The Bank Merger Wave: Causes and Consequences

J. Alfred Broaddus, Jr.

I'm delighted to be here today to talk with you about the wave of bank mergers that is sweeping across our country—our hometown included. While Richmond had experienced some sizable ripples earlier, as we all know, the really big wave last year and earlier this year left very few local institutions in its wake. Many Richmonders are still adjusting to the loss of their banks and to the new, North Carolina-based, financial landscape. Residents of other U.S. cities—including large, proud cities like Philadelphia and San Francisco—are experiencing similar adjustments and emotions due to bank consolidations.

Turn back the calendar 28 years to see how times have changed. In 1970, the year I began working at the Richmond Fed, the largest bank in the country—the Bank of America with assets of $27 billion—was located in California; Charlotte, North Carolina, was a not-particularly well-known Southern city on the opposite coast. How many of us imagined then that Charlotte would later be headquarters to one of the world's largest banking companies, with assets of almost $600 billion? By virtue of being home to NationsBank and First Union, Charlotte has become a focal point of the rapid banking consolidation that is now extending across the whole of the United States.

Banking consolidation is big news these days, with a new megamerger announced almost monthly. The proposed Citicorp-Travelers union could break new ground on banking-insurance combinations, and the NationsBank-BankAmerica merger will produce a huge, truly national banking franchise. With change of this magnitude, however, come concerns, and people are concerned about a lot of things regarding these developments. They are concerned about higher fees and lower levels of service. They are concerned about credit

This article is adapted from an address by J. Alfred Broaddus, Jr., president of the Federal Reserve Bank of Richmond, to the Henrico Business Council in Richmond, Virginia, on September 17, 1998, and to the Mortgage Bankers of the Carolinas in Hilton Head, South Carolina, on September 18, 1998. Jeff Lacker, John Weinberg, John Walter, and Nita Jones contributed substantially to the article. The views expressed are those of the author and not necessarily those of the Federal Reserve System.

availability and disrupted banking relationships. In my remarks this morning I want to address some of these concerns and what I believe are some of the major forces behind these events.

1. HISTORICAL CONTEXT

To understand the developments I have just described, it is helpful to review briefly a bit of the history of American banking—particularly the history of restrictions on bank branching. Turn back the calendar once again. In the early years of the post–World War II period, strict and quite limiting branching restrictions were common throughout the United States. Obviously, consolidation in banking could not occur until these restrictions were removed. The restrictions had two sources. The first was a powerful sentiment that can be traced to the earliest years of our nation's history: a deep-seated distrust of large, centralized organizations—large financial institutions in particular. Subsequently, efforts to shield smaller banks by limiting competition from branches of larger banks became a factor as well.

As you will recall from your American History courses, the fear of concentrated financial structures became dramatically apparent during the early-nineteenth-century debates over whether to renew the charters of the First and Second Banks of the United States. Many were afraid that these large federal institutions would concentrate financial power and be used to benefit their owners at the expense of the broader public. As a result, neither charter was renewed, and after the demise of the Second Bank in 1837, no comparable replacement was chartered.

Despite these apprehensions, branching was not uncommon in the South before the Civil War, and several Midwest banks had branches as well. But there was little interest in establishing new branches following the war, since the technology that would allow inexpensive long-distance communication had yet to appear. Seeking approval from distant headquarters for local lending decisions would have been prohibitively costly before the widespread availability of the telephone. These communications costs argued for small, locally run banks. As a consequence—and this will be a surprising statistic for many of you—while there were approximately 13,000 banks in the United States at the turn of the century, there were only 119 bank branches in the entire country. In the last few years of the nineteenth century, advances in communications technology stimulated new interest in branching. But with increased interest came strong opposition, much of it from smaller banks and much of it successful, which ultimately produced widespread branching restrictions at both the federal and state levels.

One impediment to branching was the general belief that the National Banking Act passed during the Civil War prohibited it. To remedy the situation, legislative proposals were offered in the late 1890s that would have

allowed national banks to branch, but these proposals met with fatal opposition from several congressmen and the Comptroller of the Currency, who regulates national banks, on the grounds that they would concentrate banking power. As an alternative to *branches*, an act was passed in 1900 that lowered the minimum capital necessary to form a new national bank in a small town. In response, the number of *banks* almost doubled in the next ten years. The newcomers were primarily small unit banks: that is, single-office banks. These banks formed an antibranching fraternity that slowed the spread of branch banking for decades.

From 1900 until the early 1920s, the Comptroller prohibited national bank branching with few exceptions, and unit bankers lobbied successfully to contain the spread of branching by state-chartered banks. At the annual convention of the American Bankers Association in 1922, unit banks argued that "branch banking concentrates the credits of the Nation and the power of money in the hands of a few." During the 1920s, a number of states, including Virginia, imposed significant restrictions on the branching powers of state-chartered banks. Importantly, though, a handful of states bucked the trend and passed fairly liberal branching laws, among them notably—I could say prophetically—North Carolina.

For all of the strength of antibranching sentiment in this period, the high failure rate of unit banks throughout the 1920s and in the early years of the Great Depression revealed quite starkly a significant downside to branching restrictions: namely, the susceptibility of unit banks to runs generated by shocks to their local, usually relatively undiversified loan portfolios. Failure rates for the branching banks that existed were generally much lower, motivating some policymakers to call for liberalized branching as a means of diversifying individual bank portfolios and strengthening the banking system.

A number of states *did* liberalize their branching laws between 1929 and 1939. Further, in 1932 Senator Carter Glass, who as you know was one of the founding fathers of the Federal Reserve, proposed enhanced branching powers for national banks to allow both statewide and limited interstate branching. The momentum of this trend, however, was largely undercut early on by the passage of national deposit insurance in 1933, which guaranteed the stability of the banking system via an alternative route. Insurance allowed branching restrictions to continue with relatively marginal changes from the end of the Depression until the 1980s. During that 50-year period, the number of bank mergers was relatively modest: generally between 75 and 150 per year.

Since 1981, however, merger activity has exploded, with close to 600 mergers occurring in 1997 alone. Over this same period the opposition to branching that was so robust in the preceding 100-plus years eroded, and restrictions on branching collapsed in three steps. First, during the 1980s, 20 states, including Virginia, liberalized in-state branching laws. By 1990, 36 states authorized statewide branching, and only two prohibited it. Second, in the early 1980s, states began passing laws allowing bank holding companies from other states

to purchase banks within their borders. North Carolina, South Carolina, and Virginia did so in 1985 and 1986. By 1990, all but four states allowed at least some cross-border purchases. With this change, bank holding companies gained the ability to purchase banks outside of their headquarters' states, although they were required to operate these interstate acquisitions as separate banks—so interstate branching, for the most part, was still not possible. These two steps broadened in-state and interstate expansion opportunities markedly, and sizable banking companies began to form. As you will recall, it was during this period that NCNB in North Carolina began to grow rapidly, purchasing banks throughout the Southeast and Texas, and ultimately renaming itself NationsBank.

In short, the consolidation of the banking industry was well under way when the third step was taken: passage of the Riegle-Neal Interstate Banking Act of 1994, which finally eliminated interstate banking restrictions. This historic legislation gave banks the right to have branches nationwide and set the stage for the dramatic acceleration in merger activity here and elsewhere during the past two years.

2. EXPLAINING THE MERGER WAVE

Certainly the current merger wave would have been impossible without the elimination of branching restrictions, and at one level it is tempting to conclude that their removal explains the large number of consolidations. But state legislatures and the U.S. Congress were simply responding to pressures from banks wishing to pursue mergers. Consequently, rather than telling us what is driving the mergers, the easing of branching restrictions—while an essential precondition—simply begs the question.

One popular hypothesis is that individual banks merge in order to increase their market power, and it is true that national market shares have been steadily increasing in banking. The top ten banking firms increased their aggregate share from 22 percent in 1980 to 34 percent in 1997, while the five largest banks have almost doubled their share. But banking is still relatively fragmented nationally and is much less concentrated than many other major industries. Consider, for example, the soft drink and automobile industries. Both are far more concentrated than the banking industry, with the top two soft drink firms holding 74 percent of the market and the top three automakers controlling 71 percent. Yet most would agree that there is intense competition in these two industries.

More importantly, banking is still predominantly a *local* service, and measures of concentration at the local level have been virtually constant for the last two decades, even as the industry has consolidated nationally. The reason is that mergers have generally occurred *across* local markets rather than *within*

them—no accident, given that federal bank regulators scrutinize every bank merger for its effects on local concentration. Additionally, as long as new bank entry into particular local markets is largely unrestricted, competition should prevent abuses of market power and ensure consumer choice. In the last five years almost 670 new banks have been chartered throughout the United States, which has intensified competition in many markets. Closer to home, in North Carolina, South Carolina, and Virginia, 50 such banks have formed. In these circumstances it seems unlikely that the recent spate of bank mergers has been motivated in any material way by expectations of enhanced, exploitable market power.

So what *is* driving the merger wave? In brief it seems to me that the extraordinary advance in communications and data processing technology over the last two decades is the single most important underlying force—hardly an original insight but a powerful one. A prime example is the database-management software for mainframe computers that automated the recordkeeping that is the core of the banking business. The development of personal computers and the software that manages networks made it possible for banks to provide widespread access to their records at branches and automated teller machines (ATMs). Cost savings came as these advances were exploited to manage information databases far less expensively and more efficiently. A key point here is that these cost savings accrue most significantly in the management of very *large* databases: in sharing information among a large number of users and over wide distances. In other words, the benefits of the technology revolution accrue most fully to very large-scale banks. The ability to share customer and product information via computer networks has greatly lowered the cost of maintaining far-flung branches and of operating centralized call centers. All this has increased the *relative* advantage of being a big bank. More narrowly—but also on a technology note—some recent mergers may have been motivated in part by the desire of some banks to share the costs of Year 2000 compliance.

It seems clear then that technology is the fundamental force driving the merger wave. At first glance, this force and the consolidation that has resulted from it appears to have picked winners and losers rather arbitrarily. Charlotte becomes a major national banking center while Richmond loses most of its larger independent financial institutions. As I noted a minute ago, however, North Carolina permitted statewide branching well before most other states, and it seems clear in retrospect that this circumstance played at least some role in the emergence of the state as a banking center. Beyond the direct effect of consolidation on particular states and cities, however, keep in mind that the technological advances I have just described in conjunction with the demise of branching restrictions has greatly increased potential banking competition—and the benefits that result from it—in *all* local markets, including Richmond. It is now not only legally permissible but operationally feasible for any bank in the United States to establish a presence in Richmond, or, for that matter,

Charlottesville, Farmville, or Lexington. Local competition should increase even while the national banking industry consolidates.

3. RESPONDING TO ANXIETIES

While technological progress and heightened competition are typically thought to be good for consumers, the banking merger trend has been greeted with anxiety by many if not most Americans. Recent attention has focused on three such fears: diminished service, higher fees, and decreased credit availability—particularly for small businesses.

Diminished Service

When a bank is taken over, its customers often complain that the quality of service is not what they had come to expect from their old bank. And this may well be true for at least two reasons. First, the mix and pricing of products is likely to change with the merger, so customers preferring the old product mix will be less satisfied. The economies of scale that make large banks cost-effective depend on the standardization of products and service. Without standardization the information sharing that drives mergers would be inefficient at best. And cost savings would be lost if, with each merger, the acquirer added a new set of products or different versions of the same product. But this standardization can be a significant minus to customers who are accustomed to tailored services and want them to continue.

Second, as firms grow in size there occurs a natural numbing effect on service quality and initiative. A big-box retailer cannot offer the individualized service of the small retailer. Because the larger store's employees are responsible for a much broader line of products, they likely do not have the intimate knowledge of each product that is often found in smaller, more specialized shops. As banks merge into larger companies, there are similar results.

In today's more competitive market, however, many banks are eager to provide the antidote to standardized banking. New community banks are forming at an increasing rate here and elsewhere. Many of these banks enter a market precisely to capitalize on the shortage of "high-touch" banking created by recent consolidations and aggressively pursue the dissatisfied customers of merged institutions. These smaller banks can tailor products and service levels specifically to the demands of these customers.

Before the current merger wave, banks were relatively protected from competition and set service levels to appeal to the average customer. But today's open competition is forcing banks to differentiate themselves by service level. Large banks exploit the economies of large-scale production of standardized, "low-touch" banking. High-touch community banks focus on high-quality tailored services.

The additional choices in the new environment will almost certainly improve consumer well-being. Consumers will have more options from which to choose: high-touch community banks, on the one hand, and, on the other, large megabanks that offer less tailored services but a wide array of cost-effective products in a wide variety of locations. Although the number of banking organizations has declined by 42 percent since 1980, the number of banking *offices* has increased by 35 percent. This means that even after accounting for population growth, the number of banking offices per capita has increased by almost 15 percent. In the aggregate the banking industry has been expanding services even while consolidating.

Having said all this, it is certainly true that in the transition to the new banking structure some customers will be adversely affected by the disruption of established banking relationships. Suppose, for example, that you are a 70-year-old, high-balance customer or a small business, accustomed to a high-service banking relationship focused specifically on your needs. When a large bank with a very different approach to customer service buys the smaller bank you have dealt with for years, your initial reaction very likely will be dissatisfaction with the merger results. In the worst-case scenario, you may face the costs and inconvenience of switching your account to a bank that offers more personalized or company-specific services. Such costs are regrettable. The bright side is that they should prove to be temporary stumbling blocks in the transition to more robustly competitive banking markets.

High Fees

Attention has also been directed at the new or higher fees some customers must now pay for some banking services, which has led many to believe that the new merged banks charge unreasonably high fees. Clearly, banks have become more aggressive in their assessment of service charges and fees over the last decade, and big banks have moved to increase these charges sooner than smaller banks. Service charges on deposits as a percentage of deposits have risen by 42 percent for all banks and by 67 percent for the largest.

But I'm suspicious of the notion that banks in today's highly competitive banking environment can get away with charging fees significantly out-of-line with costs. My guess is that many of the fees have resulted from an unbundling of services: that is, charging explicitly for particular services rather than providing a bundle of services to all customers at one price. Customers who are more costly to serve are now charged higher fees, which allows lower-cost customers to be charged lower fees than would otherwise be possible. In the less competitive banking market of the past, banks covered most of their costs via their interest margin rather than by charging fees. They paid below-market rates of interest for deposits but invested them at market rates. They compensated depositors for the low deposit rates by offering them a largely undifferentiated

bundle of free services. Before the early 1980s, ceilings on deposit interest rates reinforced this arrangement. But equal service levels for all customers meant that high-balance customers were often subsidizing low-balance customers.

This comfortable world of cross-subsidies and minimal fees is no longer sustainable. Competition between banks intensified in the early 1980s as interest rate ceilings were removed and branching restrictions fell. Competition between banks and other financial institutions also intensified as nonbanks like Merrill Lynch offered market interest rates to attract depositors traditionally served by banks, especially higher-balance customers. Banks were forced to begin differentiating among customers, charging fees and varying interest rates according to customer balances and activity. Over time, this shift to matching interest payments and fees more closely to the costs of serving customers should result in a more efficient and equitable banking system. It will reduce cross-subsidies and encourage the industry to devote more resources to producing the most highly valued services. In many respects the greater incidence of fees so widely attributed to mergers is merely an acceleration of this already well-established trend.

Of all the new bank fees, none has received more attention than ATM fees, which some critics have called "unconscionable" and "outrageous." In fact, though, ATM fees, like other bank fees, appear to be an example of unbundling. Users are now required to pay for the convenience of this costly service. When a bank charges no specific fees for ATM use, customers who make little or no use of the machines subsidize other customers who are frequent users. Similarly, if customers of other banks pay smaller fees or no fees for ATM use, then customers of banks with extensive ATM networks subsidize noncustomers. Arguments like these are of little interest to ATM users who are accustomed to inexpensive or free access, and Senate Banking Committee Chairman D'Amato has introduced legislation that would ban certain fees. Most observers expect the fees to remain in place, however, which will encourage the installation of additional machines and promote the added customer convenience that accompanies them.

Unbundling, however, has also produced fallout beyond the dissatisfaction with ATM fees. When banking was less competitive, it had a public utility aspect—offering wide payments system access to all customers at the same price, while inevitably subsidizing some customers at the expense of others. As heavy competition eliminates cross-subsidies and rationalizes pricing, low-balance customers, in particular low-income individuals and households, are experiencing price increases. A backlash has developed and produced calls for federal legislation requiring the provision of low-fee accounts to small-balance depositors. No such action has been taken to date, but this issue is likely to receive further attention in the period ahead.

Credit Availability

Finally on the list of anxieties produced by the merger wave, some observers worry that the trend could adversely affect the availability of credit, particularly for small businesses. Smaller banks are a primary source of small-business credit. As large banks absorb small banks, who will make small-business loans?

Again, technology and competition are forcing banks to specialize in the way they serve customers, including small-business borrowers. Large banks, for the most part, are not abandoning small business. Rather, they are now offering small businesses a menu of standardized, quick-turnaround loan products. Because of the cost advantage in offering homogeneous products, large banks are likely to dominate such lending. These plain-vanilla loans have features that will suit many small businesses quite well. They offer speed: credit-scoring software accelerates creditworthiness decisions and loans can be approved within 24 hours. They offer convenience: loan applications can be made over the phone or, in some cases, over the Web, representing the ultimate in "low-touch" lending. They offer low interest rates: because these providers must compete with other large lenders offering similar products, rates are low. And finally they are amenable to comparison-shopping: standardized loan products vary little and are offered by many banks, so comparisons are easy to make.

Notwithstanding these attributes, standardized loans obviously are less suitable for small-business borrowers that require financial products tailored to their unique circumstances. Community banks retain an advantage over large banks in serving these customers, since smaller banks enjoy short lines of communication between lending officers and borrowing company owners and managers. This close communication permits community banks to customize products and employ borrower information in ways that large bank reporting and monitoring systems cannot easily accommodate. Three types of small-business borrowers can be expected to gravitate toward the community banks: (1) those lacking complete financial histories because of the newness of their operations or the uniqueness of their product; (2) those for whom the information needed to determine their creditworthiness is hard to summarize numerically for automated evaluation and requires face-to-face meetings to verify; and (3) those who want detailed and specialized financial advice. In sum, we can expect to see large banks specializing in standardized small-business lending and community banks in more tailored lending.

On balance, there is an excellent chance that, rather than reduced availability, small businesses will find a wider array of loan products to choose from going forward—in other words a more efficient loan market with no loss of availability. Here, as in some of the other areas I have discussed, the mergers currently taking place may create transitional costs as long-standing banking relationships are lost or altered in reorganizations. Ultimately, though, small

businesses should benefit from a broader selection of lending institutions to meet their specific credit needs.

4. CONCLUDING REMARKS

You may wonder where the Fed's main interest in all of this lies. Briefly, the Fed's goal and responsibility regarding bank mergers—and my personal goal and responsibility as a senior Fed official—is to ensure that these changes in the structure of banking institutions and markets are consistent with relevant banking law and, most fundamentally, that they serve the public interest rather than detract from it. So where do I come out on the issues I've raised?

In broad terms, I like what's going on, undoubtedly in part because I have a visceral aversion to efforts by governments to prevent, regulate, or slow market-driven change. In my view, the recent bank megamergers represent the structural adaptation of the banking industry, unfettered by archaic geographic restrictions on competition, to the opportunities afforded by new and emerging technologies. While some may suspect that the megamergers are motivated by a desire to monopolize markets and raise prices, there is no evidence that banking markets in fact are becoming more monopolized. On the contrary, the banking industry remains far less concentrated than many others we consider highly competitive. Moreover, competition has been enhanced by the recent entry of hundreds of new banks into particular local markets and the entry of a large number of existing banks into *new* local markets they had not served before. Although inevitably there will be costly disruptions of established banking relationships in the transition, this heightened competition offers the prospect of increased consumer and business choices among banking products and institutions, and decreased costs. These changes are squarely in the public interest. I might note here that I am well aware of the concerns some local community leaders have expressed regarding the potential impact of mergers on community reinvestment. The Board of Governors has given these concerns very careful attention in its consideration of particular merger applications, and it will continue to do so.

Having said all these generally favorable things about bank mergers, let me mention in closing one significant risk in this trend. This risk doesn't get much attention in the media when particular mergers are announced—indeed, it gets almost no attention—but it is quite important nonetheless. Unlike most other businesses, banks enjoy what is often called a federal financial safety net, specifically deposit insurance and access to the Fed's discount window and payment services. This safety net serves the public well most of the time.

Here's the risk: when a bank's balance sheet has been weakened by financial losses, the safety net creates adverse incentives that economists usually refer to as a "moral hazard." Since the bank is insured, its depositors will not necessarily rush to withdraw deposits even if knowledge of the bank's

problems begins to spread. In these circumstances the bank has an incentive to pursue relatively risky loans and investments in the hope that higher returns will strengthen its balance sheet and ease the difficulty. If the gamble fails, the insurance fund and ultimately taxpayers are left to absorb the losses. I am sure you remember that not very long ago, the savings and loan bailout bilked taxpayers for well over $100 billion.

The point I want to make in the context of bank mergers is that the failure of a large merged banking organization could be very costly to resolve. Additionally, the existence of such organizations could exacerbate the so-called "too-big-to-fail" problem and the risks it presents. Consequently, I believe the current merger wave has intensified the need for a fresh review of the safety net—specifically the breadth of deposit insurance coverage—with an eye toward reform. But that's another speech best left for another day.

Using the Term Structure of Interest Rates for Monetary Policy

Marvin Goodfriend

The term structure of interest rates, i.e., the yield curve, has long been of interest to monetary policymakers and their advisers. The transmission of monetary policy is conventionally viewed as running from short-term interest rates managed by central banks to longer-term rates that influence aggregate demand. A central bank's leverage over longer-term rates comes from the fact that the market determines these as the average expected level of short rates over the relevant horizon (abstracting from a term premium and default risk). Working in the other direction, the long bond rate contains a premium for expected inflation and, thus, serves as an indicator of the credibility of a central bank's commitment to low inflation.[1]

Different theoretical perspectives support the two above-mentioned uses of the term structure for monetary policy: John Hicks's (1939) expectations theory of the term structure supports the first, and Irving Fisher's (1896) decomposition of nominal bond rates into expected inflation and an expected real return supports the second.[2] The two views are compatible in principle, although reconciling them creates difficulties of interpretation in practice. For example, does a steepening yield curve indicate a loss of confidence in the central bank's commitment to low inflation, or does it indicate that markets expect tighter

The author is Senior Vice President and Director of Research. This article is an edited version of a paper written for the book *Money and Interest Rates*, I. Angeloni and R. Rovelli, eds., Macmillan 1998, proceedings of a conference sponsored by Banca d'Italia and IGIER, University Bocconi. Macmillan holds the copyright. The comments of Mike Dotsey, Bob Hetzel, Andy Olmem, John Walter, and participants at the Bank of England workshop on "Extracting Information from Financial Markets" are greatly appreciated. The views are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] See, for example, Goodfriend (1993), King (1995), and Svensson (1992).

[2] The idea that the term structure of interest rates can be explained by investors' expectations about future short-term interest rates dates back at least to Fisher (1896), but the main development of the theory was done by Hicks (1939).

policy in the form of a higher path of short rates pursued by the central bank? The yield curve contains information of use to monetary policymakers, but it needs to be interpreted in light of judicious subsidiary "identifying" conditions, together with other data on the economy. Some circumstances lend themselves to clearer interpretations than others, and there are many pitfalls.

Whether or not one regards longer-term interest rates as economic indicators or as part of the transmission mechanism for policy, or both, the term structure plays a potentially important role in the policymaking process. In spite of its complexity, the term structure cannot be ignored.

This article addresses some issues involved in using the term structure to conduct monetary policy. I begin by discussing the long bond rate as an indicator of inflation expectations. Second, I comment on the role that bond rates have played in recent U.S. monetary history. Third, I explain how information in the yield curve can be used to overcome what I call the "policy in the pipeline problem." Fourth, I review recent empirical evidence supporting the two theoretical views underlying our understanding of the term structure. I explain how "peso problems" associated with "inflation scares" in the bond market may help to account for a serious empirical failure of the expectations theory of the term structure. I also discuss evidence supporting the view that significant movements in long-term interest rates are largely driven by expected inflation. Finally, I point out some pitfalls of using the term structure to make tactical policy decisions.

1. PURSUING LOW INFLATION

The Chairman of the Federal Reserve Board, Alan Greenspan, supports a long-run goal for price stability such that "the expected rate of change of the general level of prices ceases to be a factor in individual and business decisionmaking."[3] The long bond rate is particularly well suited to help a central bank assess the degree to which it has achieved low inflation defined in that way. One could compare the behavior of the yield on a long-term nominal bond to its behavior in a past period in which inflation was very low and the public was reasonably confident that it would stay low. For instance, in the United States the 30-year nominal bond rate ranged from around 3 percent to a little over 4 percent from the mid-1950s until the mid-1960s, a period in which inflation averaged around 1 to 2 percent, and presumably, long-term inflation expectations were no more than that.[4] One would think that if the Federal Reserve (the Fed) were to achieve full credibility for low inflation, the long bond rate would once again move down to the 3 to 4 percent range. Most of the nominal bond yield would

[3] Greenspan (1990), p. 6.

[4] See Salomon Brothers and Hutzler (1969). Inflation as measured by the consumer price index actually jumped temporarily to the 3 to 4 percent range from 1955 through 1957.

then reflect an expected real return in the neighborhood of 3 percent.[5] Consistently low bond rates would constitute a key piece of evidence that inflation expectations had ceased to be a factor in individual and business decisions.

In addition, we would expect to see bond rates display the kind of indifference to incoming data that they showed in the low-inflation period of the 1950s and early '60s. Long bonds were then one of the most conservative investments, with stable bond prices and a dependable real return.[6] In sharp contrast, bond prices and ex post real returns became increasingly variable in the period of high and fluctuating inflation and inflation expectations. The variability of returns was also due in part to the increased range of short-term (real) rates that the Fed had to sustain from time to time in order to bring rising inflation under control.[7] For example, both factors were at work when interest rates peaked in 1981. With inflation then above 10 percent per year, long bond rates were double and bond prices were about half of what they had been in the mid-1970s.

The Fed succeeded in bringing inflation down to 4 percent by 1983 and has brought it down below 3 percent in the 1990s. Yet long bond rates continued to be sensitive to incoming data that raised the odds of higher future inflation and Fed action on short rates to head it off.[8] Bond rate volatility caused by the 1994 inflation scare suggests that the Fed did not then have full credibility for low inflation. Even if *actual* inflation remains low, the low inflation goal will not have been achieved until the United States has low and *stable* bond rates more characteristic of the last period in which the Fed had nearly full credibility for low inflation. The one percentage point decline in long bond rates in 1998 to below 5 percent indicates that the Fed has moved closer to full credibility for low inflation.

The U.S. government recently introduced 5-, 10-, and 30-year index bonds whose market yields reflect an expected real rate of interest over these horizons. The yield gap between an index bond and the comparable-maturity conventional (nominal) bond is a direct market estimate of expected inflation. Going forward, the size and stability of the yield gaps will help the Fed assess the extent to which it has achieved price stability.

---

[5] Theory and evidence both support the view that the expected real return on default-free long-term nominal bonds varies in a range within a percentage point or so of 3 percent. Quantitative work by Ireland (1996) that ties the ex ante real rate to expected consumption growth, which varies little over long horizons, suggests that the long real rate should range near 3 percent. And evidence from U.K. index-linked securities and U.S. index bonds also supports that view.

[6] See Ibbotson (1994).

[7] See Goodfriend (1993, 1997).

[8] See Borio and McCauley (1996) and Gerlach (1996) on the 1994 bond rate volatility. Compare the 1994 experience to the volatility described in Kessel (1965) and Salomon Brothers and Hutzler (1969).

Preemptive Policy

One of the most important lessons learned by central bankers in recent decades is that credibility for low inflation is the foundation of effective monetary policy.[9] The Fed has acquired credibility since the early 1980s by consistently taking policy actions to hold inflation in check. Experience shows that the guiding principle for monetary policy is to preempt rising inflation. The go-stop policy experience of the 1960s and '70s taught that waiting until the public acknowledges inflation to be a problem is to wait too long. At that point, the higher inflation becomes entrenched and must be counteracted by corrective policy actions more likely to depress economic activity.[10]

One might wonder why a preemptive strategy should apply more to fighting inflation than to unemployment? The answer is this. A central bank naturally has more credibility for fighting unemployment when the economy is weak than for fighting inflation when the economy is strong. The reason is that when the economy is weak, the public applauds an easing of policy because the obvious benefits in employment come immediately while any costs in higher inflation come later. On the other hand, tightening policy to preempt a rise in inflation invariably draws criticism because the risks of lower employment come immediately, while the benefit to stabilizing inflation is difficult to perceive.

To be preemptive, monetary policy must be forward-looking. That puts a premium on a forward-looking indicator, especially one that embodies a direct measure of inflation expectations such as a long bond rate. As I will point out below, the bond rate has not been a particularly good forecaster of changes in trend inflation, and so it certainly needs to be used in conjunction with other economic indicators. Yet there is evidence that the long-term nominal bond rate moves primarily as a result of inflation expectations. Sharp bond rate movements ought to be taken as evidence of worsening or improving credibility on inflation, as the case may be, and taken into account in making decisions on short-term policy.

2. THE ROLE OF BOND RATES: THE U.S. EXPERIENCE

In discussing the role of bond rates in recent U.S. monetary history, I present three examples of large bond rate movements that probably influenced policy actions by signaling sharply changing inflation expectations. I also comment on the fact that longer-term rates often seem to lead short-term rates over the business cycle.

[9] See Goodfriend and King (1997) for a formal justification of inflation targeting within what they call the New Neoclassical Synthesis macroeconomic model.

[10] See Goodfriend (1997).

Influential Bond Rate Movements

Significant bond rate movements probably influenced the timing and size of subsequent Fed policy actions on these three occasions.

The February 1980 Inflation Scare

After having tightened monetary policy sharply in the fall of 1979, the Fed, based on evidence of a weakening economy, held short rates relatively steady in January and February of 1980. Meanwhile, the 30-year bond rate jumped by 2 percentage points between December and February. The inflation scare was primarily the result of three factors: (1) inflation as measured by the implicit price deflator was nearly 2 percentage points higher in the first quarter of 1980 than in the previous quarter, partly due to the second oil shock; (2) the Soviet Union's invasion of Afghanistan destabilized financial markets; and (3) the Fed hesitated with its policy tightening in the face of a weakening economy. The Fed's hesitation probably created some doubt about whether the Fed would hold the line on inflation. At any rate, the Fed's decision to resume its policy tightening with a huge 3 percentage point increase in the federal funds rate in March was probably influenced significantly by the sharp prior increase in the long rate.

The 1984 Inflation Scare

The economic recovery from the 1981–82 recession was robust. Real GDP grew by 5 percent in 1983–84. Although inflation was only around 4 percent, the long bond rate rose from about 10 percent in the summer of 1983 to peak the following summer at around 14 percent. Amazingly, this was only about 1 percentage point short of its peak in 1981 even though by mid-1984 inflation was 5 or 6 percentage points lower. The Fed raised short-term interest rates in line with the long rate, and the yield curve remained flat. Although there were clearly other good reasons to tighten monetary policy at the time, the sharp rise in the long rate probably contributed to the Fed's inclination to raise short rates as much and for as long as it did.

The 1985 Acquisition of Credibility

The Fed held short rates in the 7 to 8 percent range in 1985 and early 1986, while real GDP grew at 3.3 percent and prices increased at about 3.5 percent. In early 1986, oil prices moved down from around $28 a barrel to less than $15 a barrel. Against these developments, the 30-year bond rate declined from around 12 percent to 7 percent between January 1985 and April 1986, half of the decline coming *before* the collapse of oil prices. The huge 5 percentage point drop in the long rate signaled a big jump in the credibility of the Fed's commitment to low inflation and probably contributed to the Fed's inclination to move short rates down about 2 percentage points in 1986.

Does the Fed Follow the Bond Market?

Economists and financial market analysts have noted that longer-term rates have a tendency to lead short rate movements over the business cycle. In other words, the Fed often appears to follow the market. Some observers argue that the Fed is obliged to follow longer rates and exerts little independent influence of its own. Others recognize that the Fed has considerable discretion over short rates, but they interpret the evidence as indicating that the Fed follows long rates because these are taken to indicate the direction the short rates *ought* to follow for stabilization purposes. This second view is often accompanied by a plea that the Fed should not blindly follow the bond market.

In fact, the Fed has considerable discretion to influence the evolution of short rates. It moves short rates to stabilize inflation and unemployment with the help of a variety of economic indicators, including bond rates. The Fed does not automatically follow longer-term rates though. It only appears to do so at times. The fact that long rates are determined in good part (according to the expectations theory of the term structure) as the average of expected future short rates causes the bond market to try to predict future Fed interest rate policy actions. To the extent that Fed policy contains "systematic follow-through," bond rates move ahead of future changes in short rates. On the other hand, on those occasions when long bond rates jump sharply due to an inflation scare, or fall sharply due to the acquisition of credibility for lower inflation, the Fed might follow with higher or lower short rates, respectively. But the Fed would only take such policy actions if it interpreted the information in long rate movements as consistent with other information signaling a sharp and persistent change in inflation expectations.

Bond Market Vigilantes

The forward-looking nature of bond rates has led some commentators to argue that "bond market vigilantes" are capable on their own of stabilizing the economy against inflation. The argument implies that central banks are now largely irrelevant. This point makes no sense and is actually quite dangerous. Long rates often rise ahead of central bank actions because they reflect a higher expected future path of short rates. If a central bank were to disappoint the bond market by not following through, then bond rates would likely not rise as much in the next potentially inflationary episode. In effect, bond markets are vigilantes only when they are "trained" to be so by credible anti-inflationary monetary policy.

Bond markets would cease to be vigilantes if the central bank ceased to follow a credible low-inflation policy. In such an environment an increase in long rates could reflect higher inflation expectations, i.e., an inflation scare. Rather than acting to restrain spending and inflation, an inflation scare would signal a loss of confidence in the central bank's commitment to low inflation. A central bank might then have to react with a higher path for short-term real

rates to hold the line on inflation.[11] Any way you look at it, bond markets are not capable on their own of automatically maintaining low inflation.

3. POLICY IN THE PIPELINE

It is difficult for a central bank to know when and how much to change short-term interest rates to hold the line on inflation or to resist a recession. In practice, a central bank moves short rates in steps so it can observe the consequences of its actions and assess sequentially the need for each incremental rate change. Policymakers know that it takes some months for the effects of a given change in rates to be felt by the economy. Policy can cumulate "in the pipeline," so to speak, as a sequence of policy actions lengthens. As a tightening proceeds, for example, central bankers become more cautious about taking further actions for fear of overdoing it, and creating a recession. Of course, the opposing risk is that excessive caution might allow inflation to rise.

The term structure of interest rates can play a useful role in assessing how much policy is in the pipeline. If a central bank has credibility as an inflation fighter, then markets may guess correctly that an initial increase in the short rate is likely to be followed by further increases. The expected future path of short rates will be built immediately into the term structure of interest rates. As Dahlquist and Svensson (1996) show, it is possible to extract the expected sequence of future short rates from the spot rate yield curve; the constructed sequence of future rates can then be displayed as a corresponding forward rate curve. Under the assumption of negligible term premia, the forward curve shows the time path of the market's expectation of future short-term interest rates.

Using the forward rate curve, a central bank can see that its initial rate increase carries with it expectations of a whole sequence of increases. Thus, not only the first rate increase but a whole sequence of projected increases in short rates is put into the pipeline the moment a series of tightenings is initiated. Indeed, to the extent that markets begin to expect a sequence of tightening actions before they begin, policy is put into the pipeline before a central bank actually raises short-term rates.

To the extent that a central bank's subsequent interest rate increases were predicted, they would not constitute new policy impulses. A central bank could confidently follow through without being deterred by policy in the pipeline. On the other hand, the central bank could use the forward rate curve to gauge the extent to which the actual path for short rates departed from the initially predicted one. It could thereby keep track of new policy impulses it was putting into the pipeline.

[11] See Goodfriend (1993) and Mehra (1997).

The above discussion can be made more concrete by reference to the 1994 policy tightening. Campbell (1995) constructs and reports a set of forward rate curves extracted from the corresponding U.S. spot yield curves at different dates in 1994. The Fed raised short-term interest rates in a series of seven steps from 3 percent beginning in early February 1994 to 6 percent in early February 1995. The first point of note is that one-year-ahead forward rates rose from 3 to 4 percent in early January 1994, indicating that the bond market expected a significant tightening well before it began. Second, just after the Fed first increased the short rate by 25 basis points in February 1994, the market expected the Fed to raise short rates to 5 percent by early 1995. By early May 1994 the market was looking for 6 percent short rates in May 1995. In mid-May 1994, after having moved the spot short rate up to 4.25 percent, the Fed announced its belief that further policy moves would be unnecessary in the short term, and the forward rate curve fell, indicating that the market then expected a May 1995 short rate of around 5.25 percent.

Judging by the behavior of the forward rate curves, it seems fair to say that the bulk of the Fed's policy impulses were delivered in three major steps—the first percentage point increase by early January, an additional percentage point in early February, and a third by early May. The announcement in mid-May constituted an impulse for easier policy, and so on.

The bottom line is this. Most of the seven federal funds rate policy actions did not put much new policy into the pipeline at the time that they were actually taken. The actions merely supported longer-term interest rate increases that had already happened. Generally speaking, an uncoupling of policy actions and impulses should be expected to characterize episodes of policy tightening or easing. Using the term structure to distinguish between actions and impulses is the first step in dealing with the policy in the pipeline problem. Of course, it will take considerable effort to work out a comprehensive method for dealing with the problem in practice.

4. EVIDENCE ON THE DETERMINATION OF BOND RATES

The two theoretical foundations of our understanding of long bond rates, the expectations theory of the term structure and the Fisher decomposition, have been extensively assessed on empirical grounds. Some recent work shows how the monetary policy perspective complements the finance-theoretic understanding and interpretion of the behavior of the yield curve.

The Expectations Theory of the Term Structure

Empirical work that tests the expectations theory of the term structure with U.S. data finds some unsettling results. Campbell and Shiller (1991, p. 505)

summarize the main findings this way: "The change in the long-term rate does not behave as predicted—the slope of the term structure almost always gives a forecast in the wrong direction for the short-term change in the yield [to maturity] on the long-term bond, but gives a forecast in the right direction for long-term changes in short rates." In other words, long rates seem to overreact to short rates.

We can understand the force of Campbell and Shiller's comment by reviewing the logic behind two key implications of the expectations theory of the term structure. The first implication begins with the idea that, in equilibrium, the interest rate on a long-term bond must equal the average expected level of short rates over the relevant horizon (abstracting from a term premium and default risk). If the long rate were above the expected average of future short rates, then investors would prefer to hold a long-term bond rather than a sequence of short-term securities. But that calculation on the part of investors would cause the bond price to rise until the long-term interest rate fell enough to equate the expected returns on the two investment strategies. The upshot is that when the short rate is below the long rate, the expectations theory of the term structure says that future short rates must be expected to rise, and vice versa. Assuming that market expectations are formed rationally, the theory predicts that short rates will actually rise on average in this case or fall if the short rate is above the long rate. This is the first important implication of the expectations theory of the term structure.

The second implication follows by comparing the expected one-period return to holding a short-term security with the return of a long-term bond held for one period. By holding the short-term security, an investor would earn the short-term interest rate. There is no risk of capital gain or loss on the short security because it matures after one period. Now consider a long bond that makes a constant coupon payment each period and matures a few periods in the future. The one-period return on the long bond has two components. The first component is the coupon divided by the bond price, i.e., the interest yield. The second component will equal the one-period expected appreciation (or depreciation, if any) of the bond price divided by the bond purchase price.

Once again, theory tells us that these two one-period returns must be equal in equilibrium. If the current bond price is such that the one-period interest return on the long bond is above the short rate, then the market must be expecting the bond price to fall. Since the coupon payments are fixed, the lower future bond price, in turn, must imply that the yield to maturity on the long bond is expected to move still higher (because both the interest and the price appreciation components of the bond move higher). The upshot here is that when the short rate is below the long rate, the expectations theory of the term structure says that long bond returns must be expected to rise. Assuming that expectations are formed rationally, the theory predicts that the long rate will actually rise on average when it is higher than the short rate or fall if the short

rate is above the long rate. This is the second important implication of the term structure of interest rates. It is this implication that Campbell and Shiller point out is not observed in the data. Instead of being followed by a change in the long-term interest rate in the same direction as the sign of the slope of the yield curve (long rate minus short rate), the long rate tends to move in the opposite direction.

Bekaert, Hodrick, and Marshall (1997) offer an explanation for this empirical failure that is driven by small-sample anomalies caused by peso problems in the data analysis. They explain how the interest rate data could have been generated by investors who behave according to the expectations theory of the term structure and form their expectations efficiently. Bekaert et al. (1997, p. 13) explain the peso problem this way: "Suppose that short-term interest rates can evolve in three different regimes, with the mean and volatility of rates increasing together as we move across regimes. Further, suppose that any shock that increases the short-term rate also increases the probability of switching to a higher-rate regime. Then, as short rates rise, the term spread may rise as agents rationally forecast a transition to a higher-rate regime. However, if in a particular sample, the higher-rate regimes are observed less frequently than their unconditional probabilities, this increase in the spread will appear unjustified ex post. Thus, increases in the spread are subsequently followed by surprising persistence of lower-rate regimes. At the same time, short rates immediately following the shock will tend to be higher than their unconditional value even if rates stay within a low-rate regime, since they are highly serially correlated. This could account for the puzzling ability of the term structure to predict the direction of short rates but not long-bond returns mentioned above."

High and volatile interest rates were more common in recent decades in the United Kingdom than in either Germany or the United States. According to the peso-problem view, one would expect there to be less evidence against the expectations hypothesis of the term structure in countries with a sample that is more representative of the population distribution. Bekaert et al. emphasize that the evidence supports the peso-problem view since there is only weak evidence against the expectations hypothesis in U.K. data.

Bekaert et al.'s peso-problem interpretation of U.S. interest rate data fits nicely with the idea, emphasized in Goodfriend (1993), that the inflation-scare concept helps understand the behavior of bond rates in the United States. To appreciate the connection, consider this: As Bekaert et al. (1997, p. 2) put it, underlying the peso-problem interpretation of U.S. data is the idea that the true "data generating process includes a low probability, usually catastrophic, state that generates extreme disutility to economic agents. Because the state has low probability, it is unlikely to be observed in a given small sample of data. Because it is catastrophic, the possibility that this state may occur substantially affects agents' decisions, which in turn determines equilibrium prices and rates of return. . . . When a peso problem is present, the sample moments calculated

from the available data do not coincide with the population moments that agents actually use when making their decisions."

Although Bekaert et al. do not mention it, from the inflation-scare point of view the catastrophic state can be interpreted as one with a high trend rate of inflation, perhaps much higher than the 13 percent inflation rate the United States experienced temporarily in the early 1980s. According to the inflation-scare interpretation, long bond rates in the United States jumped sharply on many occasions, reflecting an increased likelihood of a transition to a higher trend inflation state that never materialized because the Fed happened to take countervailing action to resist it in this small data sample. The bond rate came down after the inflation scares, but future short rates remained higher for a while because a higher path for short-term real interest rates was needed to restore credibility for low inflation.

The Fisher Decomposition

Irving Fisher (1896) pointed out that a nominal interest rate on a security is composed of an expected real return and a premium to compensate investors for inflation expected over the life of the security.[12] The introduction of index-linked bonds in the United Kingdom in the early 1980s has by now created a reasonably long time series of direct evidence on the Fisher decomposition of nominal bond rates. Barr and Campbell (1997) report the results of an empirical study of the expected real interest rate and the expected inflation components of the bond rate, assuming that the log version of the pure expectations hypothesis holds. Their major findings are these. Somewhat surprisingly, short-maturity nominal bonds are less risky than short-maturity real bonds, but long-maturity nominal bonds are riskier than long-maturity real bonds. They recognize that this pattern is explained by the large negative short-run correlation between real interest and expected inflation. At longer horizons this correlation is very weak and has little effect on the variability of nominal bond returns.

At longer horizons the real interest rate becomes less variable, leaving expected inflation as the dominant factor driving bond returns. Almost 80 percent of the movement of long-term nominal rates in the United Kingdom appears to be due to changes in expected long-term inflation. The series on expected inflation computed using the indexed and nominal bonds forecasts actual inflation better than the nominal bond rates.[13] The regressions for short horizons

[12] See Ireland (1996) for a modern exposition of Fisher's idea.

[13] Breedon (1995) also reports that medium-term expectations of inflation derived from the U.K. index bonds in conjunction with the nominal bonds predict changes in future inflation reasonably well, though they exhibit a consistent tendency to overpredict future inflation itself.

Using the above technique, the Bank of England's Inflation Report for May 1997 reports that expected inflation at ten years fell by about 50 basis points on the announcement of the Bank of England's independence.

confirm Mishkin's (1990b) finding for the United States that the term structure of six months or less contains no information about the path of future inflation.

The above-mentioned findings indicate that long bond rate movements are largely driven by expected inflation. The findings support the idea that sharp long rate movements can be interpreted as indicative of shifts in the credibility of the central bank's commitment to low inflation. At the same time, although the expected real interest rate becomes less variable at longer horizons, there still appears to be room for a central bank to exercise a degree of influence on longer-term real rates through its management of short rates. Finally, the large short-run negative correlation between expected inflation and the expected real rate is consistent with the fact that central banks manage short-term nominal rates closely and smooth them against shocks. With short nominal rates kept constant by a central bank, a shock to the inflation expectations component of the rate implies an equal and opposite movement in the expected real rate.

5. PITFALLS IN USING THE TERM STRUCTURE

There are serious pitfalls in using bond rates to gauge the inflation risk in the outlook for the economy, or in gauging the degree to which a series of short-term interest rate policy actions will be transmitted to the economy through longer-term interest rates. I discuss these briefly below.

Bond Rate Forecasting Failures

The long bond is arguably a good indicator of the market's perception of a central bank's commitment to low inflation. That alone makes significant bond rate movements deserving of the attention of central bankers. A related but separate question is the extent to which bond rates actually have proven to be good forecasters of future inflation trends. As discussed above, an ongoing inflation trend is reflected in higher bond rates. And the term structure does contain information for forecasting cyclical swings in inflation.[14] But when it comes to foreseeing changes in the trend rate of inflation, bond rates have not done as well. For instance, U.S. bond rates did not foresee the big jump in trend inflation that occurred in the late 1960s and early 1970s. Rates did move up, but only in line with the actual deterioration in current inflation.

As another example, consider that the U.S. 30-year rate was roughly in the same 8 percent range in early 1992 as it was in early 1977, in spite of the fact that inflation was 3 percentage points lower in 1992 than in early 1977. Assuming a real long-term interest rate of around 3 percent, the long-term expected rate of inflation would have been about 5 percent in both years. Apparently, investors perceived the 6 percent inflation rate as temporarily high in early

[14] See, for instance, Mishkin (1990a) and references contained therein.

1977, and they perceived the 3 percent inflation rate in 1992 as temporarily low. However, the five years beginning in 1977 saw the worst inflation of the period, and inflation has fallen by a percentage point or more since 1992.

Even more spectacular, the fact that the U.S. long rate rose to around 14 percent in the summer of 1984 seems incredible in light of the fact that trend inflation since then has remained around 4 percent or less. Clearly, bond rates have not always been very good predictors of changes in inflation trends.

Policy Actions and Long Rates

The Fed moved short-term rates up by about 3 percentage points from the spring of 1988 to the spring of 1989, but the 30-year bond rate increased relatively little, and the yield curve was inverted for most of 1989. In contrast, the Fed again moved short rates up by 3 percentage points from February 1994 to February 1995. Yet in this latter case the long rate moved up from a trough of less than 6 percent in October 1993 to peak at over 8 percent in November of 1994, and the yield curve did not invert.

The two episodes of policy tightening were similar in magnitude and not far removed in time. Moreover, inflation rose only modestly in the late 1980s and actually held steady at around 3 percent in 1994–95. Yet the behavior of the long rate differed enormously in the two periods. What should one conclude? Apparently, the effect of a policy tightening on long rates can differ widely depending on the circumstances. This suggests that the transmission of a sequence of interest rate policy actions to the economy depends on underlying factors such as the state of the business cycle or the nation's commitment to low inflation.

An alternative interpretation might be this: In fact, the long rate actually jumped by 2 percentage points from January to September 1987 just before the stock market correction. The bond rate registered an inflation scare in 1987, but perhaps the Fed's response was delayed by the transition from Chairman Volcker to Chairman Greenspan, which took place in the summer, and later by the October stock market correction. Under this interpretation a 2 percentage point bond rate move accompanied a 3 percentage point short rate increase in both the 1988 and 1994 periods. One might conclude that the only difference is that the policy tightening associated with the bond rate rise was delayed by a year in the earlier period.

Even if these two episodes can be seen as reflecting similar correlations between the bond rate and the short rate, is there any reason to expect the correlation to be stable in the future? Clearly the answer is no. Long rates varied relatively little with short rates in the low-inflation 1950s and '60s.[15]

[15] See the nice demonstration of this point in Chadha and Ganley (1995), who show that the correlation between short and long U.K. interest rates in the low-inflation 1870–1914 period is much smaller than in the high-inflation 1970–1995 period.

Inflation expectations were then securely anchored, and the range in which the Fed varied short rates to stabilize the economy was smaller in the low-inflation period than it was in the 1970s, ’80s, and ’90s. If the Fed succeeds in acquiring full credibility for low inflation in the years to come, then short and long rates should once again co-vary as in the earlier period. In retrospect, the late 1980s and mid-1990s may be seen as a transition period in which short and long rates continued to exhibit the kind of covariation observed in the period of high inflation.

Direct Policy Leverage on the Long Real Rate

Monetary policy transmission is conventionally viewed as running from short-term real interest rates managed by central banks to longer-term real rates that influence aggregate demand. There are two major pitfalls to overcome in estimating such direct policy leverage on the long real rate. First, as discussed in the policy in the pipeline section above, one must distinguish policy actions from policy impulses. Interest rate policy actions that have been anticipated clearly would not be expected to affect longer-term rates much, if at all. One should construct and use a sequence of policy impulses in order to gauge the effect of policy on longer-term rates. Second, when current inflation is stable and inflation expectations are well-anchored, then it is reasonable to interpret the effect of a nominal short rate policy impulse on the nominal long-term rate in real terms. While those conditions were probably satisfied in the 1950s and early ’60s United States, they probably have not been satisfied completely since then. Actual inflation has been well-behaved in the 1990s, but the relatively large movements in long bond rates suggest that inflation expectations are still not firmly anchored.

With those caveats in mind, consider some simple evidence on the leverage that short rate policy actions exert on long rates. Cook and Hahn (1989) found for the United States in the late 1970s that a 100-basis-point increase in the Fed’s nominal federal funds rate target increased the nominal 30-year rate by 13 basis points on average. Cook and Hahn used a narrow day or two time window in their calculations. Two rough calculations in my 1993 paper suggest a larger 25-basis-point effect on the 30-year rate per 100-basis-point short rate policy action in 1979 and 1980.[16] Assuming that both actual inflation and inflation

[16] The sharp 2.3 percentage point federal funds rate rise from September to October 1979 pulled the long bond rate up 0.7 percentage points. And the sharp 8.6 percentage point funds rate reduction between April and June 1980 pulled the long rate down 1.6 percentage points. Averaging the two effects yields about 25 percent as the fraction of aggressive funds rate policy actions transmitted to the long rate. Among all the sequences of aggressive policy actions in the period I studied, these two seemed the best candidates to gauge the size of direct leverage of policy over the long rate in real terms. They were very large moves, surprising in their timing and magnitude. Further, they were taken in only a few weeks’ time when inflation and, plausibly, inflation expectations were relatively unchanging. See Goodfriend (1993), p. 13.

expectations were relatively unchanging on average when these policy actions were taken, we can interpret these estimates of policy leverage in real terms.[17]

Taken as a whole, the year-long episode of policy tightening in 1994 suggests the potential for much greater direct policy leverage over the long real rate. As mentioned above, the nominal long rate moved about two-thirds as much as the nominal short rate in 1994. Since inflation held steady, the 3 percent increase in nominal short rates was entirely real. The one to two-thirds leverage, however, should be considered an upper bound on the direct term structure effect running from real short to real long rates because the long rate rise almost certainly included an increase in inflation expectations as the inflation scare ran its course.

We can say more. As it happened, the 1994 long bond movements in the United Kingdom paralleled those in the United States: both rose by about 2 percentage points.[18] Using the U.K. index bond, Barr and Campbell (1997) show that the 1994 rise in the U.K. ten-year nominal bond rate was due in equal parts to a rise in expected inflation and a rise in the expected real yield. Applying a similar decomposition to the rise in the U.S. long rate cuts the apparent direct leverage of short-term real rates over the long real rate down to 30 basis points per 100-basis-point short rate action, more in line with the evidence described above.

It bears repeating that the leverage exerted by short rates over long rates is regime dependent. Policy leverage will depend on the market's expectation of what a given central bank policy action implies for the expected path of future short rates. For the 1970s period in the United States examined by Cook and Hahn, the Fed was not yet in a full-fledged inflation-fighting regime. That might explain why the leverage found by Cook and Hahn is smaller than for the early 1980s or for 1994. Moreover, one might think that policy leverage in the 1950s was relatively weak too, since policy actions needed to stabilize the economy were relatively small and of short duration.[19] The point is that

[17] A change in the central bank's short rate target can itself convey information that simultaneously influences long-run inflation expectations one way or the other. Markets could become more concerned about future inflation because the central bank has revealed *its* concern; or markets could feel more confident of price stability because the central bank *is* taking action against inflation. The statement in the text assumes that these two effects cancel each other. Clearly, this question needs to be addressed in a more sophisticated way, controlling explicitly for changes in expected inflation, perhaps using index-bond market data.

[18] See Borio and McCauley (1996).

[19] An important point to keep in mind is that a given short-rate policy action will exert a greater effect on the long rate the shorter the average life of the bond as measured by its duration. The duration of a coupon bond may be thought of as the term to maturity of an equivalent zero coupon bond that makes the same total payments and has the same yield. The duration of a very long-term bond selling near par is approximately $1/r$, where r is the per-annum yield to maturity. The duration of the 30-year bond, for instance, is only about 12 years at an interest rate of 8 percent, but it rises to 33 years at a 3 percent interest rate. Other things the same, policy leverage

relatively aggressive short rate actions are required to *restore* credibility for low inflation after it has been compromised, whereas policy actions taken to *maintain* credibility for low inflation can be quite modest.

6. SUMMARY

The term structure of interest rates can play an important role in the making of monetary policy. Long rates indicate the extent to which a central bank has achieved price stability. Significant bond rate movements influence the timing and magnitude of monetary policy actions. On the other hand, the ability of bond rates to forecast changes in inflation trends is not particularly good. Moreover, the influence of policy actions on longer-term rates can be quite variable. In particular, the degree of restraint transmitted by policy is difficult to manage in a transition between high- and low-inflation regimes. The effect of policy on the economy becomes more predictable once low inflation is secure.

The peso-problem interpretation of some anomalies in the empirical assessment of the expectations hypothesis literature squares nicely with the inflation-scare interpretation of sharp movements in bond rates. Recent empirical findings on the Fisher decomposition of nominal bond rates also accord well with the influence of inflation scares and central bank interest rate smoothing on interest rates.

Some points about the use of the term structure for making tactical policy decisions are worth reiterating: (1) the need for policy to preempt a rise in inflation and inflation expectations puts a premium on the long bond rate as an indicator of credibility for low inflation; (2) policy leverage on long rates is regime dependent and, in particular, will vary with a central bank's commitment to price stability and its credibility for low inflation; (3) policy often follows long rates because long rates embody expectations of future short rate policy actions and because long rate movements often signal changing inflation expectations that may precipitate a policy reaction; (4) bond market vigilantes do not make central banks irrelevant; and (5) the yield curve can be employed usefully to distinguish policy actions from policy impulses in order to tell how much policy is in the pipeline.

The alert reader may have noticed that I have not discussed how the term structure might help a central bank forecast the risk of recession. There is a literature showing that term spreads are useful for predicting recessions as much as two years ahead. Bernard and Gerlach (1996) document the evidence for eight countries over two decades. While this finding seems robust, it is of less use to central banks than one might think. The reason, as Bernard

---

over longer-term rates will be much smaller at low interest rates such as those observed in the 1950s and early '60s.

and Gerlach recognize, is that over this sample period most recessions follow periods in which central banks have tightened monetary policy to fight inflation. A term spread that is inverted by a deliberate tightening of monetary policy may contain little additional information of use to central bankers.

REFERENCES

Barr, David G., and John Y. Campbell. "Inflation, Real Interest Rates, and the Bond Market: A Study of U.K. Nominal and Index-Linked Government Bond Prices," *Journal of Monetary Economics*, vol. 39 (August 1997), pp. 360–84.

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall. " 'Peso Problem' Explanations for Term Structure Anomalies," Stanford University Graduate School of Business Research Paper 1145, July 1997.

Bernard, Henri, and Stefan Gerlach. "Does the Term Structure Predict Recessions? The International Evidence," Bank for International Settlements Working Paper 37, September 1996.

Borio, Claudio E. V., and Robert N. McCauley. "The Economics of Recent Bond Yield Volatility," Bank for International Settlements Economic Paper 45, July 1996.

Breedon, Francis. "Bond Prices and Market Expectations of Inflation," *Bank of England Quarterly Bulletin*, vol. 35 (May 1995), pp. 160–64.

Campbell, John Y. "Some Lessons from the Yield Curve," *Journal of Economic Perspectives*, vol. 9 (Summer 1995), pp. 129–52.

_____, and Robert J. Shiller. "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, vol. 58 (May 1991), pp. 495–514.

Chadha, Jagjit S., and Joe A. Ganley. "Monetary Policy, Inflation Persistence and the Term Structure of Interest Rates: Estimates for the U.K., Germany, and the U.S." Mimeo, Bank of England, 1995.

Cook, Timothy, and Thomas Hahn. "The Effect of Changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s," *Journal of Monetary Economics*, vol. 24 (November 1989), pp. 331–51.

Dahlquist, Magnus, and Lars E. O. Svensson. "Estimating the Term Structure of Interest Rates for Monetary Policy Analysis," *Scandinavian Journal of Economics*, vol. 98 (June 1996), pp. 163–83.

Fisher, Irving. "Appreciation and Interest," *American Economic Review Publications*, vol. 11 (August 1896), pp. 331–442.

Gerlach, Stefan. "Monetary Policy and the Behavior of Interest Rates: Are Long Rates Excessively Volatile?" Bank for International Settlements Working Paper 34, January 1996.

Goodfriend, Marvin. "Monetary Policy Comes of Age: A 20th Century Odyssey," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 83 (Winter 1997), pp. 1–22.

_____. "Interest Rate Policy and the Inflation Scare Problem: 1979–1992," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 79 (Winter 1993), pp. 1–24.

_____, and Robert G. King. "The New Neoclassical Synthesis and the Role of Monetary Policy," in Olivier J. Blanchard and Stanley Fischer, eds., *NBER Macroeconomics Annual 1997*. Cambridge, Mass.: MIT Press, 1997.

Greenspan, Alan. Statement before the U.S. Congress, House of Representatives, Subcommittee on Banking, Finance and Urban Affairs. *Zero Inflation*. Hearing, 101 Cong. 1 Sess. Washington: Government Printing Office, 1990.

Hicks, John. *Value and Capital*, 2d ed. London: Oxford University Press, 1939.

Ibbotson Associates. *Stocks, Bonds, Bills, and Inflation 1994 Yearbook*. Chicago: Ibbotson Associates, 1994.

Ireland, Peter N. "Long-Term Interest Rates and Inflation: A Fisherian Approach," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 82 (Winter 1996), pp. 21–35.

Kessel, Reuben A. "The Cyclical Behavior of the Term Structure of Interest Rates," National Bureau of Economic Research Occasional Paper 91, 1965.

King, Mervyn. "Credibility and Monetary Policy: Theory and Evidence," *Bank of England Quarterly Bulletin*, vol. 35 (February 1995), pp. 88–91.

Mehra, Yash P. "A Federal Funds Rate Equation," *Economic Inquiry*, vol. 35 (July 1997), pp. 621–30.

Mishkin, Frederic S. "The Information in the Longer Maturity Term Structure about Future Inflation," *Quarterly Journal of Economics*, vol. 105 (August 1990a), pp. 815–28.

_____. "What Does the Term Structure Tell us about Future Inflation?" *Journal of Monetary Economics*, vol. 25 (January 1990b), pp. 77–95.

Salomon Brothers and Hutzler. *An Analytical Record of Yields and Yield Spreads*. New York: Salomon Brothers and Hutzler, 1969.

Svensson, Lars E. O. "Targets and Indicators with a Flexible Exchange Rate," in *Monetary Policy with a Flexible Exchange Rate*. Sveriges Riksbank, 1992.

The Predictive Content of the Interest Rate Term Spread for Future Economic Growth

Michael Dotsey

Predicting economic activity is important for numerous reasons. It is important for business firms because it aids in deciding how much capacity will be needed to meet future demand. It is important for various government agencies when forecasting budgetary surpluses or deficits. And it is important for the Federal Reserve (the Fed) in deciding the stance of current monetary policy. One set of variables that are potentially useful in forecasting economic activity are financial variables.

Financial market participants are forward-looking, and as a result the prices of various securities embody expectations of future economic activity. This pricing behavior implies that data from financial markets may reasonably be expected to help forecast the growth rate of the economy. Using financial variables to aid in economic projections, therefore, is fairly commonplace. In particular, the yield curve spread between long- and short-term interest rates has received a lot of recent attention. Although not the first to consider the implications that the spread has for predicting economic activity, Stock and Watson (1989) provided much of the impetus for further research by finding that the spread was an important component of their newly constructed index of leading economic indicators. Estrella and Hardouvelis (1991) also thoroughly document the significant relationship between interest rate spreads and future output growth.

Unfortunately, one of the spread's major predictive failures occurred immediately after the publication of these influential articles. Namely, the spread failed to predict the 1990–91 recession. In light of that occurrence, a number

I wish to thank Mark Watson for several useful discussions and Jeff Walker for valuable research assistance. I also wish to thank Arturo Estrella for providing some of his computer programs. Tom Humphrey, Yash Mehra, Pierre Sarte, and Roy Webb provided a number of useful suggestions. The views expressed herein are the author's and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

of papers reinvestigated the spread's predictive content. Among these are the works of Estrella and Mishkin (1997, 1998), Haubrich and Dombrosky (1996), Plosser and Rouwenhorst (1994), and Dueker (1997). These studies mainly concluded that the spread still contains significant information for predicting economic activity.

This article reinforces the view that the spread is generally a useful variable in predicting future growth in real GDP but also indicates that it has become less useful in recent years. In particular, the recent accuracy of the spread's prediction of GDP growth, both in-sample and out-of-sample, is less precise than over earlier sample periods. In fact, adding the spread to a VAR containing lagged output growth and short-term interest rates increases the root mean squared error of the out-of-sample forecast errors over the period 1985 to 1997.

After briefly reviewing relevant literature, I informally characterize the joint behavior of output growth and the spread. From this characterization it is clear that there is a relationship between the two variables, although that relationship is far from perfect. I then attempt to expand on the existing literature by analyzing the predictive content of the spread along a number of new dimensions. In particular, I examine whether there are nonlinearities in the relationship and whether the predictive content of the spread is closely associated with the stance of monetary policy. Further, the results here indicate the important differences between evaluations based on in-sample versus out-of-sample predictive power. Presumably, it is the latter that is most relevant for judging the ability to forecast.

1. RELATED LITERATURE

There is a wide and growing literature that examines the term structure of interest rates' predictive content for economic activity. The review given here is selective and focuses on articles that significantly influenced the statistical tests carried out later in this article.[1] One of the most influential studies is that of Stock and Watson (1989), which systematically attempts to construct a new index of leading economic indicators. Their approach is to examine combinations of 55 various macroeconomic variables and select the combination that best predicts future economic activity. To make their search manageable, they limit their index to seven variables—as does the current National Bureau of Economic Research (NBER) list of leading indicators. One of the variables that is an important component of their leading economic indicator is the spread between the ten-year and one-year U.S. Treasury bond. Because their search

[1] Other papers that look at the predictive content of the spread for real economic activity include Laurent (1988, 1989), Harvey (1988), Frankel and Lown (1994), Bonser-Neal and Morley (1997), and Kozicki (1997).

for a leading indicator series is fairly exhaustive, the finding that the yield spread is an important element of their indicator lent impetus to exploring the predictive content of this variable in isolation.

One article that supports using the spread alone in predicting economic growth is by Estrella and Hardouvelis (1991). Examining data over the period 1955 to 1988, they document that the spread between the yield on the ten-year Treasury bond and the three-month Treasury bill is a useful predictor of both cumulative economic growth up to four years in the future and marginal economic growth rates up to seven quarters in the future. They also find that the spread contains information for future economic growth not already embodied in the current level of real interest rates, in current economic growth, in the current growth rate of the index of leading economic indicators, or in the inflation rate. Further, they find the spread useful in forecasting the probability of a recession. An important implication of this article is its rule of thumb applicability. By concentrating largely on the spread's predictive content, the article's forecasting message is easy to apply and doesn't require sophisticated econometric tools or the application of large economic data sets.

Immediately after these two articles were written, the economy provided another test of the predictive power of the spread. In this case, although the spread narrowed and predicted somewhat weaker economic activity, it failed to predict the 1990–91 recession. As a result, other researchers revisited the issue. For example, Estrella and Mishkin (1997) examine the period 1973 to 1994 and find that the basic results of Estrella and Hardouvelis (1991) continue to hold in the United States as well as in a number of European countries. Haubrich and Dombrosky (1996) also find that over the period 1961:1 to 1995:3, the yield spread is a relatively accurate predictor of four-quarter economic growth but that its predictive content has changed over time. For example, they find that the yield spread was not a very good predictor of economic activity over the period 1985 to 1995.

Plosser and Rouwenhorst (1994) examine the predictive content of the spread between various maturities of long-term bonds and the three-month bill rates for a variety of countries over the period August 1973 to December 1988. A novel feature of their paper is the use of discount equivalent yields and the fact that they match the maturity structure of the interest rate spread with the forecast horizon being studied. They find that the term spread has significant in-sample predictive content for future cumulative changes in industrial production of up to five years but that this predictability is largely due to the spread's ability to predict activity at horizons of up to two years. Also, by looking at the effects of the term spread on forward rates, they are able to show that information in the longer end of the term structure is useful in predicting future economic activity.

Other papers have concentrated on another feature of the Estrella and Hardouvelis (1991) paper, namely, the ability of the term spread to signal

the probability of a recession. Estrella and Mishkin (1998), for example, using data over the period 1959:1 to 1995:1, show that the spread between the yield on the ten-year and three-month Treasury securities is the best out-of-sample predictor of the probability of a recession occurring in the next four quarters. For shorter horizons, they find that adding movements in various stock price indexes improves forecast accuracy. Dueker (1997) also finds that the yield spread is a relatively good in-sample predictor of recessions. He adds a lagged-state-of-the-economy variable and finds that it helps his model predict the severity and duration of big recessions; but as in other studies, he finds that milder recessions are harder to predict.

2. THE SPREAD AND ECONOMIC ACTIVITY

Before beginning a detailed statistical analysis, it is instructive to take a more casual view of the data and to consider why the spread may be a good predictor of economic activity. Figure 1 displays the behavior of (1) the spread between the discount equivalent yield on the ten-year U.S. Treasury bond and the three-month Treasury bill and (2) the four-quarter growth rate of real GDP. The NBER recession dates are shaded in. The first thing to notice is that movements in the spread precede changes in real GDP growth and that these two series are positively correlated. Thus the spread seems to indicate whether future output growth will be strong or weak. Also, prior to a number of business cycle peaks, namely, the 1969:4, 1973:4, 1980:1, and 1981:3 peaks, the spread inverted with the short-term rate exceeding the rate on the long bond. The spread also remained negative over most of these recessions. The spread flattened significantly prior to the 1990:3 peak, but as the recession progressed, the yield curve steepened. Such behavior typically indicates renewed strength in the economy. Consequently, it appears that the spread did not perform quite as well in this episode. Less-than-perfect performance is also observed around the 1957:3 and 1960:2 peaks. Further, one notices that the spread became negative in late 1967, and the economy remained strong.

Figures 2a and 2b highlight the behavior of the spread around business cycle peaks and troughs. Figure 2a reemphasizes the point that prior to most recessions, the yield curve becomes inverted and usually remains inverted for a good part of the recession. Figure 2b indicates that the yield curve, although inverted during most recessions, begins to steepen prior to each business cycle trough. Thus it seems reasonable that economic forecasters would find the yield spread a useful but imperfect guide of future economic activity.

The imprecision associated with the spread can be gauged by looking at Table 1. In this table, I record the number of true and false signals of recessions over the period 1956 to 1996. I look at two definitions of a signal. The first definition labels the signal as true if the yield curve is inverted and a recession occurs either contemporaneously or within one-to-four quarters of the signal.

Figure 1 Real GDP Growth and Spread 1956–1997

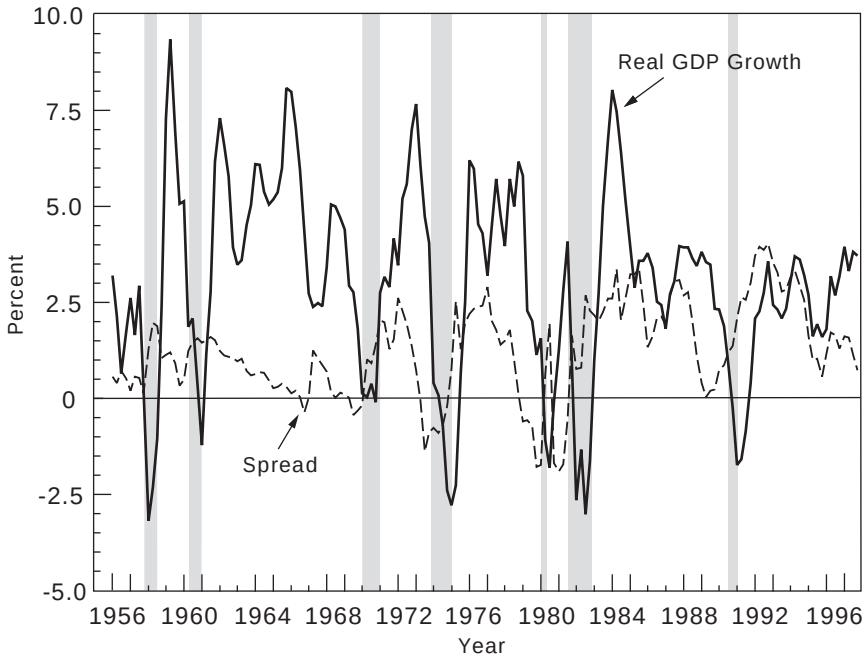


Table 1 Signal Value of the Term Spread

	Spread < 0	**Spread < 0.25**	**Spread < 0 $ff(t) - ff(t-2) > 0.5$**	**Spread < 0.25 $ff(t) - ff(t-2) > 0.5$**
True Signals	18	23	15	17
False Signals	2	13	2	8
Total Signals	20	35	17	25
True Signals within recession	8	6	2	2
Pr (True Signal)	0.833	0.552	0.867	0.652

Notes: The sample is quarterly from 1955:1 to 1995:4. The spread equals the ten-year Treasury bond rate minus the monthly average of the three-month Treasury bill rate. $ff(t)$ is the federal funds rate at time t.

Figure 2a Behavior of Spread around Business Cycle Peaks

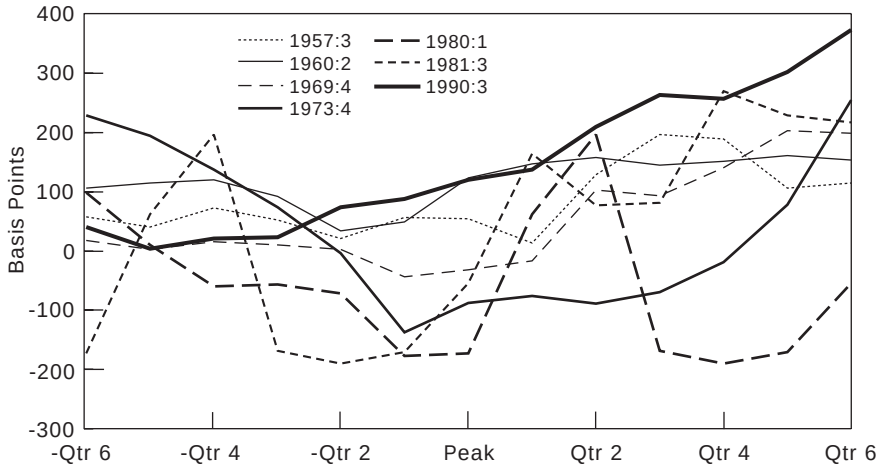
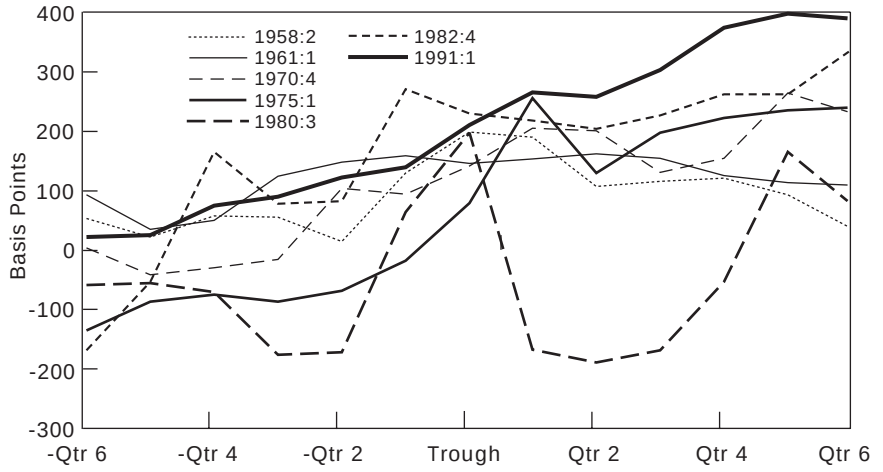


Figure 2b Behavior of Spread around Business Cycle Troughs



The second definition uses a 25-basis-point cutoff. A signal is labeled false if no recession occurs despite one of the above signals occurring. Looking at the relative frequency of true and false signals will help establish the reliability of the yield curve for predicting recessions. Note that this procedure says nothing

about instances when the yield spread failed to flatten or invert prior to a recession. The exercise lets us determine if the yield curve is like the boy who cried wolf or, in other words, if it correctly predicts a weakening in the economy.

I also investigate whether adding an indicator of monetary policy helps refine the signal. In this case a signal is labelled true if the spread inverted or was less than 25 basis points, respectively, and the funds rate was increased by more than 50 basis points in the preceding two quarters. The results in Table 1 confirm the graphical analysis that the spread is a useful but imperfect indicator of declines in economic activity. Looking at column 1, the spread inverts 18 times over the sample period, and on only two occasions does it erroneously signal a recession. Those occasions are in 1966:4 and 1979:1. The latter is labeled false only because it occurred five quarters prior to the onset of a recession. The true signals are clustered around the peaks. There are two true signals prior to and including the 1969:4 peak, three predate the 1973:4 peak, four precede the 1980:1 peak, and four precede the 1981:3 peak. Five of the occurrences are during recessions, which trivially do not signal an impending recession. Therefore, if the yield curve inverts, there is a high probability (83 percent) of an impending recession. The other columns confirm the yield curve's value as a strong signaler of a recession. Generally, most of the false signals occur in the mid- and late 1960s. Also, the character of the signals is not very different when an indicator of monetary policy is used. Consequently, there is not much evidence that the stance of monetary policy contributes to the quality of the signal.

While at first glance it appears that the spread contains information about future economic activity, it is not clear why this is the case. I am unaware of any formal economic model that investigates this issue. The spread contains direct information on a number of economic variables. Because it is a difference in nominal interest rates on bonds of different maturities, it is composed of a real term spread, the expected difference in inflation, and a term premium. Also, only temporary changes in these variables affect the spread. A permanent increase in either inflation or the real rate of interest will have the same effect on both the long- and short-term interest rates.

Often when there is an increase in expected inflation, as depicted by a steepening of the yield curve, the Fed engages in contractionary monetary policy by increasing short-term rates. In many of these episodes the long rate also initially rises, but not by as much as the short rate, and the spread narrows. Subsequently, as inflationary expectations subside, the long rate often falls and the yield curve inverts. The result of the monetary tightening is often a recession.[2] Correspondingly, when economic activity is weak, the Fed often

[2] An excellent documentation of a number of such episodes is provided by Goodfriend (1993).

loosens monetary policy by decreasing the short-term interest rate. This action generally causes the yield curve to steepen, and if an increase in inflationary expectations results from the easing, the yield curve may steepen substantially. Monetary easing often results in an increase in economic growth. Thus the result of easy monetary policy is often a steepening of the yield curve and increased economic activity. If these were the only reasons that movements in the spread were associated with economic activity, then adequately capturing the stance of monetary policy would leave little additional explanatory power for the spread in forecasting economic growth.

There are, however, other reasons why the spread may communicate future economic behavior. For example, Plosser and Rouwenhorst (1994) note that the spread's behavior is consistent with real business cycle theory. In a real business cycle model, relatively high expected future growth would imply rising real interest rates and a steepening of the term structure. The converse would occur if growth was expected to slow. Accordingly, the spread could signal expected changes in the economy that are due to nonmonetary shocks.

3. IN-SAMPLE PREDICTIVE CONTENT

In this section I examine the in-sample predictive content of the term spread between the discount equivalent yield on a ten-year U.S. Treasury bond and a three-month Treasury bill. The sample period begins in 1955:1 and extends to 1997:4. Data on the discount equivalent yield is obtained by splicing McCulloch's (1987) data set with data received from the Federal Reserve Board. The Board's data set begins in June 1961 and is not calculated in exactly the same way as McCulloch's; but except for a few years in the mid-1970s, the two series are indistinguishable. Discount equivalent yields are used for comparability purposes. I also use the ten-year, three-month spread to be comparable to most other studies but generally find that the main results of the analysis are not sensitive to the particular spread used. Results using the two-year, three-month spread and the five-year, three-month spread are similar to those reported below.

Simple Regressions

First, let's examine regressions that analyze the predictive content of the spread and various transformations of the spread for future GDP growth. I explore how sensitive the results are over different sample periods. The main finding is that the spread has predictive content for future output growth but that the regression coefficients change somewhat over different sample periods. I analyze the predictive content both cumulatively, up to two years, and marginally. Specifically, the regressions for cumulative growth are of the form

$$(400/k)\ln(y_{t+k}/y_t) = \alpha_0 + \alpha_1 s_t + e_t, \tag{1}$$

Table 2 Cumulative and Marginal Predictions of GDP Growth
(t-statistics in parentheses)

		Cumulative			Marginal		
Sample Period	***k***	α_1		$\overline{R^2}$	α_1		$\overline{R^2}$
1955:1–1997:4	2	0.96	(3.70)	0.134			
	4	0.88	(3.86)	0.182	0.80	(3.10)	0.092
	6	0.70	(3.25)	0.170	0.38	(1.68)	0.016
	8	0.53	(2.74)	0.127	−0.01	(−0.04)	−0.006
1955:1–1973:4	2	2.03	(3.24)	0.151			
	4	1.98	(4.19)	0.249	1.95	(3.40)	0.133
	6	1.41	(3.80)	0.190	0.22	(0.36)	−0.013
	8	1.29	(3.50)	0.186	0.43	(0.67)	−0.009
1973:1–1989:4	2	1.43	(6.37)	0.418			
	4	1.27	(7.04)	0.516	1.10	(3.16)	0.235
	6	1.07	(5.99)	0.528	0.71	(2.30)	0.092
	8	0.88	(6.50)	0.457	0.17	(0.54)	−0.010
1973:1–1997:4	2	1.06	(5.36)	0.274			
	4	0.95	(5.40)	0.348	0.85	(3.03)	0.172
	6	0.83	(4.98)	0.376	0.60	(2.39)	0.084
	8	0.67	(4.94)	0.322	0.15	(0.60)	−0.004
1985:1–1997:4	2	0.47	(1.82)	0.079			
	4	0.50	(1.66)	0.139	0.51	(2.53)	0.097
	6	0.55	(1.77)	0.243	0.57	(2.57)	0.125
	8	0.56	(1.94)	0.327	0.43	(1.40)	0.064

where y is quarterly real GDP and s is the spread. Values for k are 2, 4, 6, and 8. The regressions for marginal predictability are of the form

$$(400/2)\ln(y_{t+k}/y_{t+k-2}) = \alpha_0 + \alpha_1 s_t + e_t \tag{2}$$

and analyze whether the spread helps predict two-quarter output growth k periods in the future.

The first set of regression results are shown in Table 2. With the exception of the 1985 to 1997 sample period, the spread is significant at the 5 percent level in predicting cumulative output growth up to two years into the future.[3] In the latter period it is only significant at the 10 percent level. One notices, however, that the coefficients on the spread vary over different sample periods as does the informativeness of the spread as measured by the regression's adjusted R^2.

[3] All standard errors have been adjusted using the methodology suggested in Newey and West (1987). I also look at sample periods that conform to high and low inflation environments, namely, 1955:1 to 1972:4, 1973:1 to 1983:4, and 1984:1 to 1997:4, without any significant change in the nature of the results.

For example, the spread is an exceptionally good predictor of output growth over the 1973 to 1989 period.

The marginal predictive power of the spread is documented in the last two columns of Table 2. Consistent with the results in Estrella and Hardouvelis (1991) and Plosser and Rouwenhorst (1994), the spread generally has predictive content for economic growth only up to six quarters. That is, it is helpful at predicting two-quarter growth rates two quarters in the future and four quarters in the future. The spread is not informative about two-quarter growth rates at more distant horizons. Consequently, its ability to predict cumulative growth two years into the future is solely due to its strong association with near-term growth. As in the cumulative regressions, the flavor of the results would not be changed by using a spread that is composed of two-year or five-year long bond rates.

Alternative Specifications

For several reasons, one might expect that the predictive content of the spread could be improved by analyzing some alternative specifications. First, many of the episodes in which the spread inverts are also associated with contractionary monetary policy. It may be that combining an increase in the funds rate with a narrowing of the spread indicates tight monetary policy, and it is only these episodes in which the spread has predictive content. Thus the spread's signal value could be enhanced by adding an interactive term that incorporates tight monetary policy. Second, as mentioned, the spread contains a term premium that may add noise to any signal that the spread provides about the expected course of real interest rates. If this is so, then extreme values of the spread may have more predictive content than the spread itself. Also, if only large and unexpected changes in monetary policy significantly affect real economic activity, then it may be that only large movements in the spread are associated with changes in economic growth. By decomposing the spread into three components—unusually high values, normal values, and unusually low values—and by testing to determine if these different ranges imply a different relationship between the spread and economic growth, one could uncover nonlinearities in this relationship.

Specifically, the regression used for analyzing the combined effect of a monetary tightening and the spread is given by

$$(400/k)\ln(y_{t+k}/y_t) = \alpha_0 + \alpha_1 d_t s_t + \alpha_2 s_t + e_t, \tag{3}$$

where d_t is a dummy variable that takes on the value of 1 if the funds rate is raised by 50 basis points or more over the preceding two quarters.[4] To investigate the presence of nonlinearities, I run the following regression:

[4] Using cutoffs of 75 basis points or 100 basis points produces similar results.

$$(400/k)\ln(y_{t+k}/y_t) = \alpha_0 + \alpha_1 hs_t + \alpha_2 ms_t + \alpha_3 ls_t + e_t, \quad (4)$$

where hs_t takes on the value of the spread when the spread exceeds its average value by more than 0.425 standard deviations and is zero elsewhere. Similarly, the variable ls_t equals the spread when the spread is below its mean by more than 0.425 standard deviations. Otherwise it takes on the value zero. The variable ms_t equals the spread when each of the previous variables is zero and is zero elsewhere. The value 0.425 is chosen so that each variable equals the spread approximately one-third of the time. Also, the sum of the three variables is the spread itself. By dividing the spread into high, low, and intermediate values, one can check if output growth is more responsive to extreme values of the spread.

The results of these two investigations are depicted in Table 3. The sample periods are representative of the general results. The top half of the table shows that including tight monetary policy into the regressions does not significantly affect the forecasting ability of the spread. When the interactive term $d_t s_t$ is entered by itself, the adjusted R^2 is lower than in the comparable regressions using the spread by itself. Also, when both variables are entered simultaneously, only the spread retains its statistical significance.

The bottom half of the table shows the results of the analysis regarding nonlinearities. One can make a case for nonlinearity in the relationship between future output growth and the spread. Output growth responds more strongly to low values of the spread. This result may be due to the short, sharp nature of recessions, which tend to be associated with inversions in the yield curve. Both high values and intermediate values of the spread are significant over the entire sample, but high values are more likely to be significant in each subsample. Indeed, intermediate values do not have a statistically significant effect on output growth over the periods 1973:4 to 1989:4 and 1985:1 to 1997:4. For the entire sample period one can reject the equality of the coefficients. Equality, however, cannot be rejected over any of the subsamples.[5] The case for nonlinearities is, therefore, not overwhelming.

A Closer Look at the Information Content of the Term Structure

In this section I explore the additional information contained in the spread. Previous works, for example, Estrella and Hardouvelis (1991), Plosser and Rouwenhurst (1994), and Estrella and Mishkin (1997), have investigated this issue to some extent. Basically, these papers have simultaneously included

[5] The relevant statistic for the test of equality among the coefficients is distributed Chi-squared with 2 degrees of freedom. The test statistic is significant at the 5 percent level for the whole sample; the levels are 0.017 for $k = 2$ and 0.011 for $k = 4$. For the period 1955:1 to 1973:4, the significance levels are 0.843 and 0.962. For the sample 1973:1 to 1989:4, they are 0.780 and 0.775. And for the sample 1985:1 to 1997:2, they are 0.117 and 0.690.

Table 3 Predictive Content of Term Spread Using Alternative Specifications

Monetary Tightening
(t-statistics in parentheses)

Sample Period	k	α_1		α_2		$\overline{R^2}$
1955:1–1971:4	2	1.04	(2.86)			0.069
	2	0.50	(1.42)	0.81	(3.82)	0.143
	4	0.66	(2.05)			0.043
	4	0.09	(0.33)	0.85	(3.48)	0.178
1973:1–1989:4	2	1.46	(3.56)			0.208
	2	0.38	(0.92)	1.27	(5.23)	0.417
	4	1.14	(3.53)			0.205
	4	0.13	(0.50)	1.20	(5.32)	0.503

Nonlinearities

Sample Period	k	α_1		α_2		α_3		$\overline{R^2}$
1955:1–1977:4	2	0.73	(3.41)	1.33	(2.56)	2.44	(3.64)	0.181
	4	0.70	(2.97)	1.27	(2.75)	2.07	(3.69)	0.236
1955:1–1973:4	2	2.04	(3.34)	2.17	(3.35)	2.93	(2.14)	0.131
	4	2.02	(4.73)	2.00	(3.93)	2.41	(1.81)	0.229
1973:1–1989:4	2	1.31	(4.82)	1.77	(1.44)	1.76	(2.21)	0.409
	4	1.13	(4.00)	1.19	(1.47)	1.61	(3.12)	0.504
1985:1–1997:4	2	0.70	(2.00)	0.75	(1.20)	4.87	(1.99)	0.096
	4	0.62	(1.56)	0.78	(1.02)	2.45	(0.97)	0.113

another leading indicator or an index of indicators, a contemporaneous short-term interest rate or monetary aggregate, or the current growth rate of output. None of them have added a number of lags of other economic variables as is typically done in the VAR literature. Here I add two lags of output growth and four lags of the short-term nominal interest rate and test if the spread retains any significant predictive ability. The tests are performed with respect to cumulative output growth two and four quarters into the future. Thus a typical regression is given by

$$(400/k)\ln(y_{t+k}/y_t) = a_0 + \sum_{j=0}^{1}(400/k)b_j \ln(y_{t-jk}/y_{t-(j+1)k}) \qquad (5)$$

$$+ \sum_{j=0}^{3} c_j r_{t-j} + ds_t + e_t,$$

Table 4 Additional Information in Spread
(t-statistics in parentheses)

Sample Period	k	Σb		Σc		d		$\overline{R^2}$
1955:1 – 1971:4	2	0.20	(1.62)	−0.29	(−2.47)	0.48	(1.83)	0.23
	4	0.05	(0.31)	−0.30	(−2.71)	0.32	(1.22)	0.31
1955:1 – 1973:4	2	0.15	(0.80)	−0.29	(−0.99)	1.71	(1.91)	0.18
	4	0.24	(0.82)	−0.34	(−1.10)	1.74	(1.97)	0.27
1973:1 – 1997:4	2	0.25	(1.73)	−0.11	(−0.70)	0.60	(2.04)	0.27
	4	0.05	(0.32)	−0.14	(−0.99)	0.22	(1.11)	0.28
1985:1 – 1997:4	2	0.33	(2.20)	−0.50	(−1.91)	−0.28	(−0.91)	0.17
	4	0.13	(0.30)	−0.48	(−1.52)	−0.07	(−0.14)	0.14

where r is the interest rate on the three-month Treasury bill.[6] The results of this experiment are reported in Table 4. Over the entire sample period the spread is significant at the 10 percent level when predicting growth six months ahead but not statistically significant when predicting growth four quarters ahead. The spread is helpful in predicting two-quarter- and four-quarter-ahead growth rates over the 1955:1 to 1973:4 period and in predicting six-month growth over the 1973:1 to 1997:4 period. However, for this latter period the coefficient on the spread is insignificant when predicting four-quarter-ahead growth. This outcome is somewhat surprising given the results in Tables 2 and 3. Consistent with the results in Haubrich and Dombrosky (1996), the spread does not appear to be statistically significant over the most recent sample period of 1985:1 to 1997:4.

Hence, the results of this exercise indicate that the information content of the spread is reduced once other variables such as past output growth and past levels of short-term interest rates are taken into account. One must be a little guarded about the last statement. Estrella and Mishkin (1997), among others, stress that in-sample and out-of-sample predictive content are two very different things. Their work indicates that although parsimonious specifications may not perform as well in-sample, they often provide more accurate out-of-sample forecasts. In the next section, therefore, I investigate the out-of-sample predictive properties of the various models considered above.

[6] A distributive lag of past spreads was statistically insignificant. Also, longer lag lengths on past output growth were generally insignificant as well.

Table 5 Root Mean Squared Errors
Forecasts 1970:1 to 1997:4

	Start Date 1955:1					
Specification	**RMSE**	**DM**		**RMSE**	**DM**	
	1970:1–1997:4			1985:1–1997:4		
Equation 5 without the spread	2.171			1.802		
Equation 5	2.170			2.274	2.17	(0.03)
	Start Date Advances					
Equation 5 without the spread	2.171			1.848		
Equation 5	2.081	0.56	(0.58)	2.215	2.11	(0.04)
Equation 1	1.950	1.02	(0.31)	2.437	2.87	(0.00)
Equation 3	1.990	0.85	(0.40)	2.455	2.37	(0.02)
Equation 4	1.926	1.12	(0.26)	2.199	2.38	(0.02)

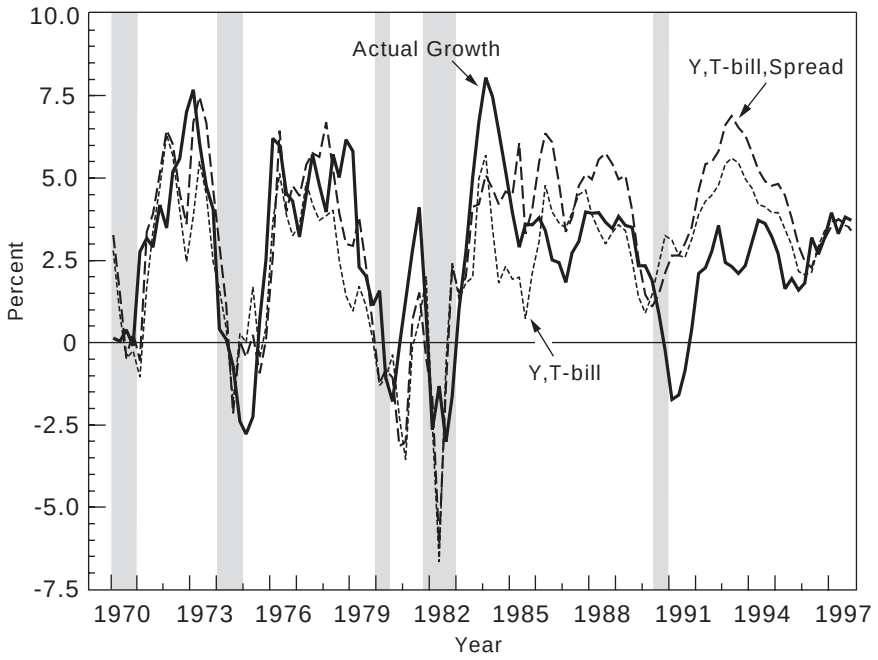
4. OUT-OF-SAMPLE FORECASTS

I now look at the out-of-sample forecast accuracy of one-year-ahead output growth for the variety of specifications considered in the previous section. The forecasts and the actual data are presented in Figures 3 through 5, and the root mean squared errors (RMSE) of the forecasts are given in Table 5. Forecasts are made over the period 1970:1 to 1997:4. The comparative predictive accuracy of the forecasts is analyzed using the methodology of Diebold and Mariano (1995) on differences of the squared forecast errors. The value of their test statistic and its significance level is reported in the columns labeled DM. This comparison is made for the entire forecasting period and for the more recent period of 1985:1 to 1997:4.

In Figure 3, the start date for the regressions is kept fixed, and the end date is continually advanced. Hence, the forecast for output growth over the period 1969:1 to 1970:1 uses data available up to 1968:4. I first examine forecast accuracy using the specification in equation (5), with and without the spread. As one sees from the two forecasts and the reported RMSE's, adding the spread does not significantly improve the out-of-sample forecasts. The root mean squared error declines almost imperceptibly from 2.171 to 2.170.

The in-sample regressions examined in the previous section, however, indicate that the coefficient on the spread varies over different sample periods. This behavior implies that a better forecasting procedure might be to roll the starting date of the regression forward as well to allow the estimated coefficients to change more rapidly. The results of this experiment are depicted in Figure 4. Here there is some improvement, with the RMSE declining from 2.171 to

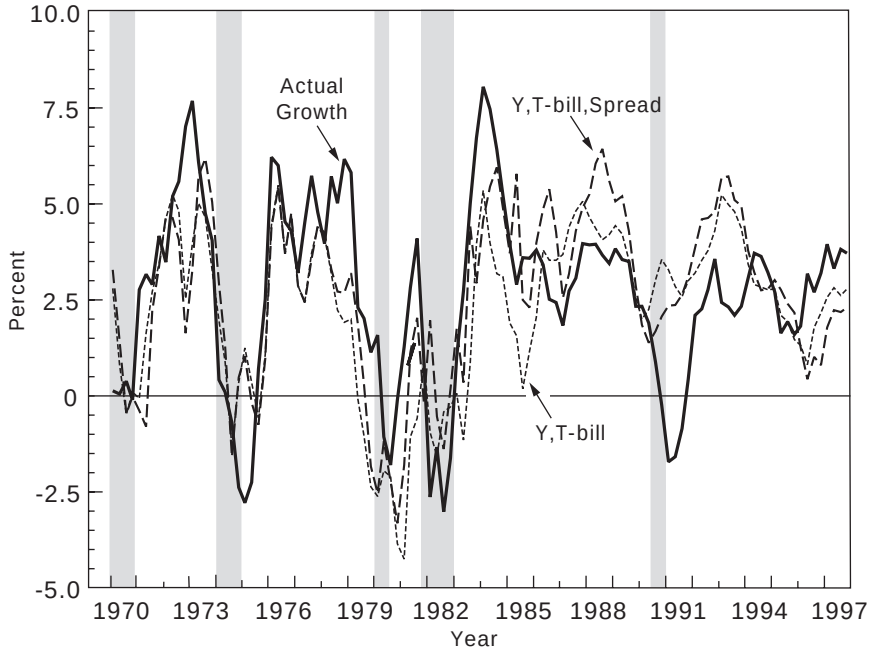
Figure 3 Actual and Out-of-Sample Predictions of Output Growth



2.081. The forecast including the spread does not overpredict the depth of the 1980 recession by quite as much as the specification without the spread, and it does not predict a sharp decline in output in 1985. The specification with the spread also indicates a slightly weaker economy in 1990 and 1991, but neither specification comes close to predicting a recession. On net, including the spread produces only a small gain in forecasting accuracy, and this gain is not statistically significant.

Surprisingly, over the entire forecasting period, only the nonlinear specification produces better out-of-sample forecasts than the spread by itself, and the improvement is minor (an RMSE of 1.926 as opposed to 1.950). Although the spread by itself produces a 10 percent increase in forecasting accuracy, as compared with a model that uses lagged values of output growth and lagged values of short-term interest rates (see Figure 5a), this increase in forecasting accuracy is statistically insignificant using the DM test statistic. Much of this gain is due to the improved forecasts in the early 1980s. Including a dummy variable that indicates tight monetary policy, as in equation (3), does not improve out-of-sample forecasting performance. Consequently, even though a parsimonious

Figure 4 Actual and Out-of-Sample Predictions of Output Growth (Rolling Regression)



specification that uses only the spread produces superior forecasts, the forecasts are not statistically significantly better.

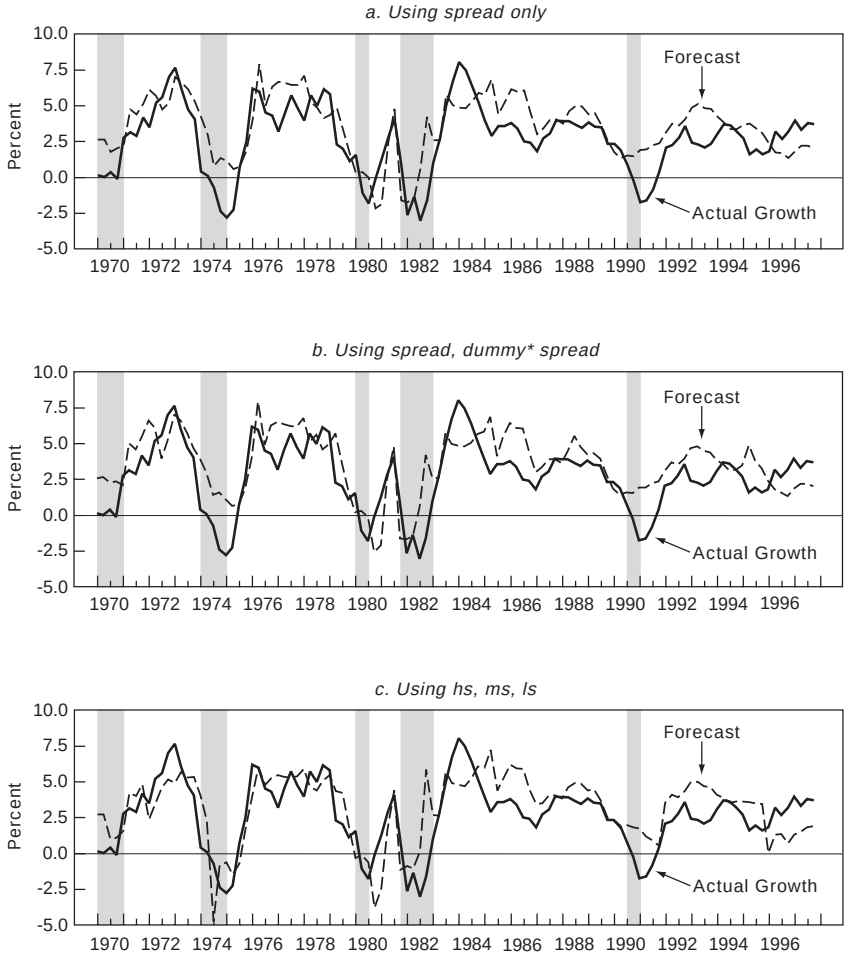
Over the more recent sample period, the results are strikingly different. Here the VAR model without the spread produces the most accurate forecasts, and these forecasts are significantly better.

5. PREDICTING RECESSIONS

In this section, I look at the ability of the spread to predict the onset of a recession using the probit model described in Estrella and Mishkin (1998). Based on the preceding section, the analysis concentrates on out-of-sample predictions but first analyzes some in-sample predictions. The relative ability of the various specifications given in equations (1), (3), (4), and (5) to accurately forecast recessions is indicated by the pseudo R^2.[7] Its values are displayed in Table 6,

[7] The pseudo R^2 is given by $1 - [\log(L_u)/\log(L_c)]^{-(2/n)\log L_c}$, where L_u is the log of the unconstrained likelihood function and L_c is the log of the maximum value of the likelihood function under the constraint that all coefficients except the constant term are zero.

Figure 5 Actual and Out-of-Sample Predictions (Alternative Regression Specifications)



as is the significance of the various coefficients in the probit regressions.[8]

As one sees from the table, the spread by itself predicts the in-sample probability of a recession relatively well. Adding a term that incorporates tight monetary policy does not help forecast recessions, nor does a specification that

[8] The significance levels for individual coefficients are corrected using the procedure in Estrella and Mishkin (1998). I wish to thank Arturo Estrella for sharing his code. The significance levels for joint tests of the coefficients on the lags of GDP growth and the T-bill rate were calculated using likelihood ratio tests that were not corrected for serial correlation.

Table 6 Significance of Variables for Predicting Recessions Using Probit Model

Specification	Variables	Significance Level	Pseudo R^2
1	spread	0.0000	0.277
2	spread	0.0004	0.281
	d*spread	0.45	
3	2 lags of GDP	0.85	0.171
	4 lags of T-bill	0.0000	
4	2 lags of GDP	0.126	0.317
	4 lags of T-bill	0.604	
	spread	0.0000	

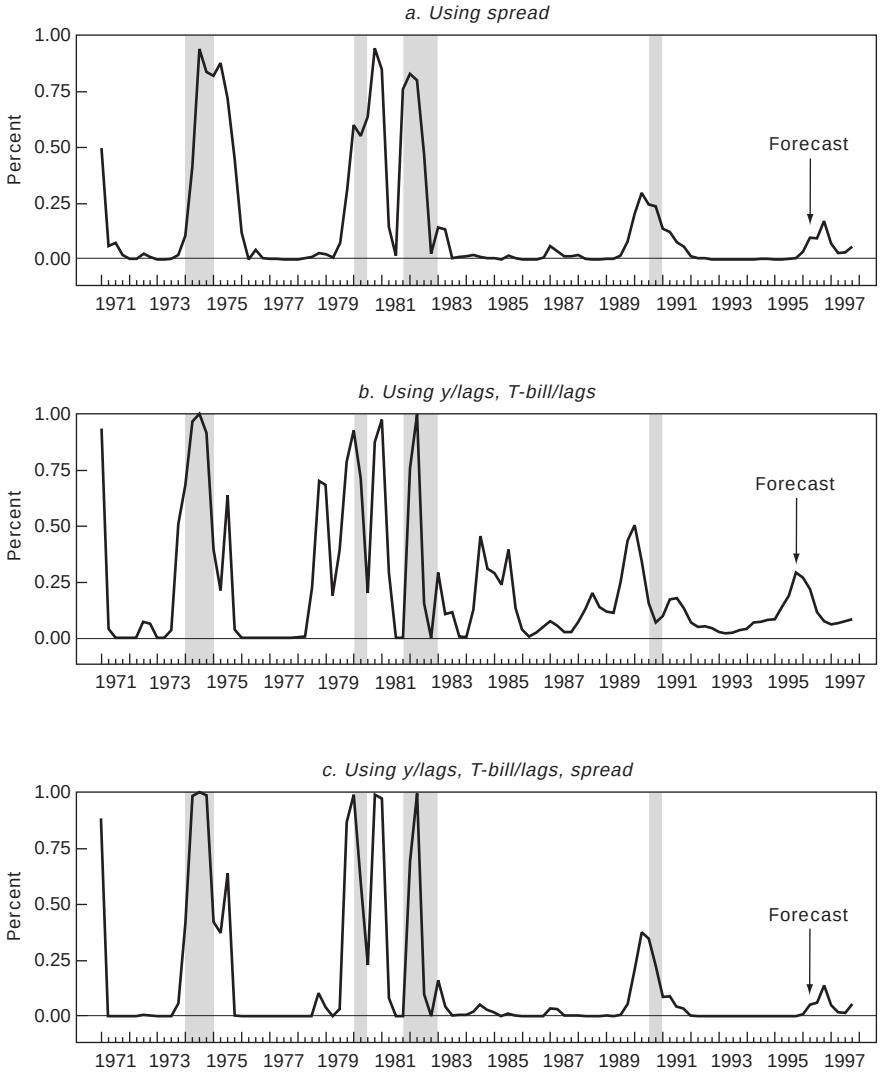
allows for nonlinear effects of the spread (the latter experiment is not reported). Adding the spread to a specification that includes lagged values of GDP growth and lagged values of the Treasury bill rate noticeably improves the in-sample forecasts of a recession.

The out-of-sample forecasts for specifications 1, 3, and 4 are shown in Figures 6a, 6b, and 6c. Only the pseudo R^2 for the specification using the spread by itself is positive and equals 0.324. The reason the pseudo R^2 is negative for the latter two out-of-sample forecasts is that the measure imposes a significant penalty for predicting a high probability of recession, when in fact no recession occurs. Also, the penalty is nonlinear, rising steeply for big forecast errors. These errors are more frequent in the latter two specifications. In some sense, though, the penalty is overly harsh because it is imposed equally whether the prediction of a recession is off by one quarter or whether the prediction occurs in the middle of an economic boom.

The three figures indicate that using the spread reduces the chance of falsely predicting the onset of a recession. This feature is particularly evident in comparing Figures 6b and 6c, where using the spread significantly reduces the probability of a recession during the mid-1980s. One also notices that while prior to the recessions in the 1970s and 1980s the three specifications forecast a high probability of recession, none of the specifications accurately signaled the 1990–91 recession. This evidence is consistent with that reported in Dueker (1997) and Estrella and Mishkin (1998).

As a final check on the spread's ability to forecast recessions, I compared its performance with that of a naive forecasting model that predicts the economy will be in its current state one quarter into the future. Even though the naive forecast uses more current information, the forecasting ability of the spread is noticeably better than the naive model. The DM statistic, which is based on

Figure 6 Out-of-Sample Predictions of Recessions



squared forecast errors, is 2.30, and the forecasts are, therefore, statistically different at the 2 percent significance level.

6. CONCLUSION

This article has investigated the forecasting properties of the yield spread for economic activity. It mainly concludes that the spread contains useful

information—information not contained in past economic activity or past monetary policy. Combined with the work of other authors, most notably Estrella and Hardouvelis (1991), Estrella and Mishkin (1997, 1998), and Plosser and Rouwenhorst (1994), the article adds to the evidence that the spread has been a useful leading indicator of economic activity. That conclusion must be tempered, however, by the observation that over more recent periods the spread has not been nearly as informative as it has been in the past. It is impossible to say whether its reduced predictive content is a function of some permanent change in the economy, or is only transitory, or is simply an outcome of examining a small sample period characterized by relatively little output variability. Given the spread's long history as a useful forecasting tool and the simplicity of its use, it will probably continue to receive significant attention in both the financial press and academic research.

REFERENCES

Bonser-Neal, Catherine, and Timothy R. Morley. "Does the Yield Spread Predict Real Economic Activity? A Multicountry Analysis," Federal Reserve Bank of Kansas City *Economic Review*, vol. 82 (Third Quarter 1997), pp. 37–53.

Diebold, Francis X., and Roberto S. Mariano. "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, vol. 13 (July 1995), pp. 253–63.

Dueker, Michael J. "Strengthening the Case for the Yield Curve as a Predictor of U.S. Recessions," Federal Reserve Bank of St. Louis *Review*, vol. 79 (March/April 1997), pp. 41–51.

Estrella, Arturo, and Gikas A. Hardouvelis. "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, vol. 46 (June 1991), pp. 555–76.

Estrella, Arturo, and Frederic S. Mishkin. "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review*, vol. 41 (July 1997), pp. 1375–1401.

_____, and _____. "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics*, vol. 80 (February 1998), pp. 45–61.

Goodfriend, Marvin. "Interest Rate Policy and the Inflation Scare Problem: 1979–1992," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 79 (Winter 1993), pp. 1–24.

Haubrich, Joseph G., and Ann M. Dombrosky. "Predicting Real Growth Using the Yield Curve," Federal Reserve Bank of Cleveland *Economic Review*, vol. 32 (First Quarter 1996), pp. 26–34.

Kozicki, Sharon. "Predicting Real Growth and Inflation with the Yield Spread," Federal Reserve Bank of Kansas City *Economic Review,* vol. 82 (Fourth Quarter 1997), pp. 39–57.

Laurent, Robert. "An Interest Rate-Based Indicator of Monetary Policy," Federal Reserve Bank of Chicago *Economic Perspectives,* vol. 12 (January/February 1988), pp. 3–14.

Newey, Whitney, and Kenneth West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* vol. 55 (May 1987), pp. 703–08.

Plosser, Charles I., and K. Geert Rouwenhorst. "International Term Structures and Real Economic Growth," *Journal of Monetary Economics*, vol. 33 (February 1994), pp. 133–56.

Stock, James H., and Mark W. Watson. "New Indexes of Coincident and Leading Economic Indicators," in Olivier J. Blanchard and Stanley Fischer, eds., *NBER Macroeconomics Annual 1989.* Cambridge, Mass.: MIT Press, 1989, pp. 352–94.

Historical Origins of the Cost-Push Fallacy

Thomas M. Humphrey

By the end of 1998, the disinflation of the 1990s had brought the U.S. price level close to absolute stability. The same disinflation witnessed a remarkable resurgence of what used to be called cost-push theories of price-level movements to explain it. Such theories, of course, attribute inflation and disinflation to a host of nonmonetary, supply-oriented influences that alter the unit cost and profit markup components of the prices of individual goods.

Cost-push theories form an integral part of the so-called *new economic paradigm,* or *new economy thesis,* which American pundits report predominantly in the popular rather than the scholarly press. Proponents of that paradigm cite such cost-reducing forces as increased global competition and rapid technological progress as the chief factors holding inflation at bay. Other frequently mentioned sources of cost disinflation—all seen as exerting downward pressure on rates of wage increase—include (1) worker job insecurity, (2) increased competition in labor markets, and (3) the declining power of labor unions in the United States.

Even these factors hardly begin to exhaust the list. Deregulation, falling computer prices, falling growth rates of health care costs: all have been proffered as cost-disinflators. Most recently, cost-push explanations of disinflation have emphasized falling import costs stemming from the Asian financial crisis, with its associated distress sale of Asian goods and plummeting foreign exchange value of Asian currencies. With such pressures holding inflation in check, cost-pushers feel free to recommend that monetary policy become expansionary in pursuit of rapid growth.

For valuable comments, the author is indebted to Roger Backhouse, Bob Hetzel, Rowena Johnson, Elaine Mandaleris, Ned Prescott, and John Walter. This article is dedicated to Professor Denis P. O'Brien, a superb scholar and historian of economic thought, who took early retirement from the University of Durham in late 1997. The views expressed herein are the author's and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

Opposed to the cost-push view is the standard monetary theory of price movements. It sees underlying monetary conditions, manifested in shifts in money supply and demand, rather than real cost-push pressure as the fundamental cause of such movements. The standard theory holds that the price level P is determined not by real cost-push but rather by the nominal stock of money relative to the real demand for it or, equivalently, by velocity-augmented money MV per unit of real output O as expressed in the celebrated equation of exchange $P = MV/O$. Conventional monetary theorists have always had a problem with the cost-push view. In their opinion, cost-push can at best explain relative prices. It cannot, however, explain the behavior of the aggregate, or general, price level. That is, it cannot do so unless it can show how cost pressures in specific sectors of the economy can markedly influence the money stock, its velocity, or the aggregate level of output—the three variables that jointly determine the general price level. Since there is no reason to think that sectoral cost pressures would materially affect these aggregate magnitudes for any substantial length of time, there is little reason to believe that cost-push theories offer a valid explanation of general price-level movements.[1] Here then is the cost-push fallacy: it confounds relative with absolute prices and sectoral real shocks with economywide nominal ones.[2] It says nothing about money's role in price determination.

Seasoned scholars accept recent cost-push theories of disinflation with a sense of déjà vu. They know that exactly the same theories—albeit with signs reversed—flourished in the 1950s, 1960s, and 1970s. Those decades saw cost-pushers attribute wage and price inflation to such forces as the increased monopoly power of trade unions, oil price shocks, the competitive struggle for relative income shares, crop failures, commodity shortages, and even the disappearance of anchovies (a key ingredient of livestock feed) off the coast of Peru. Indeed, economist George Perry (1987) of the Brookings Institution gives a fine account of the prevalence of such explanations 30 years ago.

Nevertheless, a historian would be remiss in tracing the roots of cost-push theory back no farther than the middle decades of the twentieth century. For the notion that aggregate price movements depend on real disturbances affecting the production costs (and profit markups) of particular goods is of much earlier vintage. Indeed, (1) Sir James Steuart in 1767, (2) antibullionist

[1] Of course one might argue that upward cost pressure on prices, by reducing output and employment in the affected sector (or elsewhere if demand in that sector is inelastic), may induce policymakers to increase the money stock in an effort to restore full employment. Still, nothing obligates the policymakers to take such action. On the contrary, the greater their commitment to price stability, the less likely they are to do so. The upshot is that there is no necessary, automatic linkage between cost-push and the money stock.

[2] Thus global competition, while lowering the prices of internationally traded goods, should have little effect on the general price level. Especially so as the U.S. value-added component of such goods constitutes less than 15 percent of our gross domestic product.

writers during the Bank Restriction controversy of 1797–1821, (3) the Banking School's leader Thomas Tooke in the 1840s, (4) gold standard proponents during the late-nineteenth-century bimetallism debate, (5) J. Laurence Laughlin in his 1910 comments on the post–1896 rise in U.S. prices, and (6) Reichsbank spokesmen during the German hyperinflation of 1923—all were cost-pushers with a vengeance. And if cost-push is at least two centuries old, then so too is the opposing monetary view that finds such theories erroneous. Since the early 1800s, a succession of quantity theorists including David Ricardo, John Wheatley, Henry Thornton, Knut Wicksell, Irving Fisher, Gustav Cassel, and others have criticized the theory.

The following paragraphs attempt to sketch the historical development of the cost-push view and the standard monetary critique of it. Three themes emerge. First, current cost-push theories are essentially the same as their nineteenth- and early-twentieth-century counterparts and suffer from the same defect. Second, the critics had it right: monetary policy, rather than cost-push, is what determines the path of the general price level. Third, despite its flaws, cost-push theory survives today because of its simplicity, its appeal to those whose knowledge is primarily microeconomic, and its gratifying implication that the stock of monetary purchasing power can safely be allowed to expand to meet the needs of trade.

1. SIR JAMES STEUART AND THE ORIGIN OF COST-PUSH DOCTRINE

The roots of cost-push doctrine go back at least to Sir James Steuart's 1767 *Inquiry into the Principles of Political Oeconomy,* a book Lionel Robbins describes as a "sort of compendium of all subsequent anti-quantitative theories of money" (Robbins 1971, p. 102). There Steuart enunciated at least three key strands of cost-push theory. First was his concept of the price level as a nonmonetary phenomenon determined by the same forces that determine the individual prices of specific goods. Identifying these forces as competition and cost, Steuart declared that he had "laid it down as a principle," that they determine "the standard price of every thing" (Steuart 1767, Vol. 1, p. 399; see also Screpanti and Zamagni [1993], p. 52). Increased competition, he said, forces sellers to lower prices just as falling costs also lower them. Here is the notion that real forces drive individual and aggregate prices alike.

The second strand of Steuart's cost-push doctrine supplements the first. It states that because general prices are real phenomena, they move independently of money. It denies money (metallic coin in Steuart's day) any role in price determination. "Let the specie of a country . . . be augmented or diminished in ever so great a proportion," Steuart wrote, and the prices of "commodities will still rise and fall according to the principle" of competition and cost,

"but never upon the quantity of coin" (p. 345). To explain why money has no effect on prices, Steuart advanced two arguments. First, idle hoards absorb excess coin from circulation just as they release into circulation additional coin to correct a monetary shortage. Consequently, there can be no monetary excess or deficiency to spill over into the commodity market to affect prices. The hoarding-dishoarding mechanism ensures as much. Second, changes in the stock of money that do spill over into the commodity market induce matching shifts in commodity demand and supply. In so doing, such shifts and the resulting changes in output absorb any excess coin that manages to elude the hoarding mechanism. Either way, prices remain unchanged.

The third strand of Steuart's cost-push doctrine follows logically from the second. Having denied that money drives, or governs, prices, he argued that causation runs in the opposite direction from prices to (velocity-augmented) money. Positing a two-step process, he said that cost and competition first determine prices. Then, with prices settled, the turnover velocity, or rate of use, of money adjusts to render the existing stock of coin just sufficient to accommodate real activity at the given prices. If the stock of coin is excessive, wealth holders will remove the excess (which of course being redundant yields no return in the form of convenience or liquidity) from active circulation, melt it down, and hoard it in the form of utility-yielding plate or "treasures" so that velocity falls (p. 350). Conversely, if coin is deficient, the resulting recourse to "symbolic [paper] money and a thousand other inventions" allows transactors to economize on coin whose velocity therefore rises (p. 345). Via these expedients, velocity adjusts to ensure the stock of coin is just enough to purchase all the goods offered for sale at the predetermined level of prices. In this way, causation runs from prices to velocity-augmented money. Here is the origin of the notion that changes in the stock of circulating media (coin and its paper substitutes) merely validate price changes that have already occurred and do nothing to produce such changes.

2. COST-PUSH DOCTRINES IN THE BULLIONIST-ANTIBULLIONIST DEBATE

Steuart's propositions—that cost shocks drive prices, that money cannot drive them, and that causation runs from prices to money—resonated again in the famous bullionist-antibullionist debate in England in the first two decades of the nineteenth century.[3] At that time, England, under the pressure of two harvest failures and the exigencies of the Napoleonic War, had left the gold standard

[3] On the bullionist controversy, see the classic accounts of Viner (1937, Ch. 3), Fetter (1965, Ch. 2), Mints (1945, Ch. 4), and Morgan (1943, Ch. 2). For recent interpretations, see O'Brien (1975, pp. 147–53) and Laidler (1987).

for a regime of inconvertible paper currency. The departure from gold, which released the Bank of England from the obligation to convert paper into gold at a fixed price upon demand, was followed by a sharp rise in the prices of goods, gold, and foreign exchange. Led by quantity theorists David Ricardo, John Wheatley, and Henry Thornton, one group of economists, the bullionists, blamed the Bank of England for creating inflation through excessive issues of paper notes. The Bank, they said, had simply taken advantage of the suspension of convertibility to generate an inflationary overissue of the currency. Seeking to correct this state of affairs, they recommended that England return to gold convertibility at the prewar parity as soon as possible.

An opposing group of practical businessmen and bankers, known collectively as the antibullionists, rejected this monetary explanation. Instead, they attributed the price rises to such real shocks as domestic crop failures, overseas military expenditures, and the wartime disruption of foreign trade. Like Steuart, whose work some of them may have read, they highlighted cost-push influences directly affecting the individual prices of specific commodities, notably grains and other staple foodstuffs that constituted the principal component of workers budgets. These food-price increases then passed through into money wages to raise the price of all goods produced by labor. Here is the Steuart-antibullionist notion that general price disturbances stem from nonmonetary influences affecting the individual prices of key commodities.

This notion, however, hardly went unchallenged. Bullionist writers, especially David Ricardo, criticized it for confusing relative with absolute prices. Ricardo contended that, in the absence of inflationary monetary growth, aggregate nominal demand, as measured by velocity-augmented money *MV*, would remain unchanged. With total spending (and full-capacity aggregate output) fixed, a rise in the relative price of food requiring workers to spend more on that commodity would leave them with less to spend on other goods whose prices would accordingly fall. If so, then the rise in food's price would be offset by compensating falls in other relative prices, leaving general prices unchanged.

But Ricardo's argument, with its implication that inflation must be a monetary phenomenon since it cannot stem from cost shocks to the prices of particular goods, fell on deaf ears. Unpersuaded, antibullionists continued to adhere to the cost-push idea that general price inflation stems from real disturbances affecting the particular prices of key commodities. They perceived no monetary cause of inflation.

Passive Money and Reverse Causality Propositions

On the contrary, antibullionists insisted that, since real shocks by themselves fully determine the path of prices, monetary shocks cannot serve as a contributing determinant. Two considerations, they claimed, ruled out excess money growth as a cause of inflation (see O'Brien [1975], p. 152, and Corry [1962], p. 75).

First was their *real bills doctrine*, according to which money can never be excessive if issued against the security of sound, short-term commercial bills drawn to finance real goods in the process of production and distribution. Their doctrine purported to match money creation to the value of real output so that no overissue could occur.

Second, since nobody would borrow, at any positive rate of interest, money not needed, banks could never force an excess issue on the market. Borrowers would extinguish any excess immediately by returning it to the banks to pay off costly loans. In this way, interest-minimization considerations would ensure that any excess notes instantaneously would be retired from circulation so no overissue could ever develop to put upward pressure on prices.

Both arguments embodied Steuart's passive-money notion that since real output generates just enough money to purchase it at existing prices, money cannot be an independent source of inflation. Here too is Steuart's reverse causation hypothesis that because the volume of money automatically adjusts to support real activity at predetermined prices it must be the consequence rather than the cause of those prices.

Antibullionists put these ideas to work in an effort to exonerate the Bank from blame for causing the wartime inflation. The Bank, they said, was guiltless since it had restricted its issues to real bills of exchange and so had merely responded to the real needs of trade. The Bank, in other words, could not possibly have been the source of inflation because, by limiting its advances to commercial paper representing actual output, it had merely responded to a loan demand for money already in existence and had done nothing to create that demand. Over and over again, antibullionists relentlessly insisted that money passively supplied in response to a prior demand for it could never be excessive. Indeed, as noted above, they contended that since superfluous money finds no borrowers at any interest rate, the Bank could not have overissued even if it had sought to do so. Borrowers would have thwarted any such attempt by returning the excess money to the Bank to retire loans. Flowing immediately back to the Bank, the monetary excess could never have remained outstanding long enough to cause inflation.

Bullionist Critique

Bullionists, notably David Ricardo and Henry Thornton, had little trouble exposing the fallacy of these views. In so doing, they presented the definitive classical monetary critique of cost-push theorizing.

We have already mentioned Ricardo's critique of the antibullionists' relative price theory of absolute prices. Equally fallacious, bullionists thought, was the real bills doctrine.[4] For it links the nominal money stock to the nominal

[4] Not to be confused with the doctrine of the same name advanced by Thomas Sargent and

volume of bills, a variable that moves in step with prices and so the money stock itself. By linking the variables, it renders both indeterminate. Far from preventing overissue, it ensures that any random jump in prices will, by raising the nominal value of goods-in-process and so the nominal volume of bills presented as collateral for loans, cause further increases in borrowing, lending, the money stock, spending, and prices ad infinitum in a self-perpetuating inflationary spiral. In short, the doctrine fails to perceive that price increases themselves expand the needs of trade and so generate—and justify—the very monetary expansion necessary to perpetuate them. The doctrine's flaw consists of the dynamically unstable price-money-price feedback loop established when money is allowed to be governed by the needs of trade. Far from prohibiting monetary inflation, the real bills mechanism virtually guarantees it.[5]

As for the argument that the Bank could never, at any positive loan rate of interest, force an excess issue on borrowers, bullionists observed that it overlooks a crucial point. Loan demands, and hence new money advanced to accommodate them, depend not upon the loan rate of interest per se but rather on the difference between that rate and the expected rate of profit on the use of the borrowed funds. When the expected profit rate exceeds the loan rate (as occurred to an extraordinary degree during the Napoleonic wars), borrowing becomes profitable.[6] Such profitability renders loan demands insatiable. With the Bank accommodating these loan demands with fresh issues of notes and deposits—money that spills over into the commodity market in the form of excess demand for goods—prices rise without limit. And with rising prices elevating the nominal value of goods and therefore the nominal volume of bills that represent them, those bills pass the real bills test and are accepted as collateral for additional loans. In such circumstances, the supply of eligible bills becomes inexhaustible and the real bills criterion cannot prevent overissue. Here is the classic refutation of the cost-push notion that money, because it responds passively to the needs of trade, cannot be inflationary.

With these arguments, the bullionists exposed the logical flaws inherent in each component of antibullionist cost-push theory. These components—the relative price theory of absolute price movements, the real bills doctrine, the interest-avoidance reason for the impossibility of overissue—thus emerged from

Neil Wallace (1982). As David Laidler (1984) notes, the Sargent-Wallace version of the real bills doctrine shares but one feature with its classical counterpart, namely, an inability to guarantee price level stability at a unique, determinate equilibrium level. Otherwise, it is an entirely different theory.

[5] Thornton ([1811] 1962, pp. 341–42) traced this particular real bills fallacy to John Law, who sought to limit the quantity of paper money by tying it to the nominal value of land. On Law, see also Lloyd Mints (1945, pp. 15–16, 18, 20, 30–32), the foremost twentieth-century critic of the real bills doctrine.

[6] Bullionists contended that usury ceilings constrained the Bank's loan rate to 5 percent while wartime boom conditions had raised the expected profit rate well above that level.

the debate with their validity suspect. Nevertheless, they proved impossible to kill. Though flawed, they possessed the advantage of being at once simple, transparent, intuitively appealing, and consistent with the everyday experience of practical businessmen. Illustrating the adage that popular economic theories (no matter how fallacious) never die, they survived to flourish in subsequent monetary debates.

3. THOMAS TOOKE AND THE BANKING SCHOOL

Cost-push doctrines surfaced again in the mid-nineteenth-century Currency School-Banking School debate over the need for compulsory gold backing of a currency already freely convertible into gold. (Britain had returned to such a currency when it resumed gold convertibility in 1821.)[7] In opposition to the quantity theory reasoning of the Currency School, leaders of the Banking School, notably Thomas Tooke, author of the monumental six-volume *History of Prices* (1837–1857) and preeminent collector of price data in his day, adhered to supply shock and factor cost theories of price determination.

Referring to Tooke's supply shock theory, the English banking scholar Sir Theodore E. Gregory describes how a "preoccupation with the special factors influencing particular prices" of key commodities led Tooke "to take full account of particular [price] variations" while simultaneously rejecting "the rigid connection between the quantity of money and the state of the price level postulated by the Currency School" (Gregory [1928] 1962, p. 121). Gregory notes that Tooke's list of special supply shock factors included harvest failures, extraordinary weather changes, freight rate alterations, changes in tariff rates, the erection and removal of wartime trade blockades, exchange rate movements, import cost variations, and cost-reducing technological progress embodied in machines. Modern cost-pushers updated this list in the mid-1970s when they attributed the rampant inflation then occurring to such random shocks as crop failures, the disappearance of anchovies off the coast of Peru, and the OPEC-imposed quadrupling of the price of oil. Still later, in 1997–98, cost-pushers expanded the list to include favorable import price shocks emanating from the financial crisis in East Asia.

As for Tooke's factor cost theory, it asserted that general prices owe their determination to *factor incomes* consisting of "rents, profit, salaries, and wages," rather than to money per unit of real output (Tooke [1844] 1964, p. 123). Tooke did not explain how these price-determining factor incomes themselves were determined. Instead, he left their origin open to a variety of possible causes. His theory of price movements is therefore suggestive of recent theories

[7] For classic accounts of the Currency School-Banking School debate, see Viner (1937, Ch. 5), Fetter (1965, Ch. 6), Mints (1945, Ch. 6), Morgan (1943, Ch. 4), and Robbins (1958, Ch. 5). For recent interpretations, see O'Brien (1975, pp. 153–59) and Schwartz (1987).

attributing disinflation to any one of a multiplicity of nonmonetary elements in the institutional environment such as deregulation, the removal of production bottlenecks and particular supply inelasticities, increased global competition, a decline in the power of trade unions, reductions in the nonaccelerating inflationary rate of unemployment, and the like. It is also reminiscent of theories that see inflation as the outcome of the competitive struggle for relative shares in the national income (in which the claimants' shares initially total more than 100 percent). In any event, since factor incomes are simply factor quantities multiplied by factor prices, it is obvious that Tooke came perilously close to explaining prices in terms of prices.

To illustrate how factor prices drive product prices, Tooke concentrated on falling and rising interest rates. Arguing that falling rates meant lower costs of doing business, he reasoned that these cost reductions would be passed on to buyers in the form of lower prices. The result would be price deflation even if the money stock per unit of output remained unchanged. As he put it in the famous fourteenth thesis, or conclusion, of his 1844 *An Inquiry into the Currency Principle*, "a reduced rate of interest has no necessary tendency to raise the price of commodities. On the contrary, it is a cause of diminished cost of production, and consequently of cheapness" (Tooke [1844] 1964, p. 123). Conversely, Tooke noted that rising interest rates inflate prices by boosting business costs. And they do so independently of the behavior of money.

We will return to Tooke's interest cost-push argument and its definitive critique later. Suffice it to say here that it survived into the 1950s when long-time Congressman Rep. Wright Patman of Texas, economist John Kenneth Galbraith, Council of Economic Advisers Chairman Leon Keyserling, and other populist writers argued that Federal Reserve interest rate increases are inflationary because they raise the businessman's cost of capital.

Monetary Assumptions of the Banking School

Tooke and the Banking School required one final step to complete their theory. Having attributed product price determination to real shocks affecting factor prices, they had to show why monetary shocks could not also be a contributing determinant. They ruled out money by asserting the real bills doctrine and the law of reflux, both of which they took from the antibullionists and applied to convertible-currency gold standard regimes. Arguing (1) that the stock of money could never be inflationary or deflationary if issued by way of loans made to finance real transactions in goods and services, and (2) that overissue was in any case impossible because excess notes would be returned to the banks for conversion into coin and repayment of loans, Banking School writers reiterated the antibullionist doctrines of passive money and reverse (price-to-money) causality.

Indeed, it was Tooke who stated the reverse-causality proposition most forcefully as the famous twelfth thesis of his *Inquiry.* "The prices of

commodities," he wrote, "do not depend upon the quantity of money indicated by the amount of bank notes, nor of the amount of the whole of the circulating medium; . . . on the contrary, the amount of the circulating medium is the consequence of prices" (Tooke [1844] 1964, p. 123). Elaborating, Tooke argued that factor price shocks and the resulting cost-push pressure on product prices induce corresponding shifts in the demand for bank loans to finance real activity at the altered level of prices. Banks then accommodate these loan demands via variations in the note and deposit issue. In this way, prices determine the note and deposit components of the money stock, the expansion or contraction of which are the result, not the cause, of price movements. In sum, money stock movements are validating, not causal. They merely underwrite, or validate, price changes produced by other means.

Tooke Versus Wicksell on Interest Cost-Push

It would be difficult indeed to overestimate the importance of cost-push theorizing in Banking School writings. We have seen how it led Tooke, in the fourteenth thesis of his *Inquiry,* to conclude that, no matter what happened to the money stock, a reduced rate of interest per se is deflationary rather than inflationary because it lowers business costs.

Tooke's error went largely unchallenged until the Swedish economist Knut Wicksell identified it 50 years later in his 1898 *Interest and Prices* and in Volume 2 of his 1905 *Lectures on Political Economy*. Tooke had simply failed to perceive that monetary contraction—namely, shrinkage in the stock of velocity-augmented money per unit of output—and not interest rate reduction per se is the true cause of deflation. For without such contraction, aggregate monetary expenditure MV on the nation's full-capacity output of goods and services O would remain unchanged. In such circumstances, interest rate reductions would exhaust themselves in lowering relative, not absolute, prices. The prices of capital-intensive goods—goods in which interest expense forms a relatively large share of total cost—would fall, to be sure. But such falls, by reducing the amount spent on those goods so that more could be spent on non-capital-intensive goods, would produce a compensating rise in the prices of the latter. The prices of capital-intensive goods would fall relative to the prices of non-capital-intensive goods. There would be a change in the structure, but not the overall level, of prices.[8] Absolute or general prices would remain unchanged (Wicksell [1898] 1965, p. 99; [1905] 1956, p. 180).

[8] This same relative price effect admits to an alternative explanation. The fall in the price of capital-intensive goods induces consumers to demand more of them. To supply the extra quantity demanded, producers bid labor and land away from labor- and land-intensive goods-producing industries whose costs and therefore prices must rise under the impact of bidded-up wages and rents. In this way, labor- and land-intensive goods become dearer relative to capital-intensive ones and the latter cheaper relative to the former.

Having identified the foregoing flaw, Wicksell proceeded to attack Tooke's theory on three additional grounds. First, he challenged its implicit assumption that all noninterest costs remain unchanged when interest rates fall. For if this assumption were true, then indeed aggregate costs and prices would, as Tooke asserted, fully register underlying reductions in the interest rate. But Wicksell denied the validity of the assumption. Instead, he appealed to the logic of his original charge to argue that rate reductions would induce compensating rises in noninterest costs, leaving total costs unchanged. Let rate reductions initially lower costs relative to prices, thus giving entrepreneurs an incentive to expand their operations. To expand operations, entrepreneurs must hire more labor and land. Given that those resources are already fully employed, the resulting increased demand for them bids up their prices and so raises the wage and rent components of total costs. The result is a fall in interest costs counterbalanced by a rise in wage and rent costs, leaving aggregate costs and prices unchanged (Wicksell [1905] 1956, p. 183).

Wicksell's second criticism was that Tooke's theory could not explain why bank rate increases tend to correct trade balance deficits and reverse gold outflows. For according to Tooke, such rate rises should, by pushing up domestic costs and prices relative to foreign ones, check exports, spur imports, and so worsen the trade balance rather than improving it. "If Tooke's view were correct . . . the banks would take steps which, on his theory, would lead to higher prices and to a further restriction of the already too limited export of goods" (Wicksell [1905] 1956, p. 186). The widened trade deficit and the ensuing external drain of gold to cover it would force banks to raise rates again in an effort to protect their reserves. Boosted by the rate increase, prices would rise still higher, thereby exacerbating the trade deficit. Conversely, when the trade balance is in surplus, Tooke's notion that low rates cause low prices "leads to equally absurd consequences." If he were right, then lower rates should, by reducing domestic relative to foreign prices, induce additional export surpluses paid for by inflows of gold. Banks receiving the gold on deposit would, upon finding their gold reserve augmented, reduce their rates in an attempt to work off the excess reserves. "The result according to Tooke would be a still further *fall* in domestic prices . . . so that the balance of payments would become more and more favorable and money would flow in on an ever-increasing scale" (Wicksell [1898] 1965, p. 99). In short, Tooke's interest cost-push doctrine implies, contrary to fact, that the trade balance is perpetually in unstable equilibrium, with trade deficits or surpluses progressively expanding in a monotonic explosive sequence.

Wicksell's third criticism was that money and credit markets would, like the balance of payments, be dynamically unstable if Tooke's theory were true. For if falling interest rates do indeed produce falling prices, then, with lower prices, less money is required to effect a given full-employment volume of real transactions. Their money needs diminished, transactors cut back borrowing

and pay off loans. In so doing, they return the superfluous money to the banks to swell reserves. The resulting excess reserves then induce banks to lower their rates still further, causing further falls in prices, borrowing, lending, and money circulating outside the banks. Via this sequence, a flood of excess reserves would continually inundate banks, and the rate of interest would eventually fall to zero. Conversely, rising rates, by boosting prices, would lead to greater loan demands for extra money to mediate real transactions at the higher prices. Banks, accommodating these demands through note and deposit creation, would find their reserve ratios falling. In an effort to forestall reserve deficiencies, banks would raise their rates. The result would be further price and rate increases in an endless upward spiral. "In other words, the money rate of interest would be in a state of unstable equilibrium, every move away from the proper rate would be accelerated in a perpetual vicious circle" (Wicksell [1905] 1956, p. 187). That rates in fact have been spared such dynamic instability, Wicksell wrote, is clearly a stumbling block for Tooke's theory and ample reason to reject it.

Reject it, however, is hardly what Tooke's heirs did. On the contrary, we have seen how Patman, Galbraith, and other twentieth-century American populists rehabilitated Tooke's theory to complain that the Federal Reserve spurs rather than arrests inflation when it raises interest rates. The upshot is that Wicksell's devastating critique had absolutely no impact on modern populist thinking, where Tooke's theory survives today.

4. COST-PUSH DOCTRINES IN THE BIMETALLISM DEBATE

Cost-push and conventional theories of the price level competed again during the bimetallism controversy over the proposed monetization of silver in the latter decades of the nineteenth century. At issue was the cause of the secular price deflation of 1873–1896. Bimetallists generally attributed the deflation to the failure of the gold-backed money supply to grow as fast as real output. They thought a money stock backed jointly by silver and gold circulating at a fixed ratio of 15 to 1 would have a stabler value than one backed by gold alone. Supported by two precious metals, such a stock might expand sufficiently fast to reverse the price decline and restore money's value to its pre-deflation level.

Orthodox monometallists, or gold standard advocates, however, denied that slow money growth had been the cause of falling prices. Like cost-pushers of today's new paradigm persuasion, they ascribed deflation instead to cost-reducing technological progress and to increased competition. In the words of W. W. Rostow (1948, p. 60), they

> mustered enormous evidence attesting to new methods and machines, cheapened transport costs, new raw material sources, and increased competition. They tended to deprecate the alleged monetary forces. They insisted, in short,

> that individual cost curves had fallen far and shifted to the right: that the average cost of producing a given output had decreased, and that diminishing returns—rising marginal costs—set in at a further point, requiring a higher level of demand to yield rising prices. They found in the case of each market no residual movement to be explained after its unique conditions were examined. No monetary factor was required.

Wicksell's Critique

These late-nineteenth-century cost-pushers found a worthy adversary in Knut Wicksell, whom we have already met in his role as Tooke's principal critic. Ferreting out cost-push fallacies wherever he could find them, he spared his monometallist targets no more mercy than he had spared Tooke. Noting that cost-push theories were already "so widespread" that merely to question them "would seem almost paradoxical," he proceeded to describe how they had been used to explain "the fall in commodity prices in recent decades."

> The decrease in the cost of production of commodities, the improvement of transport, etc. are often put forward without further explanation as independent causes of the fall of commodity prices . . . . It is as though this kind of explanation replaces every other theory of the value of money. The reasoning is somewhat as follows: Technological progress results in a fall in the cost of production, and so in the price, first of one group of commodities then of another. The extension of this fall in price to all, or to most, groups of commodities means a fall in the general level of prices. . . . (Wicksell [1898] 1965, p. 25)

Conversely, when inflation is the problem, cost-pushers seek the explanation "in bad harvests, in an increase in the demand for particular commodities of which the supply remains unaltered, and in the effect of tariffs and indirect taxes in raising the prices of such commodities" (Wicksell [1898] 1965, pp. 25–26). Other "alleged causes of a rise in prices" in which cost-pushers "take refuge" include "the supposed screwing up of prices by cartels and trusts, the greed of middlemen, trade union claims for higher wages, etc." (Wicksell [1905] 1956, p. 154).

As he had done for Tooke, Wicksell exposed the monometallists' confusion between relative and absolute prices. Something is wrong, he declared, "when the same causes . . . cited to account for a rise or fall in the price of *any single commodity* are put forward . . . as the source of changes in the general level of prices" (Wicksell [1898] 1965, p. 26). He proceeded to identify the error: "The proposition that prices of commodities depend on their costs of production and rise and fall with them, has a meaning only in connection with *relative* prices" (p. 99). To "apply this proposition to the general level of money prices involves a generalization which is not only fallacious but of which it is in fact impossible to give any clear account" (p. 99). According to Wicksell, there is but one way

for sectoral relative price changes to affect the general price level. That way is through the velocity-augmented stock of money per unit of output. Unless relative prices alter this monetary variable, they will have no consequence for general inflation or deflation.

Wicksell commented at length on the passive-money, reverse-causality presuppositions of cost-push theory. Modern cost-pushers, he wrote, typically regard money "as a kind of amorphous, infinitely elastic, or plastic mass which adapts itself without any pressure to any price level and is therefore entirely passive in relation to the pricing mechanism whilst the latter is regulated only by circumstances concerning the commodities themselves" ([1905] 1956, p. 154). So accustomed are these observers "to seeing in the modern credit and banking system a means of satisfying any demand whatever on the part of society for a medium of exchange that they cannot conceive of money influencing prices in one direction or the other" (p. 154). Monometallist cost-pushers, Wicksell argued, simply fail to understand that it is only through accommodative money growth (or restrictive growth in the case of deflation) that relative price changes can be translated into overall price level changes. In such cases, it is precisely the monetary accommodation (or restriction) itself rather than cost-push that changes the price level. Cost-pushers accordingly are wrong in holding that monetary accommodation merely validates price changes produced by other means. Accommodation (or the lack thereof), not cost-push, is the one absolutely necessary and sufficient condition for price changes to occur.

In overlooking this point, monometallists erred in attributing the post–1873 price deflation entirely to cost-reducing productivity shocks. It was not the shocks that produced deflation. On the contrary, prices fell because the money stock failed to grow as fast as real output.

For all its cogency and persuasiveness, Wicksell's critique of the monometallists proved no more successful than had his critique of Tooke in disposing of cost-push doctrine. Thus when J. Laurence Laughlin revived the doctrine in the early decades of the twentieth century, his critic, the quantity theorist Irving Fisher, saw the need to attack it on the same grounds Wicksell had cited. It was as if Wicksell had never written a word against it.

5. LAUGHLIN VERSUS FISHER ON COST-PUSH

If Wicksell was the harshest nineteenth-century critic of cost-push, then surely its foremost early-twentieth-century champion was J. Laurence Laughlin, the first chairman of the Economics Department of the University of Chicago, founding editor of the *Journal of Political Economy,* and leading American opponent of the quantity theory of money. The deflation of the last quarter of the nineteenth century had given way to inflation when Laughlin presented his views, first in a 1909 article in the *Journal of Political Economy* and again at a

1910 American Economic Association round-table discussion dealing with the cause of the rising prices from 1896 to 1909.

Confronting Laughlin were quantity theorists who traced inflation's cause to the Transvaal and Klondike gold discoveries and to the introduction of the cyanide process for extracting gold. They argued that the resulting huge increase in the monetary gold base and the stocks of circulating media erected thereupon had, when spent on goods, forced up prices. Laughlin, however, rejected this explanation. "The causes for the remarkable rise in prices," he declared, "cannot be looked for in those influences directly affecting gold" (Laughlin 1909, p. 263). Instead, they "must be sought in the forces settling particular prices" (Laughlin [1911] 1916, p. 178). These forces included "increased wages, higher cost of materials, higher customs-duties, and monopolies, or combinations" (Laughlin 1909, p. 266).

Laughlin described three types of cost-push mechanisms, namely, wage-push, administered pricing, and commodity shortage. On wage-push, he wrote that, with the "marked advance in wages," one "of the main elements entering into the expenses of production of all kinds of goods" has "risen in cost, and had its effect in raising prices" (p. 268). He stressed the role of ratchet effects and unilateral wage setting by trade unions. Ratchet effects imply that once "a high rate of wages has been granted, it is not easy for employers to force a reduction" (p. 268). Unilateral wage setting means that there is "an influence independent of prices which has acted to raise the rate of wages. And this influence undoubtedly is" the "pressure of labor-unions, which have been very active in recent years" (p. 269). Both phenomena imply the existence of a substantial degree of labor monopoly power even though unionized workers constituted only 6 percent of the labor force at the time Laughlin was writing.

Laughlin did not stop at wage-push. Describing the second type of cost-push or markup inflation, namely that stemming from monopoly-administered pricing, he wrote that "the formation of combinations is unquestionably the strongest force in this period working for higher prices" (p. 270). The "whole *raison d'etre* of monopolistic combinations is to control prices, and prevent active competition" which tends to drive profit markups toward zero ([1911] 1916, p. 185).

The third type of cost inflation Laughlin identified is that arising from raw material shortages, crop failures, and the like. He noted that commodity shortages drive up prices directly by reducing supply and also indirectly through their feedback into wage demands. For example, a price-wage-price feedback cycle occurs when a shortage-induced rise in the price of food "wipes out all the gains of previous increases of wages, and drives laborers to repeat their demands for higher pay, thus working again to increase expenses of production" (p. 184).

Finally, Laughlin employed his theory of *antecedent pricing* to deny money a role in price determination (see Skaggs [1995]). According to Laughlin, price

setting precedes the sale of goods. With prices settled, the stock of bank money passively adapts as required to effect the sales at the predetermined prices. Causation runs from prices to money with the latter responding endogenously to meet the needs of trade.

Irving Fisher

Laughlin found a perfect foil in Irving Fisher, America's leading quantity theorist and perhaps the greatest economist this country has yet produced. In his classic *The Purchasing Power of Money*, his remarks at the 1910 American Economic Association session on the causes of inflation, and his *Stabilizing the Dollar* (1920), Fisher took Laughlin as his target and criticized cost-push theories on four main grounds.

First, he argued that such theories fail to distinguish between changes in relative prices and changes in absolute prices. The result is confusion, with cost-pushers erroneously ascribing real and sector-specific causes to what is essentially a monetary and economywide phenomenon. In Fisher's own words, cost-pushers "have seriously sought the explanation of a general change in price levels in the individual price changes of various commodities considered separately. Much of their reasoning goes no farther than to explain one price in terms of other prices" ([1911] 1963, p. 176). Elsewhere, he listed 41 frequently cited nonmonetary causes of inflation and noted that "while some of them are important factors in raising particular prices, none of them . . . has been important in raising the *general* scale of prices" (1920, p. 11). Fisher pointed out that "no explanation of a general rise in prices is sufficient which merely explains one price in terms of another price" (p. 14).

Second, Fisher argued that anything that affects the price level must do so through changes in the stock of money, its circulation velocity, or the physical volume of trade. If these magnitudes remain constant, the price level cannot change. There is no reason to believe that changes in the specific wages of unionized labor or the prices of monopoly products will affect these macroeconomic variables. Therefore, if "trade unions seek to raise prices of labor while trusts raise prices of commodities," the general price level "cannot change" ([1911] 1963, pp. 179–80). True, the individual prices of union labor and monopoly products might rise. But these changes in particular "parts of the price level may occur only at the expense of opposite changes in other parts" (p. 180).

Fisher's third criticism referred to the tendency of cost-pushers to resort to ad hoc explanations stressing temporary disturbances, random events, and other special factors. "Mere graspings at the first straw in sight that seems to offer any explanation" is how he disparaged this practice (1920, p. 16). Cost-pushers typically "pick out some particular cases with which they happen to be familiar and drag them before the public." A crop failure renders corn dear, a firm raises

its price, a union demands higher wages—"and immediately someone hails the event as a representative cause of the high cost of living" (p. 16). Fisher termed this practice "the error of selecting special cases." He argued that because such alleged causes of inflation occur only sporadically, are short-lived, and affect only a limited range of commodities, they could not explain a sustained rise in the level of all prices. As he expressed it, "special causes working on selected commodities" would not "be general enough to explain the concerted behavior of . . . changes in the *general* scale or level of prices" (p. 16). Only excessive monetary growth could account for sustained inflation. Or as he put it, "in almost all great and prolonged price movements the chief factor is the quantity of money" (p. 52).

Finally, Fisher opposed cost-push inflation theories because they lead to what are now called price and wage controls, or incomes policies. Such "vicious remedies" he wrote, "are not only futile, but harmful" (p. 75). He further noted that while incomes policies focus directly on "the problem of the size of our incomes, they are expected to solve the second problem too," that is, the problem of inflation (p. 81). Unfortunately, since incomes policies per se cannot permanently reduce inflation if money growth remains excessive, the inevitable result is that "disappointment follows their application." In short, "unless a genuine solution" to inflation is found, "a bewildered and infuriated public is apt to keep on trying every sort of alleged remedy, good, bad, or indifferent, often with disastrous results" (p. 81).

6. COST-PUSH THEORIES IN THE GERMAN HYPERINFLATION

No sooner had Fisher offered his advice than European central bankers chose to ignore it. It was only shortly after he published a particularly blistering version of his critique in his *Stabilizing the Dollar* that Reichsbank officials were employing cost-push theories to account for the German hyperinflation debacle of 1923. That episode saw the cost of a postage stamp and a newspaper rise to 90 billion marks and 200 billion marks, respectively. At the peak of the inflation, when the money supply was expanding at a rate of 1300 percent per month and 30 paper mills were working around the clock just to supply the Reichsbank with paper for its note issue, the institution's spokesmen were insisting publicly that money growth had nothing to do with the inflation.[9]

On the contrary, they blamed inflation on external real shocks and declared that money growth was the consequence not the cause of inflation. Balance

[9] On the German hyperinflation debate, see Bresciani-Turroni's classic study ([1931] 1968) and also Ragnar Nurkse's account for the League of Nations (1946). For a recent interpretation, see Holtfrerich (1986).

of payments disturbances, they claimed, had depreciated the foreign currency value of the deutsche mark, thereby raising the prices of imported commodities. Here then was the source of the cost-push pressure. For, given the foreign currency prices of Germany's food and raw material imports, the exchange rate depreciation had raised the deutsche mark price of those specific items and therefore the prices of finished goods embodying them as ingredients. Like those who attribute our current price situation to disturbances emanating from East Asia, Reichsbank officials located the root cause of the hyperinflation in the post–World War I punitive actions of the Allies. More specifically, they traced a chain of causation running from reparations burdens (and the expropriation of German export facilities) to balance of payments deficits to exchange rate depreciation to rising import prices and thence to general price inflation onward to rising money demand and finally to the money stock itself. That is, they argued that external shocks operating through the balance of payments caused the inflation, that the resulting rise in prices created a need for more money on the part of business and government to carry on the same level of real transactions, and that it was the duty of the Reichsbank to accommodate this need, a duty it could fulfill without affecting prices. Far from seeing currency expansion as the source of inflation, they argued that it was the solution to the acute shortage of money caused by skyrocketing prices.

Critics of the Reichsbank, including Costantino Bresciani-Turroni, Gustav Cassel, Walter Eucken, Gottfried Haberler, Fritz Machlup, Ludwig von Mises, L. Albert Hahn, Karl Schlesinger, Alfred Lansburg, and others, however, had little trouble demolishing these views. With respect to the link between reparations payments and exchange depreciation, they argued that Germany could pay reparations through increased exports and reduced imports with only temporary disruptions to the balance of payments. Reparations therefore should have no lasting effect on the exchange rate whose long-run depreciation must, according to the theory of purchasing power parity, be entirely due to excessive monetary growth. Similarly, with respect to depreciating exchanges and rising import prices, they noted that neither phenomenon could persist indefinitely unless sustained by inflationary money growth. Finally, with respect to import price increases and general price inflation, they denied that the former could be transmitted to the latter provided that the money stock and hence total spending were held in check. For in the absence of monetary excess, a rise in the particular prices of imported commodities would be offset by compensating reductions in other prices leaving the general price level unchanged. The critics further noted that import prices constituted too small a fraction of total prices to affect them more than minimally, anyway. With these arguments, the critics effectively severed all the links in the cost-push chain running from reparations payments to exchange rate to import prices to general prices.

Reichsbank spokesmen, however, had one card left to play. They cited empirical evidence showing that the rate of price increase had continually

outstripped, and temporally preceded, the rate of money growth throughout the hyperinflation. The temporal lead of prices over lagging money seemed to indicate that the former caused the latter, contrary to the predictions of the monetary theory.

But anti-cost-pushers replied that this state of affairs was entirely consistent with the monetary view. Prices were advancing faster than the money stock because the public had formulated expectations of higher future rates of money growth and inflation.[10] These expectations, by raising the anticipated depreciation cost of holding marks, had greatly reduced the demand for them and had stimulated a corresponding rise in their circulation velocity. This expectations-induced rise in velocity had caused prices to rise faster than the money stock.

Reichsbank officials, however, refused publicly to acknowledge as much and continued to adhere to their cost-push, passive-money, reverse-causation doctrines. Citing the real bills theorem, they insisted that their duty was to supply the growing sums of money required to conduct real transactions and support the needs of trade at the skyrocketing (and predetermined) level of prices.

7. CONCLUSION

The longevity of cost-push theory challenges the very notion of economics as a progressive science. Any scientific discipline addressed to popular and professional audiences alike should be able to rid itself of discredited ideas once and for all. In the case of cost-push, however, economics has been unable to do so. For at least 200 years, critics have repeatedly exposed the fallacies of the theory. Yet each time it has bounced back with its popularity intact. Why does it refuse to die? What accounts for its remarkable resiliency despite its defects?

One reason, of course, is the theory's simplicity, a characteristic that renders it at once transparent, intuitively plausible, and easy to grasp by those untrained in economic analysis who ask of a theory only that it conform to everyday experience and rudimentary common sense. A related reason is its appeal to observers whose practical knowledge is micro- rather than macroeconomic. Untrained in general equilibrium macromodels, such observers may commit the fallacy of composition and so mistakenly assume that what is true for the part is necessarily true for the whole when reasoning from the particular to the general. Consider a pragmatic businessman or banker keenly attuned to the forces operating in his own market but unaware of how all markets interact. He knows how costs and competition affect the individual price he can charge. He

[10] Howard S. Ellis in his classic *German Monetary Theory 1905–1933* (1934) cites Cassel, Eucken, Machlup, Mises, Palyi, Pigou, and Robertson as holding this view.

generalizes from his own firm- and industry-specific experience to assume that these same conditions drive prices economywide as well. He becomes a cost-pusher. Lacking a macroeconomic perspective, he sees no need to understand that monetary policy actually determines prices.

A more fundamental reason for the doctrine's appeal derives from its teaching that the price level is a nonmonetary phenomenon determined by the real forces of cost and competition. With these forces holding prices in check, the doctrine implies that monetary policy is free to pursue desirable nonprice objectives such as boosting growth and achieving full employment. The doctrine, in other words, promises to liberate the central bank from its price-stabilization constraint to concentrate on other goals. Here is the latest manifestation of Sir James Steuart's idea that money stock changes unabsorbed by idle hoards induce matching shifts in commodity demand and supply such that quantities alter at unchanged prices. Here too is the old real bills idea that the money supply should be free to adapt itself to the needs of trade. Finally, here is the source of the ever-popular notion that central banks should pursue low interest rate (expansionary) policies to achieve noninflationary gains in real activity and incidentally to lower the interest component of business costs. Any borrower standing to benefit from low interest rates is tempted to subscribe to a theory that justifies them.

The enduring appeal of these ideas despite evidence of their invalidity represents a triumph of hope over experience and the source of the doctrine's long life. The doctrine seems unlikely to disappear. It will persist as long as people continue to see the price level as a nonmonetary phenomenon.

REFERENCES

Bresciani-Turroni, Costantino. *The Economics of Inflation: A Study of Currency Depreciation in Post-War Germany* (1931), translated by Millicent E. Sayers, with a foreword by Lionel Robbins. England: Augustus M. Kelley, 1968.

Corry, B. A. *Money, Saving and Investment in English Economics, 1800–1850.* New York: St. Martin's Press, 1962.

Ellis, Howard S. *German Monetary Theory, 1905–1933.* Cambridge, Mass.: Harvard University Press, 1934.

Fetter, Frank W. *Development of British Monetary Orthodoxy, 1797–1875.* Cambridge, Mass.: Harvard University Press, 1965.

Fisher, Irving. *Stabilizing the Dollar*. New York: Macmillan, 1920.

_____. *The Purchasing Power of Money: Its Determination and Relation to Credit, Interest, and Crises* (1911), revised edition. New York: Macmillan, 1922, reprinted, New York: Augustus M. Kelley, 1963.

Gregory, T. E. *An Introduction to Tooke and Newmarch's "A History of Prices"* (1928), No. 16 in Series of Reprints of Scarce Works on Political Economy. London: London School of Economics and Political Science, 1962.

Holtfrerich, Carl-Ludwig. *The German Inflation, 1914–1923: Causes and Effects in International Perspective*, translated by Theo Balderson. New York: DeGruyter, 1986.

Laidler, David. "Bullionist Controversy," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave—A Dictionary of Economics*, Vol. 1. London: Macmillan, 1987, pp. 289–94.

_____. "Misconceptions about the Real-Bills Doctrine: A Comment on Sargent and Wallace," *Journal of Political Economy,* vol. 92 (February 1984), pp. 149–55.

Laughlin, J. Laurence. "Causes of the Changes in Prices Since 1896," *American Economic Review: Proceedings,* vol. 1 (April 1911), pp. 26–36, as reprinted in C. A. Phillips, ed., *Readings in Money and Banking*. New York: Macmillan, 1916.

_____. "Gold and Prices," *Journal of Political Economy,* vol. 17 (May 1909), pp. 257–71.

League of Nations. *The Course and Control of Inflation: A Review of Monetary Experience in Europe After World War I*. Geneva: League of Nations, 1946.

Mints, Lloyd W. *A History of Banking Theory*. Chicago: University of Chicago Press, 1945.

Morgan, E. Victor. *The Theory and Practice of Central Banking, 1797–1913.* London: Cambridge University Press, 1943.

O'Brien, Denis P. *The Classical Economists*. New York: Oxford University Press, 1975.

Perry, George L. "Cost-Push Inflation," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave—A Dictionary of Economics*, Vol. 1. London: Macmillan, 1987, pp. 699–701.

Robbins, Lionel. *Money, Trade and International Relations*. New York: Macmillan, 1971.

_____. *Robert Torrens and the Evolution of Classical Economics*. London: Macmillan, 1958.

Rostow, Walt Whitman. *British Economy of the Nineteenth Century*. London: Oxford University Press, 1948.

Sargent, Thomas J., and Neil Wallace. "The Real-Bills Doctrine versus the Quantity Theory: A Reconsideration," *Journal of Political Economy,* vol. 90 (December 1982), pp. 1212–36.

Schwartz, Anna J. "Banking School, Currency School, Free Banking School," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave—A Dictionary of Economics,* Vol. 1. London: Macmillan, 1987, pp. 182–86.

Screpanti, Ernesto, and Stefano Zamagni. *An Outline of the History of Economic Thought,* translated by David Field. New York: Oxford University Press, 1993.

Skaggs, Neil T. "The Methodological Roots of J. Laurence Laughlin's Anti-Quantity Theory of Money and Prices," *Journal of the History of Economic Thought,* vol. 17 (Spring 1995), pp. 1–20.

Steuart, Sir James D. *Inquiry into the Principles of Political Oeconomy: An Essay on the Science of Domestic Policy in Free Nations,* Vols. 1 & 2. London: 1767.

Thornton, Henry. *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain* (1802), Together with his evidence given before the Committees of Secrecy of the two Houses of Parliament in the Bank of England, March and April 1767, some manuscript notes, and his speeches on the Bullion Report, May 1811. Edited with an introduction by F. A. Von Hayek. London: George Allen & Unwin, 1939, reprinted, New York: Augustus M. Kelley, 1962.

Tooke, Thomas. *An Inquiry into the Currency Principle,* 2d ed. London: Longman, Brown, Green, and Longmans, 1844, reprinted as No. 15 in Series of Reprints of Scarce Works on Political Economy, London: The London School of Economics and Political Science, 1964.

Viner, Jacob. *Studies in the Theory of International Trade*. New York: Harper Brothers, 1937.

Wicksell, Knut. *Lectures on Political Economy,* Vol. 2, "Money" (1905), translated by E. Classen, edited by L. Robbins. London: Routledge and Kegan Paul, 1956.

_____. *Interest and Prices* (1898), translated by R. F. Kahn. London: Macmillan, 1936, reprinted, New York: Augustus M. Kelley, 1965.