subject of discrimination in lending, with particular emphasis on mortgage lending. It sometimes seems as though the main issues in antidiscrimination efforts have shifted from education and labor markets to the credit markets.

Two different federal laws deal with discrimination in lending: the Fair Housing Act (FHAct) and the Equal Credit Opportunity Act (ECOA). These fair lending laws prohibit lenders from discriminating in credit transactions on the basis of race, color, national origin, religion, sex, and other specified grounds.[1] But the laws provide little practical guidance for enforcement, particularly with regard to the role of the banking regulatory bodies. Congress left to these agencies and to the courts the job of working out the specifics of how to define and promote the laws' purpose—fair lending.

While discrimination has been discussed widely in the popular and professional press, an overview of the two fair lending laws and their enforcement is difficult to find. Section 1 describes these laws, their origins, and the interpretations that the courts and enforcement agencies have given some of

This article has benefited greatly from comments and suggestions from Marvin Goodfriend, Tom Humphrey, Anatoli Kuprianov, Jeffrey Lacker, Roy Webb, and John Weinberg. Alex Mendoza provided excellent research assistance. The views expressed herein are the author's and do not represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Two other laws often mentioned in discussions of fair lending are the Home Mortgage Disclosure Act of 1975 (HMDA) and the Community Reinvestment Act of 1977 (CRA). While both laws play a part in current fair lending enforcement, neither prohibits discriminatory lending, so neither is defined as a fair lending law. The HMDA requires depository institutions to report data that are frequently used in investigations and studies of lending discrimination. The CRA requires the banking agencies to consider a depository institution's efforts to meet the needs of its community when it applies to the banking agencies for permission to expand. As currently interpreted by the agencies, this has meant that a bank's fair lending performance is weighed when considering such an application. For discussions of the HMDA and the CRA, see Canner and Smith (1991) and Lacker (1995), respectively.

their provisions. Section 2 discusses the three major methods of enforcement of the laws: (1) complaints from aggrieved parties leading to investigations by enforcement agencies, (2) civil court actions, and (3) examination by the federal banking agencies.[2] When it comes to their part of enforcing the fair lending laws, the banking agencies follow procedures long used in carrying out their responsibilities for bank safety and soundness. Specifically, they employ periodic fair lending examinations of all banks.[3] Yet the examination of every institution, whether or not the agency suspects discrimination, is strikingly different from the practices of the federal agencies responsible for other areas of anti-discrimination law enforcement. Section 3 discusses the enforcement of other antidiscrimination laws. The absence of periodic exams in other areas of antidiscrimination law enforcement naturally leads one to ask whether routine use of exams for banks is the most efficient means of enforcing the fair lending laws. The concluding section raises briefly some issues that bear on this question.

1. THE FAIR LENDING LAWS

The Fair Housing Act (FHAct) was passed as part of the Civil Rights Act of 1968, which Congress enacted following urban unrest in many U.S. cities in 1965, 1966, 1967, and after Rev. Martin Luther King, Jr.'s assassination in early 1968. According to a Supreme Court opinion, Congress intended the FHAct to contribute to the elimination of ghettos by reducing discriminatory housing practices (*Trafficante v. Metropolitan Life Insurance Company,* 409 U.S. 205 [1972], cited in Board of Governors, *Consumer Compliance Handbook* [1995]).

The FHAct prohibits discrimination in many activities of the residential real estate industry besides lending. The act prohibits discrimination by race, color, religion, sex, handicap, familial status (if a household includes children), and national origin.[4,5] It prohibits the refusal to sell, rent, or negotiate for the sale or rental of housing for discriminatory reasons. Varying the terms of sale or rental in a discriminatory manner is prohibited, as is falsely claiming that housing is not available for inspection, sale, or rental. The act also prohibits real estate

[2] These agencies are the Federal Deposit Insurance Corporation, the Federal Reserve, the Office of the Comptroller of the Currency, the Office of Thrift Supervision, and the National Credit Union Administration.

[3] I will use the term "banks" throughout to refer to commercial banks, thrifts, and credit unions.

[4] As passed, the FHAct prohibited discrimination on the basis of race, color, religion, and national origin. A 1974 amendment added sex as a basis. A 1988 amendment added handicap and familial status.

[5] Unless otherwise noted, I will use the word "discrimination" throughout to mean discriminating on the basis of categories such as race, gender, or familial status.

brokerage organizations, such as multiple-listing services, from discriminating in their terms of access to the organization.

Section 805, the fair lending portion of the FHAct, makes discrimination unlawful in several aspects of home finance. Specifically, it prohibits discrimination in the making or purchasing of loans, the proceeds of which are for purchasing, constructing, improving, repairing, or maintaining a dwelling. The prohibition applies to any person or entity whose business includes engaging in residential real estate-related credit transactions.

The FHAct is enforced by the Department of Housing and Urban Development (HUD), by individuals, and by the Justice Department. Under the act, HUD may take enforcement actions against lenders based on complaints from individuals or on its own initiative. Individuals, or organizations representing individuals, may pursue civil court actions under the act for discrimination against them. A civil action may be brought in a federal court by the Justice Department whenever it believes that a lender is engaged in a pattern of discrimination.

The federal bank regulatory agencies also enforce the act. These agencies are the Federal Deposit Insurance Corporation (FDIC), the Federal Reserve (Fed), the Office of the Comptroller of the Currency (OCC), the Office of Thrift Supervision (OTS), and the National Credit Union Administration (NCUA). Since the act says little about their enforcement responsibilities, the banking agencies have been left to determine largely on their own how to enforce the FHAct.

The fair lending section of the FHAct does not specify actions or policies that are considered discrimination.[6] This means that the agencies responsible for the act's enforcement have little to guide them in the investigation of lending discrimination. At first blush, the identification of lending discrimination might seem a simple matter. The difficulty arises because discrimination, in the broad sense of the word, is a fundamental function of a successful lender. To remain profitable, a lender must avoid (discriminate against) borrowers who are unlikely to repay, and favor those who will repay. Enforcement agencies are left with the difficult job of separating appropriate discrimination from discrimination prohibited by the act. Court decisions have provided broad guidance, but detailed direction is lacking.

In 1972, a U.S. Supreme Court ruling established a procedure for determining whether an individual has been treated differently based on his minority status. Such discrimination goes by the phrase "disparate treatment." The case dealt with employment discrimination, but the courts later applied the precedent to lending discrimination. The Supreme Court set out four factors that a plaintiff

[6] This contrasts with the housing section of the FHAct, which provides an explicit list of actions which are discriminatory under the law.

must establish to make a prima facie case of employment discrimination. The plaintiff must show (1) that he belongs to a racial minority, (2) that he applied and was qualified for a job for which the employer was seeking applicants, (3) that, despite his qualifications, he was rejected, and (4) that, after his rejection, the position remained open and the employer continued to seek equivalently qualified applicants (*McDonnell-Douglas Corp. v. Green*, 411 U.S. 792 [1973], cited in Board of Governors, *Consumer Compliance Handbook* [1995]).

To translate these standards for use in lending cases, the courts simply substituted the idea of creditworthiness qualifications for job qualifications (*Hickson v. Home Federal of Atlanta*, 805 F. Supp. 1567 [N.D. Ga., 1992]; *aff'd*, 14 F.3d 59 [11th Cir. 1994], *Gross v. U.S. Small Business Admin.*, 669 F. Supp. 50 [N.D. N.Y., 1987]). Therefore, to make a prima facie case, the plaintiff must show (1) that he is a member of a protected class, (2) that he attempted to get a loan and met all relevant qualifications for doing so, (3) that the bank refused to make the loan despite the plaintiff's qualifications, and (4) that, after his rejection, the bank continued to make loans to equivalently qualified applicants. Banking agency investigation procedures are based in part on determining whether these factors are met (see Board of Governors, *Consumer Compliance Handbook* [1995], p. 1.19). The courts and the agencies are left to make difficult decisions, however. No two loan applications are exactly alike, so it is impossible to match a rejected minority application perfectly to an accepted non-minority application, and by doing so show discrimination. The courts and agencies must decide whether similar applications are enough like the rejected minority application to persuasively show discrimination. At the very least, however, this ruling provided what the statute lacked—a doctrine to guide policy.

In addition, courts and banking agencies have recognized that the FHAct's prohibitions against discrimination extend to "redlining" (*Laufman v. Oakley Building and Loan Co.*, 408 F. Supp. 489 [S.D. Ohio, 1976]; *Ring v. First Interstate Mortgage, Inc.*, 984 F.2d 924 [8th Cir. 1993]; Cloud and Galster 1993, p. 109; Board of Governors, *Consumer Compliance Handbook* [1995], p. 1.59). Redlining is the practice of denying loans for housing in certain neighborhoods, even if the loan applicants are creditworthy. According to the courts and banking agencies, redlining is unlawful when the decision to avoid making loans in a particular neighborhood is based on the race, national origin, religion, or other similar categorization of residents of the neighborhood. Redlining is not unlawful when based entirely on economic considerations, such as the location of a neighborhood in a flood plain (Interagency Task Force on Fair Lending [1994], p. 5; Board of Governors, *Consumer Compliance Handbook* [1995], pp. 1.58–1.59).

The Equal Credit Opportunity Act also prohibits lending discrimination. This act was passed in 1974 as an amendment to the much broader Consumer

Credit Protection Act, passed in 1968.[7] Specifically, the ECOA prohibits discrimination in all personal and commercial credit transactions based on race, color, religion, national origin, sex, marital status, age, and other bases.[8] The prohibitions apply to anyone regularly extending credit or arranging for the extension of credit. The ECOA is broader than the FHAct since the ECOA covers virtually all lenders while the FHAct covers only real estate-related lending. Housing lenders are subject to both statutes.

As originally passed in 1974, the ECOA only prohibited discrimination based on sex and marital status. Congressional hearings preceding the passage of the ECOA had produced testimony of lending discrimination against women and particularly of women denied credit without the signature of a male (Senate Rep. No. 589, 94th Cong., 2d Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 403, 404 [1976]; Board of Governors, *Consumer Compliance Handbook* [1995], p. 1.12). Additional bases were added by a 1976 amendment. In 1974 hearings, instances were reported of discrimination based on age and race. In addition, during 1974 the banking agencies conducted studies on loan acceptance and rejection rates for minority versus non-minority applicants. Holding creditworthiness factors constant, the studies found much higher rejection rates for minorities (Senate Rep. No. 589, 94th Cong., 2d Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 403, 405 [1976]; U.S. Congress, Senate [1993], pp. 570–71).

Besides prohibiting lending discrimination, the ECOA includes some requirements not explicitly related to fair lending. Within thirty days of an application, a lender must notify the loan applicant whether the application has been accepted or rejected. In the case of a rejected application, a lender must state the reasons for the denial or tell the denied applicant that he may have a statement of the reasons on request. Lenders also must provide, either routinely or at the applicant's request, copies of any appraisal reports.

The ECOA requires the banking agencies to ensure that depository institutions comply with the act's requirements. The ECOA delegates to the Fed the authority to "prescribe regulations to carry out the purposes of the title." These regulations are the Fed's Regulation B. Each regulatory agency is responsible for ensuring that the depository institutions it normally supervises comply with the ECOA and Regulation B. Several other federal agencies are responsible for enforcing compliance by other types of firms that make loans.

[7] The Consumer Credit Protection Act encompasses the Truth in Lending Act, the Fair Credit Reporting Act, the Equal Credit Opportunity Act, the Fair Debt Collection Practices Act, and the Electronic Fund Transfer Act.

[8] Other bases on which lenders may not discriminate include an applicant's status as a recipient of public assistance and an applicant's claim of any right under the Consumer Credit Protection Act.

These agencies include the Federal Trade Commission, the Small Business Administration, the Farm Credit Administration, and others.

The ECOA has left the agencies and the courts with much to work out concerning its enforcement. For instance, like the FHAct, the ECOA leaves largely undefined the actions and policies that are considered discrimination. In addition, the act does not tell the enforcement agencies how to uncover discriminating creditors or discriminating actions. While the FHAct outlines HUD's complaint response procedures, the ECOA does not address such procedures.

Nevertheless, like the FHAct, the ECOA stipulates that any creditor is subject to one of three enforcement actions. First, a person alleging injury may sue the creditor in a federal district court. Second, the federal agency with jurisdiction may take action against the creditor. Third, the Justice Department may pursue a civil action against a suspected violator. Such action may occur if another enforcement agency refers the case to the Justice Department or if the Justice Department suspects a pattern of violations.

Court decisions have helped to define discrimination under the ECOA as they have under the FHAct. In the legislative history of the 1976 amendment to the ECOA, Congress showed that it meant for the act to encompass a new concept of lending discrimination. Discrimination is to include not only disparate treatment (discussed above) but a discrimination standard recently developed by the courts, "disparate impact" (Senate Rep. No. 589, 94th Cong., 2d Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 403, 406 [1976]). Disparate treatment occurs when individuals receive different treatment because of their minority status. Disparate impact occurs when minority individuals and non-minority individuals receive equivalent treatment but a lending policy has a disparate effect on minorities. Making loans for amounts no smaller than a set minimum is the classic example of a policy with a disparate impact. Since, on average, minorities are over-represented in low income brackets, they will more frequently seek low value loans. Therefore, they will be rejected more frequently for loans than will non-minorities under a minimum-loan-amount policy.

The courts developed the disparate impact standard in employment discrimination cases, including a 1971 Supreme Court ruling, *Griggs v. Duke Power Co.* (401 U.S. 424 [1971]). In that case, several black employees of Duke Power Company challenged as discriminatory the company's policy of requiring a high school diploma and a passing score on an intelligence test as a prerequisite for hiring and promotion. The Court ruled that the Civil Rights Act of 1964 outlawed practices that had the effect of discriminating and that Duke's prerequisites had such an effect and were not significantly related to performance of the jobs in question (Board of Governors [1977b], p. 106; *Griggs v. Duke Power Co*. (401 U.S. 424 [1971]).

Following the 1976 amendment to the ECOA, courts extended this rule to lending discrimination. They ruled that a complainant can make a prima facie

case of illegal lending discrimination if he can show that a lender's policy has a disproportionately adverse effect on a protected class. A policy is said to fail the so-called "effects test" if it has a disproportionately adverse effect (*Cherry v. Amoco Oil Co*., 490 F. Supp. 1026 [N.D. Ga., 1980]).

Various courts have come to different conclusions over the burden of proof the lender must carry to rebut the complainant's prima facie disparate impact case. Some courts have ruled that a lender must show a "legitimate business reason" for a practice that produces the disparate impact. Other courts have ruled that the lender must show a "business necessity," a more stringent standard. In 1994 the federal agencies enforcing the fair lending laws came out in favor of requiring the more strict burden for the lender (Interagency Task Force on Fair Lending [1994], p. 7).[9]

2. ENFORCEMENT OF THE FAIR LENDING LAWS

There are three methods of enforcing the fair lending laws. First, the enforcement agencies may take action in response to complaints. Second, individuals or the Justice Department may bring civil court actions. Third, the banking agencies periodically examine every bank for evidence of discrimination. The agencies take remedial or punitive action if evidence of discrimination is found.

Complaints

An individual who believes he has suffered prohibited credit discrimination may complain to (1) the Justice Department, (2) HUD if the loan is for real estate, or (3) the federal agency with fair lending enforcement powers over the institution believed to have discriminated. The FHAct requires HUD to receive and respond to complaints of fair lending violations. The FDIC, the Fed, and the OCC are each required to maintain a consumer affairs division by the Magnuson-Moss Warranty—Federal Trade Commission Improvement Act of 1975 (U.S. Congress, Senate [1976], p. 1).[10] These divisions receive and take action upon consumer complaints, such as complaints of fair lending violations. The Justice Department also responds to lending discrimination complaints that it receives.

Each year these organizations together receive over a thousand complaints of alleged credit discrimination (see U.S. Congress, Senate [1993], pp. 668–69,

[9] For further discussion of the history and current status of the disparate impact standard in lending discrimination, see Vartanian, Ledig, and Babitz (1995), Dennis and Bachman (1995), and Board of Governors of the Federal Reserve System, *Official Staff Commentary on Regulation B, Equal Credit Opportunity* (1995), sect. 202.6(a)-2.

[10] The name of the act derives from its sponsors Senator Warren G. Magnuson and Representative John E. Moss.

680, 700, and 742; U.S. Department of Justice, annual reports of the Attorney General [1977–1987]; U.S. Congress, Senate [1990], p. 119). The agencies investigate each complaint of suspected lending discrimination. If the agencies uncover evidence of discrimination, they may attempt to resolve the complaint by obtaining agreement between the parties. Beyond this, the banking agencies and HUD may require remedial action or may penalize the violator, while the Justice Department can sue. Sometimes one agency refers the case to another agency.

Courts

Based on provisions in both the ECOA and the FHAct, an individual who believes he has suffered prohibited credit discrimination may also seek relief directly from the courts. According to the legislative history of the ECOA, Congress viewed the power of individual claimants to bring civil court cases as a primary mode of enforcement of the statute (Senate Rep. No. 589, 94th Cong., 2d Sess. 2, reprinted in *U.S. Code: Congressional and Administrative News* 403, 415 [1976]).

Private fair housing and civil rights agencies have become more active in facilitating fair lending actions taken by individuals. For instance, one of the first agencies, the Toledo Fair Housing Center, has produced a number of victories for borrowers either in court, in settlements outside court, or through complaints made to the federal enforcement agencies (Cloud and Galster 1993). Similar organizations also have become active recently, producing several large settlements for their clients, and initiating additional suits (Sweet 1995, p. 4; *Washington Post*, September 22, 1995; *American Banker*, January 10, 1995).

Beyond the ability of individuals to bring court cases, the Justice Department may bring civil court actions under both the FHAct and the ECOA. Before 1988, Justice Department enforcement powers were limited under the FHAct and the ECOA; the Justice Department was not granted the power to seek monetary awards. In 1988, an amendment to the FHAct granted the Justice Department the authority to seek monetary awards for damages and civil penalties. A 1991 amendment to the ECOA similarly expanded the types of relief the Justice Department could seek in fair lending cases brought under that statute.

Before these amendments provided the Justice Department with the authority to seek monetary awards, most fair lending actions by the Justice Department resulted in small out-of-court settlements that aroused little popular interest.[11] Since the early 1990s, however, every suit brought by the U.S. Department of Justice has been the focus of great interest, at least among lenders. Each year

[11] At least one lending discrimination case brought by the U.S. Justice Department went to trial (*U.S. v. American Future Systems, Inc.*, 743 F.2d 169 [3d Cir. 1984]).

since 1992 the Justice Department has investigated several fair lending cases resulting in large out-of-court settlements against lenders. In these settlements, the lenders agreed to significant penalties, including compensation for victims, modification of lending procedures, and increased efforts to reach minority individuals or communities.

In civil court cases brought by individuals under the FHAct, the courts may award actual and punitive damages and other equitable relief. The act sets no maximum for these damage awards. In court cases brought by the Justice Department under the FHAct, penalties for violations are the same as in cases brought by individuals, except that civil penalties beyond actual damages are limited to $50,000 for the first violation and $100,000 for any subsequent violations. Under the ECOA, individuals and the Justice Department may seek actual damages and the imposition of injunctions. An injunction is a court order requiring a party to do or refrain from doing a specified act. But lenders may not be made to pay punitive damages to an aggrieved individual greater than $10,000. In class action suits, punitive damages are limited to the lesser of $500,000 or 1 percent of the lender's net worth.

Examination

Examination by the banking agencies provides the third method of enforcing the fair lending laws. The banking agencies probe the institutions they supervise for evidence of lending discrimination in periodic, on-site fair lending examinations. For banks examined by the Fed, a fair lending examination occurs at least every two years. The other agencies have differing schedules.

Fair lending examinations have their origin in, and are modeled after, bank safety and soundness examinations. In a safety and soundness examination, examiners from a federal banking agency investigate a bank's riskiness and financial health. The agencies examine every bank periodically. The examinations include an on-site analysis of the bank's management, its policies and procedures, and its key financial factors. Additionally, examiners verify that a bank is complying with banking laws and regulations. Because of this, examiners gained responsibility for verifying compliance with the fair lending laws when these laws were passed.

Besides the fair lending laws, a number of other banking consumer protection laws, such as the Truth in Lending Act, the Fair Credit Reporting Act, and the Fair Credit Billing Act, were passed by Congress during the 1960s and early 1970s. Examiners gained responsibility for banks' compliance with these laws too. Between 1976 and 1980 the FDIC, the Fed, the OCC, and the NCUA established "consumer compliance" examinations separate from safety and soundness examinations because performing both consumer law compliance and safety and soundness tasks within the same examination was too

burdensome.[12] Examiners who specialize in consumer law compliance (compliance examiners) perform fair lending examinations. The consumer compliance examination covers the fair lending laws and the other consumer protection laws. While separate from the soundness examination, the consumer compliance examination follows the model of periodic, on-site examinations used in the soundness examination.[13]

In the fair lending portion of compliance examinations, the examiner checks for evidence of disparate treatment, redlining, and disparate impact. The examination for disparate treatment of minorities may proceed along several lines. Since the early 1990s, the banking agencies have been using statistical tests to aid them in their search for evidence of discrimination.[14] Compliance examiners may run statistical tests on the outcomes of a bank's loan approval process to decide if minority status is correlated with the frequency of denial, factors other than minority status held equal. When such a correlation is found, examiners manually review a sample of loan application files to determine whether factors omitted by the statistical testing methods explain the statistical results. The main focus of statistical testing is mortgage lending, since minority status data are available only for mortgage loans.

Relatively large numbers of mortgage applications from both minorities and non-minorities are necessary for valid testing. Without large numbers, the tests cannot produce meaningful results. Only one in ten banks receives enough mortgage loan applications to allow the use of these tests.[15] Since these banks tend to be the largest mortgage lenders, large portions of all outstanding mortgage loans are therefore made by banks subject to statistical testing.[16]

---

[12] In January 1989, separate consumer compliance examinations for savings institutions were established by the Federal Home Loan Bank Board (U.S. Congress, Senate [1990], pp. 36–37; U.S. Congress, Senate [1993], p. 693). In late 1989, Congress created the OTS to replace the Federal Home Loan Bank Board as supervisor of savings institutions, following the thrift crisis of the 1980s. The OTS continued the practice of separate consumer compliance examinations.

[13] See Board of Governors (1977a) for a discussion of the history of consumer compliance examination.

[14] In the late 1970s and early 1980s, two of the banking agencies began using functionally similar but less sophisticated tests. The OCC and the FDIC collected data from banks they supervised regarding loans accepted and rejected, race, income, and measures of creditworthiness of the applicant. They then ran statistical analyses that tested for correlations between minority status and rejection, with creditworthiness held equal. The results of these tests guided examiner efforts in searching for evidence of discrimination. While the statistical testing procedures begun in the 1990s are more sophisticated and employ more creditworthiness variables, the methodology and goals are equivalent. The FDIC dropped its use of this technique in 1982 (U.S. Congress, Senate [1990], pp. 44, 155, and 170–71). For discussion of these early techniques, see U.S. General Accounting Office (1981), pp. 50 and 84, and Milroy (1980), pp. 17–33, and 128.

[15] Calem and Canner (1995), p. 121.

[16] For discussions of fair lending testing techniques, see Calem and Canner (1995), Bauer and Cromwell (1994), and Stengel and Glennon (1995).

Examiners use only informal techniques when examining most banks, since formal statistical tests can be used only on a small portion of banks. Fair lending examinations have included informal techniques since the late 1970s. The examiner typically chooses a sample of accepted and rejected applicants and gathers key creditworthiness information for the applicants from the bank's loan application files. The examiner secures a written copy of the bank's lending decision criteria or interviews bank officials to learn the criteria. For example, lending decision criteria might include the bank's minimum down payment percentage, its maximum loan-to-value ratio, and its maximum debt-payments-to-income ratio. The examiner determines whether treatment of accepted and rejected applicants accords with the bank's written or articulated loan criteria. If the bank rejects minorities when its loan criteria suggest acceptance, or accepts non-minorities when the loan criteria suggest rejection, the examiner has reason to suspect discrimination. The examiner will then expand his investigation. For example, assume a bank has a policy of normally rejecting applicants whose debt-payments-to-income ratio will exceed 36 percent if granted the loan. The examiner might become suspicious if this bank allows higher ratios more frequently for whites than for Hispanics. Characteristics of sampled rejected and accepted applicants are also compared directly to determine if the bank accepts non-minority applicants but rejects similarly qualified minority applicants. If this comparison leads the examiner to suspect disparate treatment, he conducts a more intensive investigation.

In searching for signs of disparate treatment by a bank, the examiner also compares the racial and ethnic makeup of the bank's application and loan pools to the racial and ethnic makeup of its market area. If the proportion of applications from minorities differs significantly from the proportion in the bank's market area, the examiner might suspect that the bank is discouraging minorities from applying for loans ("prescreening" minority applicants). Alternatively, a significant differential between the proportion of minority loan approvals and the proportion of minorities in the bank's market area could be a sign that the bank is rejecting minorities for discriminatory reasons. In either case, further investigation would be called for (Board of Governors, *Consumer Compliance Handbook* [1995], pp. 1.43–1.48).

Next, the examiner reviews the bank's loan policies for any that could amount to redlining or for any discriminatory impact. If the examiner finds policies he believes may have such effects, the bank is allowed to present an explanation.

The agencies' responses to violations uncovered during examinations can vary. They may simply require the bank to change its loan policy. They may go further and require it to pay a sizeable monetary penalty either as restitution to injured parties or to the U.S. Treasury. More serious or repeated violations can earn more severe penalties. Whenever an examination uncovers a pattern of fair lending violations, a provision of the ECOA requires the agencies to notify the

Justice Department.[17] While the requirement applies explicitly only to ECOA violations, it covers most FHAct violations as well. Most fair lending violations of the FHAct will be violations of the ECOA, since the former act includes only real estate lending, while the latter covers all types of lending. Some FHAct violations will not be ECOA violations because protected classes differ slightly under the two acts. Once notified, the Justice Department may then choose to investigate and bring a civil action. If the agencies find violations of the FHAct, they must report them either to HUD or to the Justice Department.

3. ENFORCEMENT OF OTHER ANTIDISCRIMINATION LAWS

The first two sections of this article have discussed the fair lending laws and their enforcement. But these fair lending laws are only part of antidiscrimination law. For instance, laws also prohibit discrimination in employment, housing, and voting opportunities. This section contains a brief description of the enforcement activities of several major federal agencies responsible for employment and housing antidiscrimination law. In contrast to the banking agencies' practice of examining all banks for evidence of credit discrimination, the federal employment and housing agencies do not employ routine examinations to enforce antidiscrimination statutes. Instead they investigate institutions only in response to a complaint or other signal of a possible violation. Alternatively, one employment-discrimination law-enforcement agency conducts audits (examinations) of a small subset of institutions. It also responds to complaints.

The major federal agency responsible for employment-discrimination law enforcement investigates employers only if there is prior suspicion of discriminatory activity. The Equal Employment Opportunity Commission (EEOC) is responsible for enforcing four laws prohibiting employment discrimination: the Civil Rights Act of 1964, the Equal Pay Act, the Age Discrimination in Employment Act, and the Americans with Disabilities Act. The EEOC's principal means of enforcing these statutes is through response to complaints of discrimination. These complaints may come from an employee, a rejected job applicant, or an individual or organization acting on an employee's behalf. The EEOC handles complaints by investigating and making a determination on the merits of the complaint. It may seek a settlement between the parties. If the EEOC cannot produce a settlement, it may litigate the case (Bureau of National Affairs 1995, pp. 0:3101–0:3812).

Though complaints are the primary source of EEOC action, the EEOC may itself initiate an investigation. Leads for these charges may come from tips gathered during investigations of other employers. They also may come

[17] This requirement was established by a 1991 amendment to the ECOA.

from the media, other government or private civil rights groups, union and trade associations, and employment agencies. Data from employers on numbers and classifications of minority and non-minority employees can provide a lead (Bureau of National Affairs 1995, p. 8:0002).

Another leading federal employment discrimination enforcement agency is the Office of Federal Contract Compliance Programs (OFCCP), a division of the Department of Labor. The OFCCP is responsible for the enforcement of three antidiscrimination and affirmative action statutes covering institutions receiving government contracts. It enforces the statutes with complaint investigations and compliance reviews ("audits"), which are similar to the banking agencies' examinations.

Although the techniques used in an OFCCP audit are similar to those used in a bank fair lending examination, OFCCP audits only a small subset of contractors.[18] For example, in 1994 it audited 4,100 contractors, selected from a population of 192,500, for compliance with the laws (Office of Federal Contract Compliance Programs [Undated]). Contractors are selected to be audited based in part on the percentage of women and minorities employed relative to comparable percentages employed by the aggregate of all such employers (Fox 1993). Even if no problems are found, an audit is costly for the audited contractor.[19] Therefore, the threat of an audit provides an incentive for contractors to maintain the percentage of women and minorities they employ at or above their peer group's average. This helps the OFCCP meet its affirmative action responsibilities. Other factors are considered when deciding which contractor to audit. These factors are the time elapsed since the last audit, employee or job applicant complaints, negative community group comments, and whether the contractor is adding employees (Fox 1993). Unlike bank fair lending examinations, to which all banks are subject periodically, contractors may go years without an OFCCP audit, if they are audited at all.

The primary federal enforcement agency responsible for housing discrimination is HUD. Like the EEOC, HUD investigations typically result from complaints of discrimination. Occasionally HUD initiates an investigation. For example, some sign of housing discrimination, such as a newspaper story, can generate this type of investigation.

[18] During OFCCP audits, personnel files are examined to determine if comparable minority and non-minority individuals are treated similarly. Personnel policies are examined for any policies that may have a discriminatory effect. The treatment of minorities relative to policies are examined for signs of discrimination.

[19] According to Fox (1993), an audit costs a contractor an average of $25,000.

4. CONCLUDING COMMENT

Banking agencies enforce fair lending laws by examining all banks for evidence of discrimination. As we saw in Section 3, this contrasts sharply with enforcement in other areas of antidiscrimination law, where routine examinations are not employed. The banking agencies' general use of routine exams has evolved as the best means of guaranteeing the safety and soundness of the banking system. But are routine examinations the most efficient means of ensuring fair lending? This is a difficult question.

The examination of every bank, whether or not it is suspected of discrimination, is expensive for both banking agencies and banks themselves. Relying instead on other enforcement mechanisms, such as complaints and legal actions, has the great advantage of directing scarce resources to identifiable problems. Nevertheless, one can imagine several issues that ought to be considered in weighing the costs and benefits of the alternatives. Victims of discriminatory lending practices may not always be aware of the discrimination. If this is the case, then complaints and private lawsuits might not induce sufficient enforcement. On the other hand, enforcement agencies receive hundreds of complaints of discrimination each year. Moreover, the growing interest of housing groups, the press, and attorneys in credit discrimination will expand borrowers' knowledge of the laws that protect them and thus encourage complaints and lawsuits.

Matters are complicated further by the fact that fair lending legal actions seem likely to produce benefits to society beyond those received by the plaintiff alone. For example, policies of a defendant bank that are unknowingly discriminatory may be identified by a lawsuit, and eliminated as a result. Consequently, future customers will be less likely to be victims of discrimination. Since plaintiffs do not reap all the benefits of their own legal actions, the number of court cases brought may be too low from society's point of view. On this basis one might argue that routine examination of all banks, coupled with the threat of penalties, supplements the antidiscrimination benefits provided by lawsuits. But an alternative to employing routine exams to supplement legal actions would be to increase monetary awards to successful plaintiffs or to perform random examinations on a small percentage of banks as done by the OFCCP on government contractors.

From long practical experience, routine exams have been found necessary to ensure the safety and soundness of the banking system. Whether or not routine exams are also the most efficient means of ensuring fair lending is a matter deserving of further research.

REFERENCES

Bauer, Paul W., and Brian A. Cromwell. "A Monte Carlo Examination of Bias Tests in Mortgage Lending," Federal Reserve Bank of Cleveland *Economic Review,* vol. 30 (Quarter 3, 1994), pp. 27–40.

Board of Governors of the Federal Reserve System. *Consumer Compliance Handbook*. Washington: Board of Governors, 1995.

_____. *Official Staff Commentary on Regulation B, Equal Credit Opportunity.* Washington: Board of Governors, 1995.

_____. "Complying with Consumer Credit Regulations: A Challenge," *Federal Reserve Bulletin,* vol. 63 (September 1977a), pp. 769–74.

_____. "Equal Credit Opportunity," *Federal Reserve Bulletin*, vol. 63 (February 1977b), pp. 101–07.

Bureau of National Affairs. *EEOC Compliance Manual,* 1995. Washington: Bureau of National Affairs, 1995.

Calem, Paul S., and Glenn B. Canner. "Integrating Statistical Analysis into Fair Lending Exams: The Fed's Approach," in Anthony M. Yezer, ed., *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics*. Washington: American Bankers Association, 1995.

Canner, Glenn B., and Dolores S. Smith. "Home Mortgage Disclosure Act: Expanded Data on Residential Lending," *Federal Reserve Bulletin,* vol. 77 (November 1991), pp. 859–81.

Cherry v. Amoco Oil Co., 490 F. Supp. 1026 (N.D. Ga., 1980). St. Paul, Minn.: West Publishing Co., 1980.

Cloud, Cathy, and George Galster. "What Do We Know about Racial Discrimination in Mortgage Markets?" *The Review of Black Political Economy,* vol. 22 (Summer 1993), pp. 101–20.

Dennis, Warren L., and Christine DiBacco Bachman. "Are DOJ and FTC Poised for Fair Lending Attack against Credit Scoring?" *Banking Policy Report,* vol. 14 (November 20, 1995), pp. 1, 14–20.

Fox, John C. "1994 Affirmative Action Briefing," Newsletter from the Law Firm of Fenwick and West, Palo Alto, Calif. (October/November 1993).

Gross v. U.S. Small Business Admin., 669 F. Supp. 50 (N.D. N.Y., 1987). St. Paul, Minn.: West Publishing Co., 1988.

Hickson v. Home Federal of Atlanta, 805 F. Supp. 1567 (N.D. Ga., 1992). St. Paul, Minn.: West Publishing Co., 1993.

Interagency Task Force on Fair Lending. "Joint Policy Statement on Fair Lending," March 8, 1994. Washington: U.S. Department of Housing and Urban Development, 1994.

Lacker, Jeffrey M. "Neighborhoods and Banking," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 81 (Spring 1995), pp. 13–38.

Laufman v. Oakley Building and Loan Co., 408 F. Supp. 489 (S.D. Ohio, 1976). St. Paul, Minn.: West Publishing Co., 1976.

Milroy, Michael S. *Compliance Manual: Fair Housing Examinations,* The Consumer Bankers Association. Washington: The Consumer Bankers Association, 1980.

Office of Federal Contract Compliance Programs. Fact Sheet, (Undated). Washington: Office of Federal Contract Compliance Programs, U.S. Department of Labor.

Ring v. First Interstate Mortgage, Inc., 984 F.2d 924 (8th Cir. 1993). St. Paul, Minn.: West Publishing Co., 1993.

Segal, David. "African Americans Accuse NationsBank of Loan Bias," *Washington Post,* September 22, 1995.

Seiberg, Jaret. "Fair-Lending Cases Will Grab Headlines; Lawyers, Activists Say," *American Banker,* January 10, 1995.

Stengel, Mitchell, and Dennis C. Glennon. "Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis," Economic and Policy Working Paper 95-3. Comptroller of the Currency, May 1995.

Supreme Court Reporter. St. Paul, Minn: West Publishing Company, 1971, 1973.

Sweet, William J. "Fair Lending Pressures Spread to Regulations and Private Litigation," *Banking Policy Report,* vol. 14 (February 6-20, 1995), pp. 16–19.

U.S. Code: Congressional and Administrative News. St. Paul, Minn: West Publishing Company, 1976.

U.S. Congress, Senate. *Problems in Community Development Banking, Mortgage Lending Discrimination, Reverse Redlining, and Home Equity Lending*. Hearing, 103 Cong. 1 Sess. Washington: Government Printing Office, 1993.

_____. *Discrimination in Home Mortgage Lending*. Hearing, 101 Cong. 1 Sess. Washington: Government Printing Office, 1990.

_____. *Consumer Protection Enforcement Activities by the Three Commercial Bank Regulatory Agencies*. Senate Report, 94 Cong. 2 Sess. Washington: Government Printing Office, 1976.

U.S. Department of Justice. "The Attorney General's Report to Congress Pursuant to the Equal Credit Opportunity Act Amendments of 1976. Washington: U.S. Department of Justice, 1977-1987.

U.S. General Accounting Office. "Examinations of Financial Institutions Do Not Assure Compliance with Consumer Credit Laws." Washington: U.S. General Accounting Office, January 2, 1981.

U.S. v. American Future Systems, Inc., 743 F.2d 169 (3d Cir. 1984). St. Paul, Minn.: West Publishing Co., 1985.

Vartanian, Thomas P., Robert H. Ledig, and Alisa Babitz. "Revising Regulation B: Codifying the Obvious or Making New Law?" *Banking Policy Report*, vol. 14 (February 6-20, 1995), pp. 16–19.

The Increasing-Returns-to-Scale / Sticky-Price Approach to Monetary Analysis

Mary G. Finn

A new approach to the analysis of the effects of monetary policy on economic activity is developing. Its pioneers are Benhabib and Farmer (1992) and Beaudry and Devereux (1993, 1995). The combined assumptions of increasing returns to scale (IRS) in production and sticky prices identify this approach.[1] The goal is to rationalize slow price adjustment in response to monetary shocks and, consequently, strong and persistent real effects of monetary policy in a fashion consistent with market clearing.

The IRS/sticky-price theory is new, ambitious, and exciting. The traditional approach to sticky prices of the type advanced by Phelps and Taylor (1977) and Fischer (1977) assumes (with no attachment to IRS) that prices are preset for a certain period of time. Thus, when new information on economic conditions arises that was unanticipated when prices were set, those prices are necessarily inconsistent with full optimization on the part of all agents and hence with market clearing. Furthermore, the real effects of money shocks are of short duration, stemming from the short period over which prices are preset. The inconsistency with market clearing and the lack of persistence in the real effects of money are widely viewed as significant weaknesses in the traditional theory. The new theory comes to grips with both of these weaknesses. In striking contrast to the traditional theory, the IRS/sticky-price theory explains how prices are free

The author thanks Jeff Lacker, Yash Mehra, John Weinberg, and Tom Humphrey for helpful comments. The views expressed are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Strictly speaking, the Benhabib and Farmer (1992) model does not maintain IRS in production. Instead, firms' production functions are influenced by an externality factor that depends on the economy's real money balances. This externality factor, like IRS, causes equilibrium indeterminacy, creating the avenue for introducing sticky prices and propagating the effects of monetary shocks. For this reason, the Benhabib and Farmer study is included among those forming the new approach. Moreover, Farmer (1993) views the monetary externality as a proxy for IRS in production.

to change sluggishly over time. The explanation is consistent not only with optimization by all agents and market clearing but also with prolonged real effects from money. Also, the IRS/sticky-price theory marks initial progress in incorporating monetary policy into the recent prominent research area featuring the importance of IRS for explaining business-cycle fluctuations (see, for example, Rotemberg and Woodford [1995] for a review of this research).

The essential ideas of the IRS/sticky-price theory may be described as follows: IRS amplify the economy's response to shocks arising from any source. When large enough, IRS cause the economy's equilibrium to become indeterminate. The assumption of sticky prices resolves this indeterminacy, and in doing so, perhaps surprisingly, allows markets to clear. The combination of the assumptions of IRS and sticky prices is a powerful force that produces slow price adjustment and, therefore, significant as well as prolonged nonneutral effects from money shocks.

The purpose of this article is to access and to explain, as nontechnically as possible, the ideas of the IRS/sticky-price theory.[2] The discussion focuses only on the Beaudry and Devereux (1995) model (the BD model).[3,4] Not only is the BD model representative of the models in its class, but its quantitative implications have been the most fully developed, which enables it to be evaluated in terms of its ability to explain empirical evidence.

The remainder of the article is organized as follows: Section 1 considers the assumption of IRS, Section 2 discusses the sticky price assumption, Section 3 explains how sluggish price adjustment and monetary nonneutrality occur, Section 4 compares the IRS/sticky-price theory's predictions with the facts, and Section 5 presents some conclusions.

1. THE ASSUMPTION OF INCREASING RETURNS TO SCALE

IRS, which work through an externality factor in the individual production functions of firms that produce intermediate goods, magnify the parameters in the relationship between aggregate final output and the aggregate amounts of factors of production. In so doing, they amplify the response of output to the fluctuations in productive factors that are induced by exogenous shocks. To see more exactly how IRS work, consider the following: an intermediate

[2] The article draws partly on the author's discussion in Finn (1995). See Salyer (1995) for a review of IRS models that do not have money or sticky prices.

[3] The models in the studies forming the new approach, listed in the text, do differ in specification, including the dimensions involving IRS and the transactions role of money. But the logic of the explanations and assessments presented in the article easily extends to those models.

[4] There are two perfectly symmetrical economies in the BD model. For simplicity of exposition, this article's discussion pertains to one of these economies.

goods firm produces a unique intermediate good i according to the production function F:

$$x_i = F(k_i, l_i, \Lambda) = k_i^{\alpha} l_i^{(1-\alpha)} \Lambda \quad , \quad 0 < \alpha < 1, \tag{1}$$

where x_i is the output of intermediate good i, $k_i(l_i)$ is the input of capital (labor) into the production of intermediate good i, Λ is the externality factor, and α is a parameter. (For simplicity, fixed costs of production and imported capital, present in the BD model, are ignored here.) F exhibits constant returns to scale with respect to the individual firm inputs, k_i and l_i. The externality, exogenous to the individual firm because it depends on aggregate economic activity, is determined by

$$\Lambda = [K^{\alpha} L^{(1-\alpha)}]^{\gamma - 1} \quad , \quad \gamma > 1, \tag{2}$$

where K and L are the aggregate amounts of k_i and l_i, respectively, employed in the economy, and γ is the IRS parameter. The number of intermediate goods firms measures unity so that aggregate input equals per-firm input of each factor of production. Intermediate firms' output is combined by many final goods firms to produce the final good. The production function linking the aggregate output of the final good, denoted by X, to intermediate goods is

$$X = [\int x_i^{\rho} di]^{1/\rho} \quad , \quad 0 < \rho < 1, \tag{3}$$

where ρ is a parameter determining the elasticity of substitution in production between the intermediate inputs $1/(1 - \rho)$. The relationship in equation (3) displays constant returns to scale.

In equilibrium, all intermediate goods firms behave identically. That is, they employ identical amounts of the factors of production which, as mentioned above, coincide with K and L; and they produce the same amount of output, denoted by x. Noting these results in equations (1) through (3), it follows that in equilibrium the relationship between aggregate final output and aggregate factors of production is

$$X = x = F(K, L, \Lambda) = K^{\alpha} L^{(1-\alpha)} \Lambda = [K^{\alpha} L^{(1-\alpha)}]^{\gamma}. \tag{4}$$

Since $\gamma > 1$, this relationship shows IRS. The IRS can now be seen to work through the externality factor, Λ, to increase the parameters linking aggregate final output, X, to each of the aggregate factors of production, K and L. Accordingly, X is more responsive to K and L than it otherwise would be. Two important issues arise.[5]

[5] Beaudry and Devereux (1993) assume that IRS are internal to the individual intermediate goods firm. The internal IRS, just like the external IRS, work to magnify the parameters linking X to K and L. Benhabib and Farmer (1992) do not assume IRS in production. Individual firm production functions are affected, though, by an externality factor, depending on the economy's real money balances (see footnote 1 for more detail). The essence of the two discussion issues, following in the text, applies to Beaudry and Devereux (1993) and Benhabib and Farmer (1992).

Why IRS?

What explains the existence of an externality factor in individual firms' production functions that depends on aggregate economywide employment of capital and, especially, of labor?[6] For instance, how could an increase in labor employment by a motor engine-producing firm simultaneously enhance productivity in a paper-producing firm? The studies forming the IRS/sticky-price approach provide no discussion of these questions.

The most obvious possible answer invokes the learning-by-doing idea (Arrow 1962). According to this idea, the people involved in any production activity learn more efficient ways of doing it. Sometimes such learning is easily disseminated and applied to other production activities. When this happens, the higher the aggregate production activity in the economy, the higher the productivity/efficiency of any individual firm's production method. But learning-by-doing tends to occur and is disseminated fairly slowly over time. Hence, it is better captured by an externality factor that depends on the economy's slowly changing aggregate capital stock than by one that depends on the economy's quickly changing aggregate labor employment. That is, it is difficult to see how learning-by-doing would lead to the magnification of labor input's exponent in the aggregate production function.

Future research may provide a clear-cut answer to the above question. It is an important one, and an answer seems necessary before confidence can be placed in the IRS/sticky-price approach to analyzing monetary policy. That approach hinges sensitively on the size of labor's exponent in the aggregate production function, which will be explained next.

The Importance of the Magnitude of IRS

The IRS/sticky-price theory of the effects of monetary shocks crucially depends on the size of the IRS parameter, γ. Two points are at stake. First, γ must be sufficiently big to cause equilibrium indeterminacy. It is only when indeterminacy arises that the sticky price assumption can be made, thereby rendering the equilibrium determinate. Furthermore, sticky prices are at the source of the powerful nonneutral effects of monetary shocks. Second, γ must not be too big because, for high values of γ, the volatility of real variables is decreasing in the magnitude of γ. Thus, the nonneutral effects of monetary shocks diminish as γ rises. These two points are more fully explained in this section.

An essential component of the IRS/sticky-price theory is that equilibrium is indeterminate prior to imposing the sticky price assumption. A necessary condition for this indeterminacy is that, at the level of the aggregate economy,

[6] Prescott (1989) raises questions and arguments similar to those presented here.

labor's marginal productivity is increasing in labor. That is, the production function in equation (4) must satisfy the parametric restriction

$$\begin{aligned} (1-\alpha)\gamma > 1 \qquad & \text{or} \\ \gamma > \underline{\gamma} \equiv 1/(1-\alpha). & \end{aligned} \tag{5}$$

Strong empirical evidence gives a value for labor's output share, $(1-\alpha)$, that is close to 0.70 (see, for example, Lucas [1990]). Using 0.70 as an example, the condition in equation (5) is $\gamma > \underline{\gamma} = 1.43$.

Some intuition for the necessity of rising marginal labor productivity is obtained by following the BD model in considering behavior in the labor market under two alternative assumptions regarding (aggregate) returns to scale—IRS and constant returns to scale (CRS). Figure 1 shows the labor market situation for both cases.

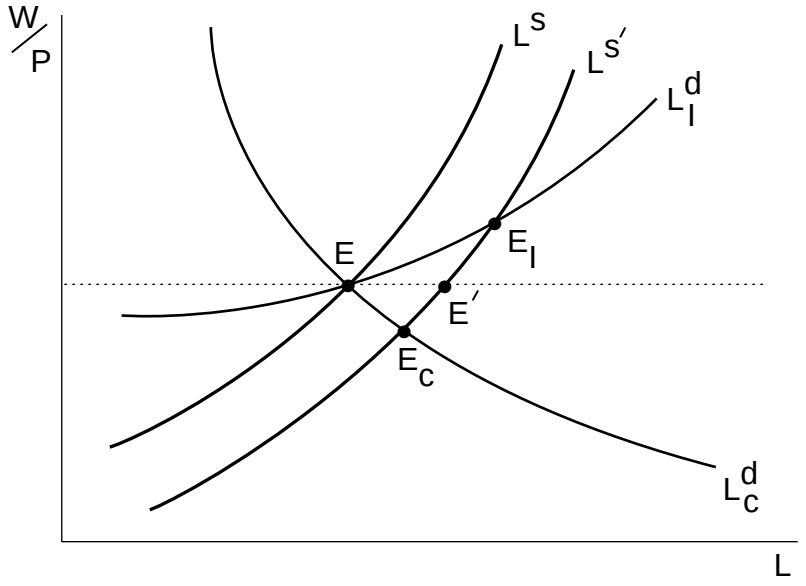
The labor supply curve, L^s, reflects agents' desire to provide an amount of labor ensuring equality between the real wage W/P and the marginal rate of substitution between leisure and consumption. An increase in W/P, which represents the opportunity cost of leisure in terms of consumption, generates a substitution effect away from leisure and toward labor, output, and consumption. Therefore, L^s is positively sloped and as labor rises along L^s, so does consumption. One important variable that can shift L^s is investment. In particular, an increase in investment, by setting in motion a force to sharply reduce the amount of output available for consumption, operates on labor incentives analogously to a negative wealth effect. Thus, at any given real wage rate, an increase in investment is accommodated by a rise in labor and output and a mitigated decline in consumption. Graphically, investment increases shift L^s to the right, and as it does, at any given real wage, consumption falls.[7]

The labor demand curves are L_c^d and L_I^d under the assumption of CRS and IRS, respectively. Both curves capture firms' decisions to hire labor up to the point that equates labor's marginal product to the real wage. With CRS (i.e., $\gamma = 1$ in equation [4]), the marginal productivity of labor is declining in labor, giving L_c^d its negative slope. But with sufficiently high IRS (i.e., γ satisfying equation [5]), labor's marginal productivity increases with increases in labor, imparting a positive slope to L_I^d.

Initially, the labor market is in equilibrium at E. Now consider an experiment in which firms suddenly wish to increase investment even though nothing fundamental has changed. The question is whether this desired increase in investment could be supported by a new equilibrium that is different from E. An increase in investment would shift L^s to the right, say to $L^{s'}$, giving rise

[7] If the marginal utility of leisure is a constant (invariant to the level of leisure) then, at any given real wage, consumption will stay constant as investment increases. In that case, at any given real wage, labor rises enough to cause output increases that exactly match the increases in investment.

Figure 1 Labor Market with IRS and CRS



to the intersection points E', E_c and E_I. Consumption is lower at E_c than at E (because consumption is lower at E_c than at E' and lower at E' than at E). This lower consumption would be consistent with a new equilibrium at E_c only if the real interest rate were higher at E_c than at E; a rise in the real interest rate reduces consumption because of intertemporal substitution. But any such rise in the real interest rate would quench the initial desired investment increase. Consequently, the CRS economy has a unique equilibrium at E. For the IRS economy, the situation is different. Consumption is higher at E_I than at E. (L_I^d is assumed to be sufficiently upward sloping so that the rise in consumption from E' to E_I dominates the fall in consumption from E to E'.) This higher consumption would be consistent with a new equilibrium at E_I if the real interest rate were lower at E_I than at E. A decline in the real interest rate operates to validate the desired increase in firms' investment. Thus, the IRS economy could achieve a new equilibrium at E_I. Multiple equilibria are possible for the IRS economy: the IRS economy's equilibrium is indeterminate.

The second point about the size of γ is that it must not be too big, because in the range satisfying equation (5), the volatility of real variables is decreasing in the size of γ (Beaudry and Devereux, Tables 2 and 3). Intuitively, increases in γ exert two opposing forces on volatility. First, for any given change in the amounts of employed capital and labor, a rise in γ implies a larger change

in output and, hence, in most other real variables, such as consumption and investment. Second, when γ satisfies equation (5), a higher value of γ causes the aggregate marginal productivity of labor to rise more quickly with any increase in labor. Consequently, smaller changes in labor and hence in capital, output, and most other real variables are needed to equilibrate the economy in response to a shock. The second force is the dominant one (again when γ satisfies equation [5]), so the volatility of real variables decreases as γ decreases.

The result of the two points concerning the magnitude of γ is that the IRS/sticky-price theory requires γ to lie in a narrow range above $\underline{\gamma}$. This outcome is somewhat troubling for the theory, at least given the current state of knowledge. Existing empirical evidence is far from precise. Studies by Hall (1990), Caballero and Lyons (1992), Eden and Griliches (1993), Basu and Fernald (1994), and Burnside, Eichenbaum, and Rebelo (1995) give point estimates of γ ranging from one to ten. But perhaps by being explicit about the microtheoretic foundations of IRS, future research will develop more precise estimates of γ or will extend the theory in a way that lessens its dependence on the size of γ.

2. THE STICKY PRICE ASSUMPTION

Once indeterminacy of equilibrium arises because of IRS, the assumption of sticky prices is made to eliminate that indeterminacy. Alternatively viewed, the sticky price assumption selects one of the multiple equilibria engendered by IRS. This section provides an intuitive explanation of how sticky prices achieve determinacy of equilibrium and discusses two characteristics of this particular approach to sticky prices.[8]

The intermediate goods market is monopolistically competitive. Each intermediate goods firm chooses the optimal price of its good each period according to the constant markup rule

$$p_t = (1/\rho)MC_t \quad , \quad MC_t = \frac{[\hat{R}_t^{\alpha}\hat{W}_t^{(1-\alpha)}]A}{\Lambda_t} \quad , \quad A > 0, \tag{6}$$

where p is the nominal price of any intermediate good, MC is the nominal marginal cost of producing any intermediate good, $\hat{R}$ is the nominal rental price of capital that firms pay, $\hat{W}$ is the nominal wage of labor faced by firms, A is a parameter, and subscript t denotes time. The marginal cost is increasing in an index of factor prices and decreasing in the externality factor that enhances the productivity of firm inputs. Noting that the absolute value of the

[8] The discussion in this section extends easily to Beaudry and Devereux (1993), where IRS are internal to firms, and to Benhabib and Farmer (1992), where the production externality is a monetary one and where firms are perfectly competitive price takers.

price elasticity of demand for intermediate goods (by final goods firms) is the same as the elasticity of substitution of intermediate goods, $1/(1-\rho)$, and is thus increasing in ρ, equation (6) shows that the markup is inversely related to that price elasticity. Equation (6) also shows that all intermediate goods firms choose the same price because marginal cost is identical across firms.

The final goods market is perfectly competitive. There, because final goods firms choose their optimal output each period, the nominal price of the final good, denoted by P_t, always equals the nominal marginal cost of its production, which can be shown to equal p_t. Therefore

$$P_t = p_t. \tag{7}$$

The sticky price assumption imposes the requirement that P_t is unresponsive or rigid with respect to any new information that becomes available at time t. That is, P_t is assumed to be predetermined at time t. Following equations (6) and (7), p_t and MC_t must also be predetermined at time t. Thus, with the sticky price assumption in place, any unanticipated movement in the externality factor at time t, Λ_t has to be matched by an equal movement in the factor price index, $[\hat{R}_t^{\alpha}\hat{W}_t^{(1-\alpha)}]$ to keep equation (6) intact. This also means that real return to capital, or the real interest rate, rises at time t if an unanticipated expansionary force increases the economy's output, and thus Λ, at time t. One example of such an unanticipated expansionary force is the sudden desire to increase investment that was considered in the experiment of Section 1. Rerunning that experiment for the IRS economy under conditions of sticky goods prices shows that the desired investment increase is vitiated by the increase in the real interest rate. Therefore, the IRS economy with sticky prices has a unique equilibrium. The assumption of sticky prices is responsible for ensuring the determinacy of equilibrium.

A unique feature of the sticky price assumption is that it is made at the last stage of the modeling procedure. All other assumptions (for instance, those describing preferences, production technologies, and market structure) are made, in standard fashion, at the beginning of the model specification. Next, the equations defining the economy's equilibrium are deduced. If that equilibrium is indeterminate, the last step of the modeling process is to obtain determinacy of equilibrium by making the sticky price assumption. This unique feature is important for at least two reasons.

First, to arrive at the point of assuming sticky prices, the economy's structure must give rise to equilibrium indeterminacy. This is somewhat of a drawback because the required structure is not necessarily easy to justify on theoretical and empirical grounds. In the IRS/sticky-price theory case, a strong stand needs to be taken on the nature and magnitude of IRS, as discussed earlier.

Second, the approach to sticky prices described above is consistent with all firms optimizing and, consequently, with market clearing. This aspect is

appealing. As mentioned before, all firms choose plans that maximize profit every period. In particular, intermediate goods firms always choose their prices according to their optimal markup pricing rule (equation [6]). By imposing sticky prices on those optimization rules, optimization continues to hold. Curiously, one may imagine, even though it is not literally true, that the intermediate goods firms are choosing prices one period in advance (i.e., choosing p_t at time $t-1$), because firms never want to adjust p_t to information that is new at time t.

This approach to sticky prices sharply contrasts with the approach taken in traditional macroeconomic models (e.g., Phelps and Taylor [1977]; Fischer [1977]). The traditional approach does not require initial equilibrium indeterminacy, and it is inconsistent with full optimization and market clearing. In that approach, by assumption, prices truly are set in advance. Consequently, when new, unanticipated information on demand and supply conditions emerges, those prices are inconsistent with all agents optimizing and thus with market clearing.

3. SLUGGISH PRICE ADJUSTMENT AND MONETARY NONNEUTRALITY

The combination of the IRS and sticky price assumptions acts as a powerful force to engender sluggish price adjustment and, therefore, monetary nonneutrality. To explain how this force operates, this section analyzes the effects on the economy of a permanent increase in the money stock.[9]

The fundamental role of money is to mitigate the transactions costs arising from the process of allocating factors of production to production activity. More exactly, the transactions costs are increasing in the economy's employment of capital and labor and decreasing in real money balances as captured by the relationship

$$\phi(N/P, X) = B(N/P)^{(1-\nu)}X^{\nu} \quad , \quad B > 0 \quad , \quad \nu > 1, \tag{8}$$

where ϕ is transactions costs (measured in units of output), N is the money stock, B and ν are parameters, and X is directly related to factor employment

[9] The main modification required to apply this explanation to Beaudry and Devereux (1993) is simply to recognize the different transactions role for money in that study. The consequential different effect of a permanent money stock increase is that nominal interest rates fall in Beaudry and Devereux (1993) while they stay approximately constant in the BD model. The explanation of monetary nonneutrality is quite different across the BD and Benhabib and Farmer (1992) models, primarily because the production externality is monetary (and not real) in nature and there is no investment in the latter model. The important point common to both explanations is that the production externality ensures that the money supply increase leads to such a sizable expansion of output (at predetermined current prices) and the real demand for money that money market equilibrium is reestablished by tiny and sluggish increases in future price inflation.

according to equation (4). Because of these transactions costs, households receive payments for providing factor services that fall short of the marginal productivities of those services. In fact, it is straightforward to show that the factor prices received by households are

$$W = MC(MP_L)[1 - \phi_2(N/P, X)] \quad \text{and} \tag{9}$$

$$R = MC(MP_K)[1 - \phi_2(N/P, X)], \tag{10}$$

where W is the nominal wage of labor received by households; R is the rental price of capital received by households; MP_L and MP_k are the marginal productivities of L and K in producing the output of the typical intermediate goods firm; ϕ_2, the marginal transactions cost of increasing output, denotes the partial derivative of ϕ with respect to X. Equations (9) and (10), together with the properties of ϕ_2, make clear that transactions costs shrink the value of factor payments to households while money holdings help mitigate the extent of that shrinkage.

If the monetary authority permanently increases the money stock at time t, what will happen at time t? With P_t predetermined at time t, by assumption, the effect of a rise in N_t is a proportionate rise in N_t/P_t. Consequently, the marginal transactions cost of hiring all factors in goods production falls (i.e., ϕ_2 declines), implying an increase in W_t and R_t (see equations [9] and [10]). This increase in factor payments invokes a rise in households' desire to supply factors to intermediate goods firms. The firms willingly hire the desired increase in factor supply—in a sense they are indifferent because, as explained earlier, both the price and marginal cost of their goods stay constant. The resultant increase in K_t and L_t expands the output of all goods; thus, X_t rises. On the opposite side of the supply response is the demand response. The increase in N_t/P_t, by raising household wealth, causes an expansion in spending for both consumption and investment purposes. Denoting aggregate consumption and investment of the final good by C_t and I_t, respectively, the rise in X_t is absorbed by the rise in C_t and I_t. Because of the presence of IRS, all of these expansions in factor employment, output, consumption, and investment are sizable. Furthermore, IRS together with the large increase in (I_t implies that all of these real expansionary effects are strongly propagated into the future.

The sticky price assumption makes current prices predetermined only at time t. Future prices are perfectly free to respond to the money supply increase, starting from time $(t + 1)$. But they do so gradually, and P_{t+1} responds hardly at all. P_{t+1} is essentially determined by its role in ensuring equilibrium in the money market at time t. The money market equilibrium condition is

$$N_t/P_t = G(X_t, i_t) \quad , \quad G_1 > O \quad , \quad G_2 < 0, \tag{11}$$

where G is a function, G_j is the first derivative of G with respect to its jth argument, $(j = 1, 2)$, and i_t is the net nominal interest rate. This condition

sets money demand equal to money supply. Stemming from the role of money in alleviating the transactions costs arising in production activity, real money demand is increasing in X_t. The opportunity cost of money is the forgone nominal interest rate on bonds and so the demand for real money balances is declining in i_t. Equation (11) captures both of these effects on real money holdings. Furthermore, an intertemporal efficiency condition governing optimal nominal (discount) bond holdings in the economy gives rise to the equilibrium asset-pricing relationship

$$(1 + i_t) = \left(\frac{P_{t+1}C_{t+1}}{P_tC_t}\right)\beta^{-1}. \tag{12}$$

The gross nominal interest rate at time t is the product of price inflation and consumption growth between time t and time $(t + 1)$ and the reciprocal of the subjective discount factor β.[10]

While P_{t+1} can in principle respond to the increase in N_t, there are two reasons why its response is negligible. First, the expansion of X_t is large enough relative to that in (N_t/P_t) to create a small incipient excess demand for real balances. Second, the strong increase in I_t propagates the increase in X_t so strongly into the future that consumption grows slightly for a time. The small increase in C_{t+1}/C_t causes a small rise in i_t that is sufficient to eliminate the incipient excess demand for money. Consequently, P_{t+1} does not need to change significantly.

IRS and capital accumulation act to strongly propagate, though at a diminishing rate, the real expansionary effect of the money supply increase well into the future. Similar reasoning explains that prices at time $(t + 2)$ and beyond are also slow to change. It is only after about four periods (at time $t + 4$), as consumption growth slows and becomes negative, that P eventually and slowly increases. At any time, the increase in future P keeps i approximately constant in face of the small negative consumption growth rates because an approximately constant i is sufficient to ensure money market equilibrium. Meanwhile, the increase in current P engineers the reductions in N/P that, each period, closely match the declines in X as the economy returns to its steady state. In the steady state, all real variables take on their original values—their values prior to the increase in N_t—and P rises in the same proportion as N_t. That is, the steady state is neutral with respect to the increase in N_t. Thus, the sluggish price adjustment and consequential monetary nonneutrality are short-run phenomena.

[10] The utility function here is assumed to be an additively separable and logarithmic function of consumption. Also, for simplicity, the expectations operator is omitted.

4. EXPLAINING THE FACTS

How well does the IRS/sticky-price theory explain the empirical evidence? To address this question, this section evaluates the theory along four key dimensions—price, output, nominal interest rate responses to monetary shocks, and the correlation between productivity and employment.

Sluggish Price Adjustment

Empirical evidence indicates that the price level exhibits a slow and lengthy response to monetary innovations.[11] The theory well captures this evidence. However, it is not entirely clear that the theory's ability to mimic the data along this dimension should be viewed as an explanation. By assumption, prices cannot contemporaneously respond to monetary shocks. Furthermore, the endogenous propagation mechanism delivers sluggish price adjustment in such a direct fashion that it is almost tantamount to building in sluggish price adjustment over time by assumption. Admittedly this is somewhat of a fine point, but it does, at least, prompt less serious evaluation of the theory along the sluggish price dimension and more serious evaluation along other dimensions.

Output Response

Stemming primarily from the assumptions of IRS and sticky prices, the BD model has strong implications for the magnitude of the output response to a money shock.[12] For example, the model implies that a 1 percentage point increase in the money stock contemporaneously causes a 7 percentage point rise in real output. An output response of this magnitude seems much too big. Sims's (1980) evidence for the U.S. economy over the postwar period shows that a 1 percent money stock innovation within the same period causes only a 0.42 percent increase in output. Part of the model's excessive contemporaneous output response may be due to the manner in which IRS works. Specifically, by magnifying labor's exponent in the production function, IRS permits the economy to instantly and significantly respond to shocks from any source.

Nominal Interest Rate Response

As described earlier, the BD model predicts essentially no nominal interest rate response to monetary shocks, which seems counterfactual. Strong evidence documents the liquidity effect—the negative effect on nominal interest rates exerted by positive monetary innovations (Cochrane 1989). However, the

[11] See, for instance, Sims (1980, 1989) for evidence pertaining to the United States.

[12] This is also true of the Beaudry and Devereux (1993) model. The same strong implications hold for the Benhabib and Farmer (1992) model, but in that case they stem from the assumptions of the monetary externality and sticky prices.

model's prediction along this dimension is not necessarily a prediction of the IRS/sticky-price theory generally. In particular, for a slightly different specification of the transactions role of money in Beaudry and Devereux (1993), the liquidity effect is indeed predicted.

Employment-Productivity Correlation

The IRS/sticky-price theory crucially rests on IRS as the cause of increasing marginal productivity of labor.[13] This phenomenon, in turn, implies that labor and its productivity should always move in the same direction. In fact, they do not. At business-cycle frequencies the correlation between labor and its productivity is close to zero (Christiano and Eichenbaum 1992). This inconsistency between the theory and the facts is important.[14] Just like the inconsistency involving the output response, it seems to underscore the importance of developing the microtheoretic foundations of IRS.

5. CONCLUSION

The IRS/sticky-price theory is a new approach to the analysis of the effects of monetary policy on the economy. Its goal is to explain, within a market-clearing framework, why prices are slow to respond to monetary shocks and, therefore, why monetary shocks can have strong real effects. Several key points are explained and assessed in this article.

First, it is not clear how IRS can in principle be justified, especially as it pertains to labor input into production. Also, existing empirical estimates do not provide solid support for placing the magnitude of IRS into the narrow range required for the theory to work. To be convincing, the IRS/sticky-price theory needs to address both of these issues. Perhaps by making explicit the theoretical foundations of IRS, future research will develop a theory that lessens its dependence on the precise magnitude of IRS or else will provide estimates that support the theory as it now stands.

Second, a fine line exists between viewing the IRS/sticky-price theory as explaining or assuming the slow adjustment of prices in response to monetary shocks. By assumption, prices are sticky; that is, they cannot contemporaneously respond to shocks. While prices are free to respond over time, the endogenous propagation mechanism produces slow price adjustment in a direct fashion—so much so that it comes close to assuming a slow price response. The existence of this fine line makes it important to evaluate the theory along dimensions different from sluggish price responses to money shocks.

[13] It is not clear whether the Benhabib and Farmer (1992) model incorporates increasing marginal productivity of labor.

[14] Prescott (1989) points out the same inconsistency.

Third, in its current state, the IRS/sticky-price theory predicts huge output responses to monetary shocks and correlations between labor and its productivity that are noticeably counterfactual. This inconsistency between the theory's predictions and the facts underscores the importance for the theory to develop the theoretical rationale for IRS.

The IRS/sticky-price theory is innovative and ambitious in tackling the difficult issue of explaining the effects of money on the economy. The points raised in this article suggest, on the one hand, that it would be premature to use this theory as a basis for understanding the monetary transmission mechanism. In particular, the theory is too inconsistent with key facts to be useful for that purpose. On the other hand, the article suggests that it may be fruitful to explore this theoretical approach further by developing the theory underlying IRS.

REFERENCES

Arrow, Kenneth J. "The Economic Implications of Learning by Doing," *Review of Economic Studies,* vol. 29 (June 1962), pp. 155–73.

Basu, Susanto, and John G. Fernald. "Constant Returns and Small Markups in U.S. Manufacturing," in *International Finance Discussion Papers,* No. 483. Washington: Board of Governors of the Federal Reserve System.

Beaudry, Paul, and Michael B. Devereux. "Money and the Real Exchange Rate with Sticky Prices and Increasing Returns," Carnegie-Rochester Conference Series on Public Policy, vol. 43 (December 1995).

_____. "Monopolistic Competition, Price Setting, and the Effects of Real and Monetary Shocks." Manuscript. University of British Columbia, October 1993.

Benhabib, Jess, and Roger E. A. Farmer. "The Aggregate Effects of Monetary Externalities." Manuscript. University of California at Los Angeles, April 1992.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. "Capital Utilization and Returns to Scale." Manuscript. University of Rochester, February 1995.

Caballero, Ricardo J., and Richard K. Lyons. "External Effects in U.S. Procyclical Productivity," *Journal of Monetary Economics,* vol. 29 (April 1992), pp. 209–25.

Christiano, Lawrence J., and Martin Eichenbaum. "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations," *American Economic Review,* vol. 82 (June 1992), pp. 430–50.

Cochrane, John H. "The Return of the Liquidity Effect: A Study of the Short-Run Relation Between Money Growth and Interest Rates," *Journal of Business & Economic Statistics*, vol. 7 (January 1989), pp. 75–83.

Eden, Benjamin, and Zvi Griliches. "Productivity, Market Power, and Capacity Utilization When Spot Markets Are Complete," *American Economic Review*, vol. 83 (May 1993), pp. 219–23.

Farmer, Roger E. A. *The Macroeconomics of Self-Fulfilling Prophecies*. Cambridge, Mass.: MIT Press, 1993.

Finn, Mary G.. "Increasing Returns to Scale and Sticky Prices: Are They Important for Real Exchange Rates? A Comment," Carnegie-Rochester Conference Series on Public Policy, vol. 43 (December 1995).

Fischer, Stanley. "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule," *Journal of Political Economy*, vol. 85 (February 1977), pp. 191–205.

Hall, Robert E. "Invariance Properties of Solow's Productivity Residual," in Peter Diamond, ed., *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's Birthday*. Cambridge, Mass.: MIT Press, 1990.

Lucas, Robert E. "Supply-Side Economics: An Analytical Review," *Oxford Economic Papers*, vol. 42 (April 1990), pp. 293–316.

Phelps, Edmund S., and John B. Taylor. "Stabilizing Powers of Monetary Policy under Rational Expectations," *Journal of Political Economy*, vol. 85 (February 1977), pp. 163–90.

Prescott, Edward C. "Comment on Building Blocks of Market Clearing Business Cycle Models," *Macroeconomics Annual*, National Bureau of Economic Research, 1989, pp. 287–91.

Rotemberg, Julio J., and Michael Woodford. "Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets," in Thomas Cooley, ed., *Frontiers of Business Cycle Research*. Princeton, N.J.: Princeton University Press, 1995.

Salyer, Kevin D. "The Macroeconomics of Self-Fulfilling Prophecies: A Review Essay," *Journal of Monetary Economics*, vol. 35 (February 1995), pp. 215–42.

Sims, Christopher A. "Models and Their Uses," *American Journal of Agricultural Economics*, vol. 71 (May 1989), pp. 489–94.

_____. "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," *American Economic Review*, vol. 70 (May 1980), pp. 250–57.