Darryl Francis and the Making of Monetary Policy, 1966-1975

R. W. Hafer and David C. Wheelock

Today, it is widely acknowledged that the fundamental mission of monetary policy is to maintain the long-run stability of the price level. Economists and policymakers generally agree that persistent changes in the price level (inflation and deflation) are, in the long run, caused by growth of the money stock in excess of the growth of total output. It is thought, moreover, that monetary policy can best promote high employment and maximum sustainable economic growth by maintaining reasonable stability of the price level. The charter of the European Central Bank, as well as legislation governing the behavior of central banks in several countries, specifies price stability as the sole objective for monetary policy. The Federal Reserve, by contrast, is assigned multiple policy objectives—"maximum employment, stable prices, and moderate long-term interest rates" (Federal Reserve Reform Act of 1977). Nevertheless, in recent years U.S. monetary policy has been consistent with a gradual reduction in the rate of inflation to the point where many economists believe that price level stability, for practical purposes, has been achieved.

The consensus about the importance of price level stability and the role of monetary policy is a fairly recent development. The macroeconomic paradigm that emerged from the Great Depression and dominated from the 1940s to about 1980 held that full employment should be the primary objective of monetary and fiscal policy. Stabilization policy was viewed as choosing from among a menu of unemployment and inflation rates along a stable Phillips curve. Many economists and policymakers viewed moderate inflation as an acceptable cost of maintaining full employment. During the 1950s the Federal Reserve frequently was criticized for paying "excessive" attention to inflation, to the detriment of employment and output growth.

Perhaps in part a response to such criticism, in the early 1960s the Fed's monetary policy generally became more expansionary. Inflation began to rise in 1965 and continued to increase through the 1970s. Unemployment fell at first, but during the 1970s the average rate of unemployment was higher than it had been during the preceding two decades. Moreover, inflation, unemployment, and real output growth all became more variable as the average rate of inflation increased.

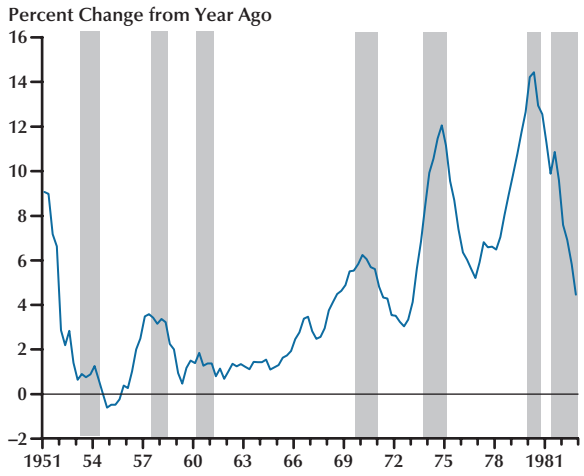
Not surprisingly, the poor performance of the macroeconomy during the 1970s brought the Federal Reserve much criticism. Among professional economists, once-dominant views about the roles of monetary and fiscal policy began to shift. Experience demonstrated the folly of those policies designed to exploit a tradeoff between unemployment and inflation and showed that expansionary monetary policy could not permanently lower the unemployment rate or increase the growth rate of real output. By October 1979, when Federal Reserve officials finally resolved to bring inflation under control, the costs of disinflating were substantially higher than they would have been earlier in the decade when inflation was lower and less entrenched.

This paper examines alternative views about monetary policy within the Federal Reserve System from the mid-1960s to the mid-1970s. We highlight the views of Darryl Francis, president of the Federal Reserve Bank of St. Louis from 1966 to 1975. In contrast to most of his Fed colleagues, Francis argued that monetary policy should concentrate on halting inflation. He believed that the influence of monetary policy on the unemployment rate was unpredictable and at best temporary. He was an early proponent of the view that the unemployment rate (and real output growth) tends toward a "natural" rate determined by factors outside the control of monetary policymakers. Francis argued that the Fed should maintain a steady growth rate of the money stock and blamed the Fed's targeting of interest rates and money market conditions for producing destabilizing swings in money stock growth.

R. W. Hafer is a professor and chairman of the Department of Economics and Finance at Southern Illinois University at Edwardsville. David C. Wheelock is an assistant vice president and economist at the Federal Reserve Bank of St. Louis. The authors thank Ted Balbach, Michael Bordo, Bob Hetzel, Garret Jones, Thomas Mayer, Allan Meltzer, Bill Poole, Bob Rasche, and Anna Schwartz for their comments. Heidi L. Beyer provided research assistance.

Figure 1

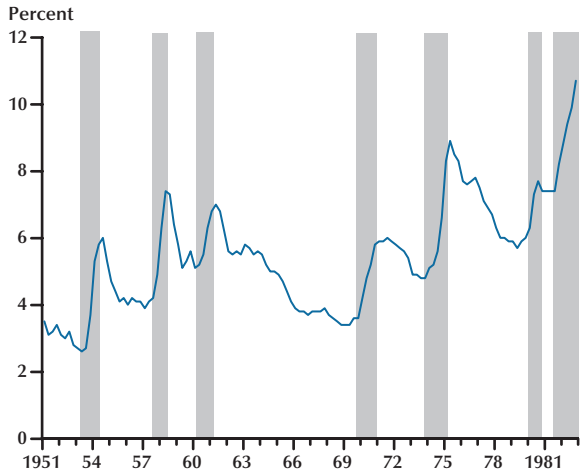
Inflation Rate
Quarterly Data, Consumer Price Index, 1951-82



NOTE: Shaded bars represent recessions.

Figure 2

Unemployment Rate
Quarterly Data, 1951-82



NOTE: Shaded bars represent recessions.

Darryl Francis's death in early 2002 prompted this historical account of his policy views and the debates within the Fed when he was president of the St. Louis Bank.[1] Reviewing the economic events and debates of this period not only provides a better understanding of the reasons behind the Fed's monetary policy actions, but also illuminates how policy views within the System evolved toward recognizing price level stability as the principal long-run objective for monetary policy.

The next two sections set the stage for our review of Francis's policy positions. First we summarize macroeconomic conditions from the 1950s through the 1970s, and then we describe the development of monetary policy from 1951 to 1966, when Francis became president of the St. Louis Fed. The subsequent section describes Francis's views about key policy issues by drawing extensively on his speeches and remarks at Federal Open Market Committee (FOMC) meetings. We highlight differences between the views of Francis and the consensus of his FOMC colleagues.

MACROECONOMIC OVERVIEW

In March 1951, the Federal Reserve and U.S. Treasury reached an agreement (the "Accord") that freed the Fed from an obligation to maintain specific yield ceilings on U.S. government securities. The agreement was sparked by a sharp increase in the rate of inflation in 1950 and early 1951 and the desire of Fed officials to halt the rise by limiting the growth of bank reserves and the money stock. Under the Accord the Fed agreed to continue to support the government securities market temporarily when the Treasury issued new debt, but yields were permitted to find their market levels as the Fed directed its focus toward containing inflation.

Inflation declined sharply in 1952 and remained low until 1956, as Figure 1 shows. After reaching an annual rate of nearly 4 percent in 1957, inflation declined to under 2 percent and remained remarkably steady until 1965. The rate of inflation then began to move upward in successive waves, with peaks in 1970, 1974, and 1980. Each peak came during a recession and followed deliberate actions by the Federal Reserve to tighten policy. In each successive cycle, however, the inflation nadir and subsequent peak were higher than those associated with the previous cycle. In 1980 the consumer price index increased at a 13.5 percent annual rate, its highest annual rate since 1947 when wartime price controls had just been lifted.

We plot the unemployment rate over the same years in Figure 2. The unemployment rate fluctuated considerably during the 1950s, then fell almost

[1] We focus on Francis's professional contributions. For more personal reflections, see Poole (2001, 2002) and Jordan (2001).

Figure 3

Phillips Curve

Annual Average of Quarterly Data, 1960-82

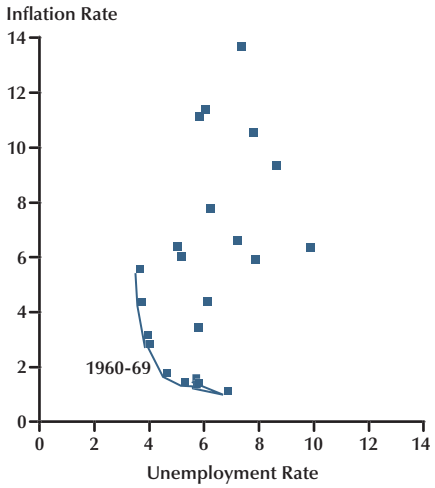
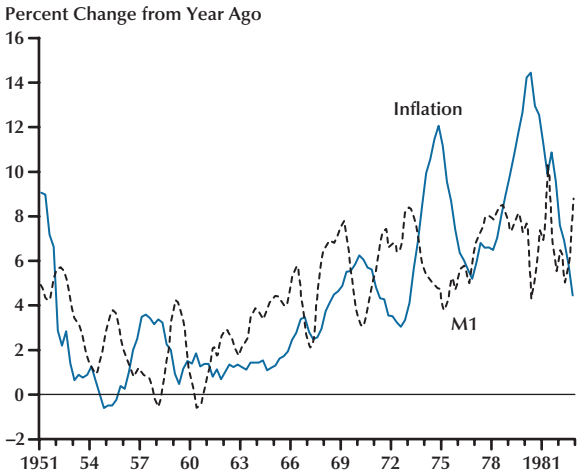


Figure 4

Inflation Rate and M1 Growth

Quarterly Data, 1951-82



continuously from 1963 to 1969 to end the decade below 4 percent. Much of the decline came as inflation was rising, suggesting that Federal Reserve officials had revised their preferences in favor of a lower unemployment rate and were willing to accept higher inflation as the cost of pushing the unemployment rate down.

The unemployment rate did not continue to fall during the 1970s, even though inflation continued to rise. As Figure 2 shows, the unemployment rate increased sharply during the recession of 1970; though it declined during the subsequent recovery, it did not fall below 5 percent. Another recession in 1974-75 pushed the unemployment rate above 8 percent. In the subsequent recovery, the rate again fell to a low point that was higher than that of the previous expansion. Finally, during the 1981-82 recession the unemployment rate peaked at over 10 percent—its highest level since the Great Depression.

Although short-run peaks in the unemployment rate tended to occur when the inflation rate was falling, the negative correlation between annual rates of unemployment and inflation that characterized the 1960s was absent during the 1970s and early 1980s. As shown in Figure 3, unemployment and inflation rates appear to have followed a predictable Phillips curve pattern—higher unemployment rates associated with lower inflation—during the 1960s. From 1970 to 1982, however, the correlation between unemployment and inflation rates was low. Moreover, both rates followed upward trends over the period, which ran counter to a view commonly held during the 1960s that expansionary monetary policy could permanently lower the average rate of unemployment.[2]

It is beyond the scope of this paper to identify the sources of specific changes in inflation or unemployment during the 1960s and 1970s. Oil price shocks in 1973 and 1979 and other supply-side disturbances are often blamed for much of the adverse movements in unemployment and the price level during the 1970s.[3] The increasing trend rate of inflation is today widely attributed to a rising average growth rate of the money stock. The association between money stock growth and inflation is illustrated in Figure 4, where we plot the growth rate of M1, a narrow monetary aggregate, alongside the inflation rate.[4] The figure illustrates that money growth and inflation moved inversely in the short run, reflecting the Fed's attempts to tighten policy in response to higher inflation. Over the longer term, however, the upward trend in the rate of inflation was associated with a similar trend in money stock

[2] Hafer and Wheelock (2001) provide a summary of alternative views during the 1960s about the association between inflation and unemployment in the long run.

[3] See Barsky and Kilian (2001) for an alternative view.

[4] We plot M1 growth because it was the aggregate favored by Darryl Francis and Federal Reserve Bank of St. Louis staff. M2 growth behaved similarly, however, during the period illustrated here.

growth. Like inflation and unemployment, M1 growth rose and fell in waves, with both growth rate peaks and troughs as high or higher than those of the previous cycle.

MONETARY POLICY FROM THE ACCORD TO 1966

We assert that neither the trend nor the variability of money stock growth, inflation, or the unemployment rate during the late 1960s and the 1970s reflected a well-designed monetary policy. To provide some information on how policy decisions were made during these years, we briefly review economic policy developments in the period preceding the "Great Inflation" of 1966-80.

During the 1950s, monetary policy focused largely on the threat of inflation. Inflation fell sharply in 1952 when the Fed began to exercise its new independence. Following the 1953-54 recession, however, inflation seemed poised to increase again. Federal Reserve Chairman William McChesney Martin vowed not to repeat the mistake of the previous expansion, when interest rates were not increased fast enough to curb inflation. Consequently, the Fed tightened policy in 1956 and maintained its stance even as economic activity began to slow. Although a few FOMC members called for an easier policy, the majority thought that continued restraint was needed to avoid "sloppy" financial markets and to contain inflationary momentum.[5]

The Fed's concern about potential inflationary momentum was heightened in 1957 when, contrary to widespread expectations, the price level failed to decline as economic activity began to slow.[6] The Fed maintained its anti-inflation posture until mid-1958, easing only when policymakers had become convinced that inflation was falling. M1 growth exceeded 6 percent during the final two quarters of 1958 after having fallen at about a 2 percent annual rate during the first half of that year. Monetary policy remained expansionary until late 1959 when a tighter policy caused M1 growth to fall. The economy entered yet another recession in the second quarter of 1960.

Even though inflation had remained low, slow economic growth and recurrent recessions—1953-54, 1957-58, and 1960-61—led to criticism of the Fed's policies. One group of economists—who were later labeled "monetarists"—blamed the Fed's "stop-go" policy actions, and resulting swings in money stock growth, for much of the instability in the macroeconomy.[7] Other economists claimed that persistently tight monetary policy had contributed to the economy's tepid growth and high unemployment, though many considered monetary policy less effective than fiscal policy for stabilizing economic activity.

The principal economic advisors in the Kennedy administration were prominent Keynesians who favored the use of fiscal policy tools to stimulate rapid economic growth.[8] In the Kennedy administration, writes Okun (1970, pp. 40-41), "the standard for judging economic performance [focused on] whether the economy was living up to its potential rather than merely whether it was advancing…As long as the economy was not realizing its potential, improvement was needed and government had a responsibility to promote it."

The *Economic Report of the President* for 1962 outlined the problem as Kennedy's advisors saw it: "Expectations in 1962 were colored by the suspicion that underutilization was to be the normal state of the American economy…[and] inadequate demand remains the clear and present danger to an improved economic performance" (1963, p. 23). The *Report* stated explicitly that "demands originating in the private economy are insufficient by themselves to carry us to full employment…[and] the Federal Government can relax its restraints on the expansionary powers of the private economy" by reducing taxes and reforming the tax system (1963, p. 32).

Where did monetary policy fit into this scheme? The consensus view, both outside and inside the Fed,

[5] Stein (1969) notes that the Fed was supported in its policy by the Eisenhower administration and by many in the academic community.

[6] The business cycle peaked in August 1957, and the downturn continued to April 1958. A common view at the time was that this recession was "only an interruption in the inflationary pressure, and the fact that it did not result in any decline of the price indexes was considered highly ominous" (Stein, 1969, p. 319).

[7] The FOMC focused on interest rates and free reserves (i.e., excess less borrowed reserves), not money stock growth, in its policy deliberations. When the Fed desired a tighter policy, it would use open market operations to reduce (or limit the growth of) bank free reserves to increase the market yields on Treasury securities. Similarly, to ease monetary policy, the Fed would increase free reserves to reduce market yields. Friedman (1960) and Brunner and Meltzer (1964), among others, argued that this approach caused undesirable swings in money stock growth that interfered with the Fed's ability to achieve the broad policy objectives of price stability, low unemployment, and economic growth.

[8] The belief that fiscal policy was a potent tool in economic stabilization was not confined to the White House. Fed Chairman Martin (1961, p. 279), testified to the Joint Economic Committee on March 7, 1961, that in the fight against inflation "undue reliance has perhaps been placed on monetary policy. I can readily agree with those who would have fiscal policy…carry a greater responsibility for combating inflation."

was that monetary policy should accommodate the needs of fiscal policy, which meant keeping interest rates low. Although nominally constrained by the continuing balance of payment deficits, monetary policy was generally consistent with the Kennedy administration's desires. Okun (1970, p. 55) writes that "the Fed did not 'lean against the wind' during 1961-65. As long as the economy continued to operate below its potential and prices remained stable, the Fed was prepared to provide the liquidity to sustain the advance."[9]

Extended summaries of the FOMC *Memoranda of Discussion* corroborate Okun's view, although there probably was more concern expressed about a possible resurgence of inflation in FOMC deliberations than in White House meetings. For example, at an FOMC meeting on December 17, 1963, Federal Reserve Chairman Martin commented that the "whole western world was again faced with the specter of inflation…and he was opposed to inflation because it led to deflation. There were those who believed that unemployment could be cured by easy money. He doubted this…Budget, fiscal and wage-price policies had more fundamental effects" (FOMC, December 17, 1963, p. 55-56).[10]

The Phillips curve was cemented firmly into the policy calculus of administration advisors and many Federal Reserve economists.[11] Using this framework, the president's advisors estimated that if the economy were operating at its potential, the unemployment rate would be approximately 4 percent and the inflation rate would be about 2 percent.[12] Fiscal and monetary policies were not considered adequately expansionary if the unemployment rate rose above 4 percent.

After declining steadily since 1961, full employment (i.e., a 4 percent unemployment rate) was achieved in 1965. Although many economists and policymakers recognized that expansionary policies could lead to higher inflation, consumer price inflation appeared to be contained. Wholesale prices, however, began to rise rapidly in 1965. With federal budget deficits also expanding, fears of higher inflation were ignited. Nevertheless, by September 1965, Fed Chairman Martin opined that price pressures were not sustainable and that "it would be desirable to keep to the status quo, with the [Open Market] Desk maintaining market conditions on as even a keel as possible at this juncture" (FOMC, September 28, 1965, p. 94).

Martin's views changed quickly. Inflation became more apparent as 1965 was drawing to a close. Despite pressure from the White House, Martin and other Fed officials began to advocate a more restrictive policy.[13] Martin stated at an FOMC meeting in late November that "if any Reserve Bank should come in with an increase in the discount rate he would be prepared to approve" (FOMC, November 23, 1965, p. 87). On December 6, 1965, the discount rate was increased from 4 to 4.5 percent. The Board of Governors was deeply divided over the increase—four members voted to approve the increase and three opposed. The Johnson administration and some members of Congress were publicly critical of the Fed's move, with some administration officials even questioning whether the Fed should have the power to act independently.[14]

Foreshadowing later episodes, the Fed's effort to contain inflation was short-lived. Monetary policy tightened further in mid-1966, but the Fed soon relented under pressure that intensified when interest-sensitive sectors of the economy began to show signs of weakness. By early 1967 monetary policy, as measured by the growth of monetary

[9] The March 26, 1963, FOMC meeting, which ended in agreement to not change course, produced a policy directive that is representative of the times: "This policy [to accommodate moderate growth in bank credit and minimize capital outflows] takes into account the continuing adverse United States balance of payments position and the increases in bank credit, money supply, and the reserve base in recent months, but at the same time recognizes the limited progress of the domestic economy, the continuing underutilization of resources, and the absence of general inflationary pressures" (FOMC, 1963, p. 47).

[10] The *Memoranda of Discussion* are not verbatim transcripts of FOMC meetings, but rather summaries of statements made by meeting participants.

[11] Some administration advisors helped develop the Phillips curve for policy use. See, for example, Samuelson and Solow (1960). See Taylor (1997) for a discussion of how use of the Phillips curve led to an inflation bias in policy.

[12] See Hetzel (1995) for additional detail. The impression one gets from interviews with former Fed officials is that the FOMC did not explicitly use the Phillips curve in its discussions. Still, the policy discussion available in the FOMC *Memoranda of Discussion* suggests that such a tradeoff was recognized and affected policy decisions. See Mayer (1995, 1999).

[13] St. Louis Fed president Harry Shuford, Francis's predecessor, argued for monetary restraint at an FOMC meeting on November 2, 1965. He and a few others recognized that "The economy was operating near capacity, and at this time the rate of increase in spending appeared to be faster than the growth in ability to produce" (FOMC, 1965, p. 23). Fed Governor Charles Shepardson concurred, stating that "the rate of recent expansion was unsustainable, and at some point steps must be taken to try to dampen it" (FOMC, 1965, p. 34).

[14] For example, Gardner Ackley, Chairman of the Council of Economic Advisors, argued that "The Federal Reserve System is part of the government, and should be responsible to the administration" (cited in Hetzel, 1995, p. 19).

aggregates, had once again become extremely expansionary.[15]

THE FRANCIS YEARS

Darryl Francis became president of the Federal Reserve Bank of St. Louis in 1966. In the tradition of his predecessors, he was an outspoken critic of the Fed's monetary policies.[16] Francis strongly supported the goal of halting inflation, but felt that the Fed's actions in late 1965 and early 1966 had been too timid. At an FOMC meeting on May 10, 1966, he noted that the monetary aggregates were continuing to grow rapidly, which he attributed to the Fed's reluctance to allow interest rates to rise. At the subsequent FOMC meeting on June 28, he pointed out that "while it was generally believed that interest rates had been rising in a restrictive manner during the past year, they had, in a very real sense, not done so. The cost of money to the borrower and the return to the saver were affected by changes in the value of the dollar. When one adjusted market interest rates for the decline in the value of the dollar [i.e., for inflation]…one found that interest costs had not risen at all in the past year…During the year market interest rate increases had provided no restriction to the excessive total demand" (FOMC, June 28, 1966, p. 65).

Measured by Francis's preferred gauge of monetary policy—the growth of monetary aggregates—policy became considerably tighter as 1966 progressed. By autumn, Francis voiced concerns that monetary policy had become too tight: "Monetary developments since last spring had been restrictive… Member bank reserves had declined moderately, growth of bank credit had slowed markedly, and the money supply had changed little on balance… Care now had to be taken to avoid becoming too restrictive…Steps should be taken to avoid any sustained monetary contraction, as well as to avoid a renewal of the rapid monetary expansion that occurred last winter and spring" (FOMC, November 1, 1966, pp. 78-79).

Francis's statements at FOMC meetings during his first year in office reflected fundamental views about monetary policy that he shared with other monetarists. The quotes above, for example, illustrate his belief that the stance of monetary policy is measured appropriately by the growth rates of monetary aggregates, not the level of interest rates, and that the Fed should keep the money stock growing at a steady pace, rather than allow it to fluctuate widely. His calls for targeting the money stock growth rate and for focusing monetary policy exclusively on containing inflation, while gaining some support in academic circles, put him at odds with most of his Fed colleagues. In this section, we examine Francis's policy pronouncements in detail and how they challenged the prevailing consensus among Federal Reserve policymakers.[17]

Monetary Policy and Employment

Many of Francis's policy views would not be controversial today, but fell outside the mainstream during his tenure at the Fed. For example, a dominant view among macroeconomists at that time was that the government should respond to any shortfalls in employment or output growth. The Fed was widely accused of having been overly concerned with preventing inflation during the 1950s, which many economists claimed had kept the unemployment rate higher than necessary.[18] Although reasonable stability of the price level was seen as desirable, many economists and policymakers argued that modest inflation was an acceptable cost of achieving high employment. Moreover, many claimed that any inflation that did occur when the economy was at less than full employment was due not to monetary policy but to "excessive" increases in wages or other costs.

Although widely held, the mainstream views about inflation and the role of monetary policy did not go unchallenged. Friedman (1968) and Phelps (1967) argued that the unemployment rate would tend toward a "natural rate" in the long run, irrespective of the rate of inflation. Friedman preached that "inflation is always and everywhere a monetary phenomenon" and argued that fluctuations in money stock growth historically had been a principal

[15] See Cagan (1972) for a detailed discussion of monetary policy during this period.

[16] Francis's predecessors at the St. Louis Bank, Delos Johns and Harry Shuford, also argued for the use of monetary aggregates in the conduct and description of policy. This tradition no doubt reflected the influence of Homer Jones, who was the director of research at the St. Louis Fed from 1958-71. Jones's influence is examined in a special volume of the *Journal of Monetary Economics* (1976).

[17] Although Francis's intellectual debt to his research staff and others should not be ignored, it was Francis who advocated these unpopular ideas and new research results in FOMC policy discussions and public venues.

[18] For example, see the views expressed by participants in a symposium on (then) recent monetary policy in the *Review of Economics and Statistics* (1960).

cause of short-run fluctuations in real output and employment. By fixing money stock growth at the long-run growth rate of real output (adjusted for the trend growth of velocity), Friedman claimed that the price level would remain stable and monetary policy would not contribute to business cycle fluctuations.[19]

Francis shared many of Friedman's views and advocated them in policy discussions. Francis decried attempts to use monetary policy to control the unemployment rate, claiming that "Use [of monetary policy] as a short-run stabilizing tool produces costs in terms of lost employment and output and undesired price level movements" (1972, p. 34). Further, he argued, "I am convinced that future stabilization of our economy depends heavily upon a moderate and stable growth of the money stock. But if the pronouncements of critics of the monetarist view are heeded, the result will most likely be erratic fluctuations in the money stock caused by attempts to 'fine tune' the economy. Such fluctuations will necessarily cause periods of inflation and will be frequently accompanied by unacceptable levels of unemployment" (1972, p. 38). In Francis's view, "stop-go" monetary policy, by which he meant abrupt shifts from slow to rapid growth of the money stock, was an important cause of fluctuations in output growth and inflation: "Only by eliminating the stop-go stabilization actions…could [monetary] policy makers permanently improve the total social welfare and avoid acting as the architects of successive waves of intensifying inflation and recession" (FOMC, April 6, 1971, pp. 77-78).

Today, Francis might be described as an inflation "hawk" because he often argued that monetary growth was too fast and inflation too high. In the fall of 1966 and again in October 1969, however, Francis pressed for an easier monetary policy because he believed that monetary growth was too slow and the danger of recession was high. On the latter occasion he argued that "The studies made at his Bank indicated…that, if the System did not permit some growth in key monetary aggregates beginning now, an unacceptable economic recession would most likely develop in 1970, which in turn might force the Committee into [an] inordinate monetary expansion" (FOMC, October 28, 1969, pp. 54).[20] Francis was prescient: The U.S. economy entered a recession in the fourth quarter of 1969.

Francis attributed the increasing trend rate of inflation that began in the mid-1960s to the Fed's persistent attempts to hold the unemployment rate below a level consistent with price stability. At the time, the consensus among most policymakers and economists was that a 4 percent unemployment rate represented full employment. In hindsight, it is now widely believed that the "natural rate" was really 5 percent or higher throughout the 1970s.[21] Even though Francis probably had no more insight about the natural rate of unemployment than any other Fed policymaker, as early as 1970 he questioned whether a 4 percent rate of unemployment could be achieved without generating higher inflation: "When spending was rising fast enough to keep the unemployment rate at about 4 percent, strong upward pressure was exerted on prices and price expectations…Much of the current unemployment was structural and could not be obviated except temporarily and with adverse price effects by stimulation of total spending" (FOMC, August 18, 1970, pp. 44-45). In 1971 he again noted that "In the last decade whenever the unemployment rate had been below 5 percent inflation had accelerated, largely because of labor market imperfections" (FOMC, April 6, 1971, p. 30).

The Cause or Causes of Inflation

Friedman and other monetarists believed that the impact of monetary policy on real output and employment was transitory: Over time, monetary policy affected only the price level, while sustained movements in the price level were caused solely by growth of the money stock in excess of total output growth. At the time, however, many economists and policymakers attributed inflation to imperfections in labor or product markets, expansionary fiscal policy, shortages of raw materials, and other non-monetary forces. Federal Reserve Chairman Arthur Burns, for example, blamed the inflation of the 1970s on increases in wages and other costs in excess of productivity gains. In a speech given in December 1970, Burns complained that "Governmental efforts to achieve price stability continue to be thwarted by the continuance of wage increases substantially

[19] See, for example, Friedman (1960) or Friedman and Schwartz (1963).

[20] Francis often cited and introduced into the record of FOMC meetings research results produced by his research staff. Of these, perhaps the most controversial was that of Andersen and Jordan (1968).

[21] For example, the *Economic Report of the President* for 1977 recognized that productivity growth had slowed substantially in the late 1960s and estimated that the "full employment rate of unemployment" was approximately 5.5 percent (1978, pp. 45-57).

in excess of productivity gains…The inflation that we are still experiencing is no longer due to excess demand. It rests rather on the upward push of costs—mainly, sharply rising wage rates." He argued, moreover, that "Monetary and fiscal tools are inadequate for dealing with sources of price inflation such as are plaguing us now—that is pressures on costs arising from excessive wage increases" (Burns, 1978, pp. 112-13).[22]

Burns often made similar comments at FOMC meetings. For example, at a meeting on April 7, 1970, he suggested that "The inflation that was occurring—and that was now being accentuated…was of the cost-push variety. That type of inflation…could not be dealt with successfully from the monetary side and it would be a great mistake to try to do so" (FOMC, April 7, 1970, p. 50). Whereas Burns viewed wage increases as the dominant cause of inflation during the 1970s, he blamed expansionary fiscal policy for the initial increase in inflation during the mid-1960s: "The current inflationary problem emerged in the middle 1960s when our government was pursuing a dangerously expansive fiscal policy …Our underlying inflationary problem…stems in very large part from loose fiscal policy" (Burns, 1978, p. 177).[23]

Francis disagreed. Unlike some of his contemporaries on the FOMC, Francis did not confuse changes in relative prices with persistent increases in the general level of prices. While monetary policy could affect the latter, changes in relative prices were caused by market forces beyond the Fed's control. In a February 1972 speech, for example, Francis argued that "In the long run the growth of output and employment is determined by the growth of resources of a society…The trend growth of prices is determined by the trend growth of money stock relative to growth in output…Deviations from a trend rate of growth of money…cause short-run deviations in output and employment…But once the adjustment is completed, output and employment will resume their longer-run growth paths" (Francis, 1972, p. 33). In another speech, Francis noted that "other factors have an influence on the movement of prices in a given year. But when we talk about the 'problem of inflation,' I think it is safe to say that the fundamental cause is excessive money growth" (Francis, 1974, pp. 6-7).[24] As a policy issue, the distinction between changes in relative prices and inflation became even more important when petroleum prices increased sharply in 1973.

The Cure for Inflation

In light of their differing views about the cause of inflation, not surprisingly Burns and Francis disagreed about how to end inflation. By the late 1960s inflation clearly was on an upward trend. As Francis pointed out in early 1969, "For about four years… the [Federal Open Market] Committee had been led into unintended inflationary monetary expansion while following interest rate, net reserves, and bank credit objectives and the even keel constraint. He suggested that, if the Committee meant business now, it should try some other guides" (FOMC, February 4, 1969, p. 47). Specifically, Francis sought an operating procedure that focused on controlling the growth of money. His view, from which he did not waver during his ten years on the FOMC, was that "the cure [for inflation] is to slow down the rate of money expansion" (Francis, 1974, p. 7).

In contrast, Burns, other members of the FOMC, and administration economists promoted wage and price controls as the only viable policy for stopping inflation. "The persistence of rapid advances of wages and prices in the United States and other countries, even during periods of recession," Burns argued, "has led me to conclude that governmental power to restrain directly the advance of prices and money incomes constitutes a necessary addition to our arsenal of economic stabilization weapons" (Burns, 1978, p. 156).[25] At an FOMC meeting on June 8, 1971, Burns argued that "Monetary policy could do very little to arrest an inflation that rested so heavily on wage-cost pressures...A much higher rate of unemployment produced by monetary policy would not moderate such pressures appreciably...He intended to continue to press [the administration]

[22] Burns made these statements in a speech titled "The Basis for Lasting Prosperity," given December 7, 1970.

[23] Burns made this statement in a speech titled "Key Issues of Monetary Policy," given July 30, 1974.

[24] Even though Burns later admitted that "Inflation cannot continue indefinitely without an accommodating increase in supplies of money and credit" (Burns, 1978, p. 208), he argued that inflation could continue well after monetary stimulus was removed, even during a period of rising unemployment: In 1971 he argued that "inflation caused by excess demand became entrenched, and remained after demand-side pressures abated. Entrenched inflation, increased militancy of labor, and willingness of business to accede to labor's wage demands, explains continued rising prices during periods of rising unemployment" (Burns, 1978, p. 126). Burns made this statement in a speech titled "The Economy in Mid-1971," given July 23, 1971.

[25] Burns made this statement in a speech titled "Some Problems of Central Banking," given June 6, 1973.

hard for an effective incomes policy" (FOMC, June 8, 1971, p. 51). On August 15 of that year, President Nixon unveiled his New Economic Program and introduced the first of three phases of direct wage and price controls.

Francis was highly critical of government controls on prices and wages, as they simply disrupted market signals. At an FOMC meeting in December 1967, he suggested that "Selective credit controls, wage freezes, and price restrictions had been advocated as alternatives [to contain inflation]. Such controls, however…raised problems of resource allocation; they interfered with freedom; and they were difficult to administer" (FOMC, December 12, 1967, pp. 54-55). In December 1970, Francis again argued at an FOMC meeting that "The adoption of administrative controls in attempting to hold down inflation, or to shorten the period of adjustment, would impose a great cost on the private enterprise economy. Serious inefficiencies would develop in the operations of the market system" (FOMC, December 15, 1970, p. 74). While such controls might mask inflation for a time, "a freeze or other control programs could not be expected to effectively restrain inflation unless accompanied by sound monetary actions" (FOMC, October 19, 1971, p. 36). In Francis's view, low rates of inflation could not be achieved over the long run unless the money stock grew at a rate approximately equal to the long-run growth rate of real economic activity. Wage and price controls, to Francis, were merely impediments to the efficient working of a free market.

Money Versus the Money Market

The money stock did not grow at anything like the steady rate that Francis and other monetarists advocated. They attributed wide swings in money growth to the Fed's strategy of targeting market (i.e., nominal) interest rates. During World War II and for several years afterward, Federal Reserve open market operations were aimed primarily at maintaining low and stable yields on U.S. Treasury securities. The Federal Reserve–Treasury Accord of 1951 removed the Fed's obligation to maintain ceilings on Treasury security yields, but both yields and the general "condition" of the Treasury securities market remained important concerns of open market policy. In particular, the Fed typically would act to prevent market yields from changing whenever the Treasury issued new debt—a policy known as maintaining an "even keel."

The Fed used this "money market" strategy to implement policy throughout the 1950s, 1960s, and 1970s. Francis was highly critical of the approach because it detracted from his preferred policy of stable growth of the money stock.[26] Moreover, he eschewed the use of market interest rates as a guide for policy because their movement did not always reflect policy actions. While rising interest rates often were considered a sign of monetary policy tightening, Francis noted that rising rates also could reflect rising inflation, the outcome of an expansionary monetary policy.

From the first FOMC meetings he attended, Francis chided the Committee for previous policies that, in his view, contributed to uncertainty over the stance of policy. For example, at a meeting on May 10, 1966, Francis observed that "There had now been ten or eleven months when the directive had continuously called for a moderation or restriction of expansion in bank reserves, bank credit, and money, and at the same time called for only slightly firmer money market conditions. Those instructions [have] been inconsistent…[and have led to] very rapid increases in bank reserves, bank credit and the money supply" (FOMC, May 10, 1966, p. 49). He expressed this view often during the late 1960s, both at FOMC meetings and in public forums. Speaking to a group of financial market practitioners in New York City in 1968, Francis argued that "Measures of money market conditions such as market interest rates and free reserves have been shown to be poor indicators of the influence of monetary actions." And, "for stabilization purposes, movements in interest rates should be viewed no differently than movements in commodity prices" (Francis, 1968, p. 8).

The FOMC never abandoned money market conditions or interest rates as policy targets. In 1970, however, FOMC policy directives began to include specific targets for the growth of money and bank credit, as well as for money market conditions. Frequently the objectives for money and credit were in conflict with those specified for interest rates, and the latter were usually permitted to take precedence. Citing such conflicts, Francis voted against two policy directives in 1973 because he did not believe that the monetary growth rates specified in those directives—which he agreed with—would be

[26] Francis was not the first Federal Reserve Bank president to criticize the money market approach. The president of the Federal Reserve Bank of Atlanta, Malcolm Bryan, argued against the approach in the 1950s in favor of targeting a monetary aggregate (Meigs, 1976; Hafer, 1999).

achieved given the money market objectives the directives also specified.[27] The failure to achieve the monetary growth objectives set by the FOMC led Francis to argue for making public the FOMC's targets and its record of achieving those targets: "The records should contain a clearer description of the whole process of making and implementing policy, including information on targets that were missed and on those that were hit" (FOMC, December 17, 1973, p. 14).

In contrast to Francis, most FOMC members were unwilling to discard interest rates or money market conditions as proximate objectives for monetary policy. Burns sometimes made statements at FOMC meetings favoring tighter control of the growth of monetary aggregates. He more frequently spoke against monetarist policy prescriptions, however, both at FOMC meetings and in public comments. For example, at an FOMC meeting in early 1971, Burns argued that "the heavy emphasis that many people were placing on the behavior of M1 involved an excessively simplified view of monetary policy" (FOMC, February 9, 1971, p. 87). And, in congressional testimony in 1975, he stated: "There is a school of thought that holds that the Federal Reserve need pay no attention to interest rates, that the only thing that matters is how this or that monetary aggregate is behaving. We at the Federal Reserve cannot afford the luxury of any such mechanical rule…We pay close attention to interest rates because of their profound effects on the working of the economy" (Burns, 1978, p. 369).[28]

Supply Shocks

Francis's policy views were shaped and supported by considerable empirical research conducted by St. Louis Fed staff, as well as economists outside the System. The St. Louis Fed formulated a simple, yet highly accurate forecasting model and began to publish forecasts in the Bank's *Review* in April 1970 (Andersen and Carlson, 1970). Like other models, however, the St. Louis model seriously under-forecast inflation in 1973-74 and the decline in real economic activity in 1974-75. Burns noted this in testimony before the House Committee on Banking and Currency in July 1974: "Inflationary tendencies and monetary expansion are not as closely related as is sometimes imagined. For example, the econometric model of the St. Louis Federal Reserve Bank, which assigns a major role to growth of the money stock in movements of the general price level, has seriously underestimated the rate of inflation since the beginning of 1973…Apparently, special factors… have been at work" (Burns, 1978, p. 176).[29]

Francis acknowledged that "With respect to inflation…the rise in prices in 1974 was just about double the increase that he would have expected to result from the policy actions that had been taken. Special factors, such as the energy and agricultural problems, had contributed to the rise in prices in 1974" (FOMC, December 17, 1974, p. 99). Francis reiterated his earlier position that observed changes in the price level caused by changes in relative prices associated with supply disturbances could not persist indefinitely. As he had argued almost a decade earlier, over time inflation was a monetary phenomenon. In a speech in October 1974, for example, he stated that he was "not willing to accept the special factor explanation of inflation because that explanation removes the focus from inflation as a monetary phenomenon" (Francis, 1974, p. 5).

In policy discussions, Francis warned against tightening excessively in response to a temporary increase in the price level caused by supply shocks, claiming that the special "factors would not continue to exert strong upward pressure [on inflation] in 1975, and the rate of inflation would subside" (FOMC, December 17, 1974, p. 99). At the same time, however, he also warned against excessive easing in response to the ongoing recession because it too had been caused by the supply shocks and not a lack of demand. At an FOMC meeting in January 1974, he argued that "the actual and prospective slowdown in economic activity resulted wholly from capacity, supply, and price-distorting constraints and not from a weakening in demand. Therefore, to ease policy and allow a faster rate of monetary growth would be to increase inflationary pressures without expanding real output or reducing unemployment" (FOMC, January 22, 1974, p. 102). And, in December of that year, he argued that "The current decline in economic activity differed from past recessions in a number of respects. First, it was one of the few declines, if not the only one, to have developed without having been preceded by stabi-

[27] See FOMC, July 17 and August 21, 1973.

[28] Burns made this statement in testimony titled "Monetary Targets and Credit Allocation" to the Subcommittee on Domestic Monetary Policy, U.S. House Banking, Currency, and Housing Committee, February 6, 1975.

[29] Burns made this statement in testimony titled "Key Issues of Monetary Policy," given July 30, 1974.

lization policy actions that brought it about. Second, there had been an absolute decline in the country's capacity to produce, caused by the agricultural and energy problems, by the distortions resulting from the wage and price controls, by the new environmental and safety standards, and by changes in foreign exchange rates" (FOMC, December 17, 1974, p. 99). Rapid money stock growth, Francis argued, could do little to affect the growth of real output or employment in such a circumstance and would result mainly in a higher rate of inflation.

CONCLUSION

Darryl Francis served as president of the Federal Reserve Bank of St. Louis during tumultuous economic times. Even so, Francis's views about monetary policy reflected an underlying set of beliefs from which he did not waver. The "four basic premises" that guided him in his policy prescriptions were set out in an early speech and reprinted in the Federal Reserve Bank of St. Louis *Review* in 1968 (Francis, 1968). These premises are as follows: "First, a predominantly market orientation." Francis firmly believed in the unequaled efficiency of free markets to allocate incomes and goods and services. "Second, quantification is essential." In contrast to most of his FOMC colleagues, Francis consistently fought for quantifiable policy rules and measures of the success or failure of policy actions. "Third, our economic system is more stable than was believed a few years ago." Francis believed that, over time, real economic growth was determined by population growth, capital formation, and technology. Monetary policy, in his view, could not reliably improve on market outcomes in the short run, or increase real output growth (or lower the unemployment rate) in the long run. Although not the accepted wisdom in his time, such a view today is fundamental. "Fourth, monetary management is more properly directed toward influencing changes in total spending." Francis questioned attempts to use monetary or fiscal policies to affect specific markets or sectors of the economy. Although actions taken to achieve price stability could impinge more on some sectors than on others, Francis argued that free markets would adjust to such actions. Allocation of goods and services or resources by market forces, he believed, was preferable to allocation by government decree.

Francis did not think about monetary policy in terms of forward-looking, dynamic rules or deep theoretical models. He had strong convictions about the efficacy of market forces and the limitations of government stabilization policies. Even though Francis believed that monetary policy could exert a powerful short-run impact on the unemployment rate, he was convinced that it could not be used to permanently steer the economy to any particular rate of growth. In the long run, Francis believed that monetary policy affected only the price level. Maintaining price stability, Francis believed, would help establish conditions that would foster maximum employment and economic growth. Although not widely shared among his contemporary Federal Reserve colleagues, today such views are mainstream.

REFERENCES

Andersen, Leonall and Jordan, Jerry L. "Monetary and Fiscal Actions: A Test of Their Importance in Economic Stabilization." Federal Reserve Bank of St. Louis *Review*, November 1968, *50*(11), pp. 11-24.

_____ and Carlson, Keith M. "A Monetarist Model for Economic Stabilization." Federal Reserve Bank of St. Louis *Review*, April 1970, *52*(4), pp. 7-25.

Barsky, Robert B. and Kilian, Lutz. "Do We Really Know That Oil Caused the Great Stagflation? A Monetary Alternative." NBER Working Paper No. 8389, National Bureau of Economic Research, July 2001.

Brunner, Karl and Meltzer, Allan H. *The Federal Reserve's Attachment to the Free Reserve Concept*. Washington, DC: House Committee on Banking and Currency, 1964.

Burns, Arthur F. *Reflections of an Economic Policy Maker: Speeches and Congressional Statements: 1969-1978*. Washington, DC: American Enterprise Institute for Public Policy Research, 1978.

Cagan, Phillip. "Monetary Policy," in Phillip Cagan et al., eds., *Economic Policy and Inflation in the Sixties*. Washington, DC: American Enterprise Institute for Public Policy Research, 1972, pp. 89-154.

Economic Report of the President (various years).

Federal Open Market Committee. *Memoranda of Discussion* (various dates).

Francis, Darryl R. "An Approach to Monetary and Fiscal Management." Federal Reserve Bank of St. Louis *Review*, November 1968, *50*(11), pp. 6-10.

_____. "Has Monetarism Failed?—The Record Examined." Federal Reserve Bank of St. Louis *Review*, March 1972, *54*(3), pp. 32-38.

_____. "Inflation, Recession—What's a Policymaker To Do?" Federal Reserve Bank of St. Louis *Review*, November 1974, *56*(11), pp. 3-7.

Friedman, Milton. "The Role of Monetary Policy." *American Economic Review*, March 1968, *58*(1), pp. 1-17.

_____. *A Program for Monetary Stability*. New York: Fordham University Press, 1960.

_____ and Schwartz, Anna Jacobson. *A Monetary History of the United States: 1867-1960*. Princeton: Princeton University Press, 1963.

Hafer, R.W. "Against the Tide: Malcolm Bryan and the Introduction of Monetary Aggregate Targets." Federal Reserve Bank of Atlanta *Economic Review*, First Quarter 1999, *84*(1), pp. 20-37.

_____ and Wheelock, David C. "The Rise and Fall of a Policy Rule: Monetarism at the St. Louis Fed, 1968-1986." Federal Reserve Bank of St. Louis *Review*, January/February 2001, *83*(1), pp. 1-24.

Hetzel, Robert L. "William McChesney Martin and Monetary Policy in the 1960s." Unpublished manuscript, Federal Reserve Bank of Richmond, April 1995.

Jordan, Jerry. "Darryl Francis: Maverick in the Formulation of Monetary Policy." Federal Reserve Bank of St. Louis *Review*, July/August 2001, *83*(4), pp. 17-22.

Journal of Monetary Economics. "Contributions in Honor of Homer Jones." November 1976, *2*(4), pp. 431-71.

Martin, William McChesney Jr. "Federal Reserve Operations in Perspective." *Federal Reserve Bulletin*, March 1961, *47*(3), pp. 272-81.

Mayer, Thomas. Interviews, special collections, general library, University of California–Davis, 1995.

_____. *Monetary Policy and the Great Inflation in the United States: The Federal Reserve and the Failure of Macroeconomic Policy, 1965-79*. Cheltenham, UK: Elgar, 1999.

Meigs, A. James. "Campaigning for Monetary Reform: The Federal Reserve Bank of St. Louis in 1959 and 1960." *Journal of Monetary Economics*, November 1976, *2*(4), pp. 439-53.

Okun, Arthur M. *The Political Economy of Prosperity*. Washington, DC: The Brookings Institution, 1970.

Phelps, Edmund S. "Phillips Curves, Expectations of Inflation and Optimal Unemployment Over Time." *Economica*, August 1967, *34*(135), pp. 254-81.

Poole, William. "Eulogy for Darryl R. Francis, 1912-2002." Federal Reserve Bank of St. Louis *Review*, March/April 2002, *84*(2), pp. 1-2.

_____. "President's Message." Federal Reserve Bank of St. Louis *Review*, July/August 2001, *83*(4), pp. 5-6.

Review of Economics and Statistics. "Controversial Issues in Recent Monetary Policy: A Symposium." August 1960, *42*(3), pp. 245-82.

Samuelson, Paul A. and Solow, Robert M. "Problems of Achieving and Maintaining a Stable Price Level: Analytical Aspects of Anti-Inflation Policy." *American Economic Review*, Papers and Proceedings, May 1960, *50*(2), pp. 177-94.

Stein, Herbert. *The Fiscal Revolution in America*. Chicago: University of Chicago Press, 1969.

Taylor, John B. "America's Peacetime Inflation: The 1970s: Comment," in Christina D. Romer and David H. Romer, eds., *Reducing Inflation: Motivation and Strategy*. Chicago: University of Chicago Press, 1997, pp. 276-80.

NAFTA and the Geography of North American Trade

Howard J. Wall

This paper estimates the effects of the North American Free Trade Agreement (NAFTA) on the geographic pattern of North American trade. Specifically, it looks at the effects of NAFTA on aggregate trade flows between subnational regions within North America and between North American regions and the non-NAFTA world. The importance of a regional analysis of the effects of NAFTA is evident from the variety of regional post-NAFTA experiences. Between 1993 and 1997, real trade between Canada and the United States increased by more than 50 percent. Over the same period, Central Canadian exports to the Southwest and Rocky Mountain regions of the United States and Eastern Canadian exports to the Southeast of the United States all increased by more than 110 percent. In contrast, Eastern Canadian real imports from the Great Lakes, Plains, and Southeast regions of the United States were actually lower in 1997 than they were in 1993. Further, although real Canadian exports to Mexico increased by 46 percent over the period, those from Western Canada rose by over 90 percent while those from Eastern Canada rose by less than 1 percent.[1]

Viner (1950) established the general principle that the welfare effect of joining a preferential trading area (PTA) such as NAFTA is ambiguous. This is because PTAs create a distortion between the trading conditions that member and nonmember countries face. In a simple partial-equilibrium model under perfect competition, a PTA will increase trade between members, whether countries or regions, because the tariff between them has been eliminated (trade creation). If the most efficient producer of a good is outside the PTA, the effect is to import more from the less efficient member-producer (trade diversion). The net effect of a PTA on trade volume (as a proxy for welfare) would depend, therefore, on the relative sizes of trade creation and trade diversion.

Despite the presumed certainty of trade creation and trade diversion, the ways in which integration affects trade are many and varied, and few fit into the simple Vinerian dichotomy. One significant non-Vinerian way for integration to affect trade volumes is through increasing returns to scale, a topic typically absent from the empirical literature, although prominent in the theoretical literature. It has also been central to the public discussion of North American integration, as, for example, Canadian firms have long argued that access to the U.S. market would allow them to exploit economies of scale. This access would allow them to increase their exports not only to the rest of North America, but also to the rest of the world. Increasing returns also affects the volume of trade in inputs and intermediate goods used by increasing-returns industries. This is because firms that expand production and exploit economies of scale need to purchase more inputs and intermediate products, which might be imported from inside or outside North America. Thus, in contrast with the Vinerian effects, with economies of scale, NAFTA may increase trade between members *and* between members and nonmembers.

The effects from trade creation, trade diversion, and scale economies arise whether one looks at trade from a national or a regional standpoint, and they would drive much of the regional variation in the effects of a PTA. As with countries, regions differ in their abilities to match their comparative advantages to the preferences of consumers in other member and nonmember regions. However, the recent literature under the rubric of the "new economic geography" suggests that things are actually much more interesting when account is taken of firms changing their locations as a response to joining a PTA. This literature, spearheaded by Krugman

[1] See Tables A1 through A4 in Appendix A for the available percent differences in real region-to-region, region-to-country, and country-to-country trade between 1993 and 1997. See also Krueger (2000) for a broader discussion of the changes in trade between NAFTA partners.

Howard J. Wall is a research officer at the Federal Reserve Bank of St. Louis. He thanks Cletus Coughlin and seminar participants at the Spring 2000 Midwest International Economics Conference at the University of Kentucky, West Virginia University, the Federal Reserve System Regional Meetings, and the Association for University Business Economic Research Conference for their comments. Ling Wang provided research assistance; Abbigail J. Chiodo provided editorial assistance.

(1991a,b), models various ways in which production patterns (and therefore trade patterns) can change with integration because of its effects on firms' optimal location decisions.[2]

One of the reasons that a PTA affects geographic trade patterns is that it alters the spatial distributions of firms' customers and suppliers. For example, consider a firm initially located in Massachusetts. By adding Mexico to the Canada–United States Free Trade Area, the spatial distributions of the firm's customers and suppliers are shifted southward, creating greater incentive for the firm to move closer to Mexico, if not into Mexico itself. If the firm relocates, regional trade patterns will change because goods that were exported from Massachusetts to Canada, Mexico, and the rest of the world would instead be exported from, say, Arizona. At the same time, because the firm has moved across the continent, it is in a better position for exporting to Asia and a worse position for exporting to Europe. Also, the firm would be more likely to import intermediate products from Asia, and the regional import pattern would change accordingly.

A second reason that a PTA affects geographic trade patterns is that it expands the set of possible places for firms to locate. Under NAFTA, Canadian and U.S. firms that move to Mexico can do so without losing tariff-free access to their domestic markets. This affects intra-NAFTA trade by switching what had been exports, say, from Canada to the United States and Mexico, into exports from Mexico to the United States and Canada. Extra-NAFTA trade would also be affected, as a firm that was exporting from Canada to the rest of the world would instead export from Mexico. The new economic geography literature suggests that these location effects are much stronger when there are cross-firm linkages whereby a firm's marginal costs are lower when other firms are nearby. With these linkages there is a tendency for linked firms to cluster, creating industry centers.

These examples are by no means exhaustive, but they do provide sufficient illustration of the theoretical inadequacies of the Vinerian dichotomy. In this vein, my estimates of the regional effects of NAFTA serve a more general purpose. Specifically, they provide support for the hypothesis that, because it does not account for the spatial or geographic effects of integration, standard customs union theory (the Vinerian dichotomy) is inadequate for capturing the effects of preferential trading areas. The geographic approach is a break from standard empirical analyses of PTAs in that it recognizes that the nation is not always the relevant unit of reference for international trade (Krugman, 1991a). I find ample evidence that the effects of NAFTA have not conformed to the Vinerian dichotomy and conclude that the customs union theory needs to be reworked to include a substantial accounting of geography and scale economies.

The empirical model that I use, the gravity model, has become the workhorse for estimating the effects of PTAs on trade volume. In a gravity model, bilateral trade is assumed to be an increasing function of the national incomes of the trading partners and a decreasing function of the distance between them. The effects of PTAs are modeled with dummy variables. For my present purposes, the gravity model has advantages and disadvantages, both arising from its simplicity. While it allows me to examine the effects of NAFTA on a large number of trading combinations, it is not versatile enough to attribute the effects on aggregate trade to trade creation, trade diversion, the mobility of firms, agglomeration, etc.

From a practical standpoint, the major advantage of the gravity model is that the researcher does not need to specify the underlying trade processes, although that it is largely *ad hoc* has meant that the gravity model has met with much suspicion by international trade theorists. Deardorff (1984, p. 504), however, concluded that gravity models tell us "something very important about what happens in international trade, even if they do not tell us why." Recently, though, the gravity model has "gone from an embarrassing poverty of theoretical foundations to an embarrassment of riches."[3] In fact, as shown by Bergstrand (1985, 1989) and Deardorff (1998), among others, the gravity model can be derived within a variety of standard theoretical frameworks. The estimates I present below demonstrate vividly the greatest strengths and weaknesses of the gravity model. On the one hand, its simplicity allows for the estimation of a large number of region-to-region NAFTA effects that would be extremely difficult to obtain using any other method. On the other hand, it provides little guidance to explain why the NAFTA effects occur. Nevertheless, the results do suggest that geography may have played an extremely large role.

[2] Also see Krugman (1998) and Fujita, Krugman, and Venables (1999). Hanson (1996, 1998a, 1998b) and Krugman and Hanson (1994) discuss the effects that previous stages of North American integration have had on location decisions.

[3] Frankel (1998, p. 2).

Two recent gravity studies also look at the effects of NAFTA on aggregate trade between NAFTA members, both using national-level data only. Krueger (1999) found that NAFTA has had no statistically significant effect on intra-North American trade, although she did find a statistically significant decrease in imports from Europe. Gould (1998), who only considered intra-North American trade, found that NAFTA has had a significant effect on trade between the United States and Mexico, but not on trade between the United States and Canada or Mexico and Canada.[4] One reason for these lukewarm results is the small number of observations of post-NAFTA national-level trade volume. As will be apparent below, this is not a problem in the present study.

THE DATA

In constructing my empirical model, many of the choices are driven by the availability of data on North American regional trade. This study is based on a unique dataset from Statistics Canada on provincial merchandise imports and exports to and from all 50 U.S. states, the District of Columbia, and most countries of the world. It is the same dataset that formed the basis of earlier studies of the effect of the United States–Canada border on trade (McCallum, 1995, Helliwell, 1996). However, because I do not wish to consider the additional complication of the border effect, I do not include data on intraprovincial trade.

My dataset includes bilateral provincial trade between all provinces and the 50 U.S. states, the District of Columbia, Mexico, and 8 non-NAFTA countries: China, France, Germany, Hong Kong, Japan, Korea, the Netherlands, and the United Kingdom. I also include data from *World Trade Flows* on bilateral trade between the non-NAFTA countries.[5] The data on trade between non-NAFTA countries are needed as a control under the assumption that trade between them has not been affected by NAFTA.[6]

The two data shortcomings are the absence of comparable state-level data on U.S. merchandise trade with countries other than Canada and the absence of Mexican state-level data of any sort.[7] Nonetheless, the dataset is extremely rich, providing a panel of 1272 bilateral trading pairs, with 11,340 observations.[8] Note that all values in the dataset are transformed into real 1992 Canadian dollars at market exchange rates.[9] I use market exchange rates rather than purchasing-power-parity exchange rates to reflect the fact that what matters for international trade is the size of a country's economy at world prices, rather than domestic prices. Thus, in the spirit of gravity models, fluctuations in the value of a country's currency are captured by fluctuations in its economic size.

In principle, I could estimate the model with every state, province, and non-NAFTA country as its own region. However, I need to collect the states and provinces into regions to yield enough observations to provide reliable estimates of the regional NAFTA dummies. Thus, using standard regional designations from the U.S. Bureau of Economic Analysis and Statistics Canada, I divide North America into 13 regions. Three are in Canada (Eastern, Central, and Western Canada), nine are in the United States (New England, Mideast, Great Lakes, Plains, Southeast, South Central, Southwest, Rocky Mountain, Far West), and Mexico is treated as one region.[10] I also divide the 8 non-NAFTA countries into two regions: Asia and Europe. So, although my model allows for the effects of NAFTA to differ across regions, the estimated effects of NAFTA are assumed to be uniform across the locations (states, provinces, or countries) within a region.

Given the dataset, there are 39 pairs of regions. Because I have data for both directions of trade for all 39 pairs, there are 78 unidirectional trading pairs—60 for intra-NAFTA trade, 8 for imports into North America, 8 for exports from North America,

[4] For estimates of the industry-level partial-equilibrium effects, see Krueger (1999), Busse (1996), Devadoss, Kropf, and Wahl (1995), Espinosa and Noyola (1997), Hinojosa-Ojeda et al. (1996), Karemera and Ojah (1998), USITC (1997), and Wylie (1995). Also, see the volume edited by Kehoe and Kehoe (1995) for applied general-equilibrium estimates.

[5] See Feenstra (2000) and Feenstra, Lipsey, and Bowen (1997) for descriptions of this dataset.

[6] Of course, in the most general of general-equilibrium models, NAFTA also would affect trade between any two non-NAFTA countries. Nonetheless, these effects are small enough to ignore for present purposes.

[7] The United States does collect state-level export data, but it is not compatible with the Canadian data. See Coughlin and Wall (2003) for an analysis of NAFTA and U.S. state exports.

[8] Of the 1272 pairs, 1020 are for trade between U.S. states and Canadian provinces (51 states, 10 provinces, 2 directions of trade), 20 are for trade between Mexico and the Canadian provinces, 16 are for trade between Mexico and the 8 non-NAFTA countries, 160 are for trade between the Canadian provinces and the non-NAFTA countries, and 56 are for trade between the non-NAFTA countries.

[9] See the data appendix for details about data sources.

[10] See the data appendix for the assignment of states, provinces, and countries to regions.

and 2 for trade between Asia and Europe. To estimate these interregional effects, I include region-pair dummy variables for all 76 of the region pairs that include at least one North American region.

ESTIMATION

I estimate bilateral trade with a gravity equation specifying the level of exports from location i to location j as a function of their gross domestic products (GDPs), the distance between them, and any number of fixed cultural and geographic measures such as language and contiguity. Departing somewhat from the standard gravity model, I do not impose the restriction that the intercepts be the same across pairs of locations and directions of trade. This follows Mátyás (1997), Bayoumi and Eichengreen (1997), Cheng and Wall (2002), Glick and Rose (2001), Pakko and Wall (2001), and Egger (2002) who argue that gravity models that restrict the intercepts to equality suffer from heterogeneity bias.

The gravity equation I estimate is

$$\ln(1 + x_{ijt}) = \alpha_0 + \alpha_{ij} + \lambda t + \beta \ln Y_{it} + \gamma \ln Y_{jt} + \delta \ln dist_{ij} + \mu' \mathbf{EU} + \delta' \mathbf{IntraNA} + \theta' \mathbf{NAImp} + \rho' \mathbf{NAExp} + \varepsilon_{ijt}, \quad (1)$$

where x_{ijt} is real exports from location i to location j in year t, α_0 is the shared intercept, α_{ij} is the trading-pair intercept (without the restriction that $\alpha_{ij} = \alpha_{ji}$), λt is the shared time trend, $dist_{ij}$ is the distance between i and j, and Y_{it} and Y_{jt} are the real GDPs of i and j.[11]

IntraNA is a 60×1 vector of dummy variables to capture the effects of NAFTA on both directions of trade between the 30 region-to-region combinations within North America. An element of **IntraNA** takes the value of one when the observation is of post-NAFTA trade from the element's exporting region to its importing region and is zero otherwise. Similarly, **NAImp** is the 8×1 vector of dummy variables to capture the effects of NAFTA on North American regional imports from Asia and Europe, and **NAExp** is the 8×1 vector of dummy variables to capture the effects of NAFTA on North American regional exports to Europe and Asia.

Because the four European countries in the sample are also members of the European Union (EU), the regression equation also includes dummy variables to control for the transformation of the European Community (EC) into the EU in 1993. Specifically, **EU** is a vector of three dummy variables for post-EU trade: one each for trade between members, trade from a nonmember to a member, and trade from a member to a nonmember. Note that, because the model has trading-pair intercepts and because the four European countries in the dataset were all members of the EC at the start of the sample period, the EU dummy variables account only for the differences between the two regimes. The effect of the EC is already accounted for by the relevant trading-pair intercepts.

The least-squares estimates are provided in Tables 1 and 2A. Distance and other standard variables in gravity models, such as contiguity and common language, cannot be estimated separately because they are all fixed over time. Because of this, they are subsumed into the trading-pair intercept, along with all other observable and unobservable fixed factors related to history, culture, preferences, etc., that would make exports from i to j differ from trade between other trading pairs.

The results in Table 1 are as expected for a gravity equation: The higher the incomes of the two partners are, the more they trade. Of the three EU dummies, only the one for the effect of the EU on EU exports to the rest of the world is statistically significant. It suggests that the change in regime from the EC to the EU increased EU exports to nonmembers by 7.8 percent ($100 \times (e^{7.5} - 1)$). In contrast, the estimated coefficients on the other two EU dummies suggest that the EU had little effect on intra-EU trade or on EU imports from nonmembers. Keep in mind, though, that because the sample is extremely limited in its coverage of European trade, these results are far from definitive. The coefficient on the time trend is positive, indicating a common trend toward more trade, even without NAFTA, although it is statistically insignificant.

My primary interest is in the signs and levels of the estimated coefficients on the interregional NAFTA dummies, which are listed in Table 2A and converted into percentage changes in Table 2B. In addition, Tables 3 and 4 provide various aggregations of the interregional percentage changes, which are obtained by applying the estimated percentage changes to the average real post-NAFTA trade volumes for 1994-98.

[11] Note that because some observations are of zero trade, the dependent variable is the log of 1 plus exports. Having censored data normally requires Tobit estimation, but for gravity models this has typically made little difference (Eichengreen and Irwin, 1998).

Table 1

Regression Results, Dependent Variable = ln(1+Exports)

	Model with heterogeneous interregional NAFTA effects		
	Coefficient*	Robust s.e.†	*t* statistic
Shared intercept (α_0)	**–5.024**	–0.330	15.205
Log of origin GDP (β)	**0.368**	0.032	11.544
Log of destination GDP (γ)	**0.507**	0.032	15.872
Trend (λ)	0.002	0.002	1.203
EU and intra-EU trade (μ_1)	0.009	0.021	0.438
EU and EU imports (μ_2)	0.013	0.025	0.512
EU and EU exports (μ_3)	**0.075**	0.025	3.080
IntraNA, NAImp, NAExp (δ,θ,ρ)		See Table 2	
11340 observations, $\bar{R}^2 = 0.981$, F(82,9986) = 66.79			

NOTE: The 1272 bilateral region-pair intercepts are suppressed for space considerations.
*Bold indicates significance at the 5 percent level.
†White-corrected standard errors.

TRADE BETWEEN NORTH AMERICAN REGIONS

Canada—United States Trade

According to my results as summarized in Table 3A, NAFTA increased Canadian exports to the United States by 29 percent and Canadian imports from the United States by 14 percent. From the perspective of the three Canadian regions, positive NAFTA effects were far from universal. As shown in Tables 2A and B, all 18 of the estimated effects of NAFTA on trade between Eastern Canada and a U.S. region are negative, and all but one are statistically significant. In total, the results indicate that Eastern Canadian exports to the United States were 9 percent lower because of NAFTA, with the largest decreases being to the Rocky Mountain and Plains regions. Similarly, Eastern Canadian imports from the United States fell by 13 percent, with imports from all U.S. regions seeing roughly similar decreases.

In stark contrast with the results for Eastern Canada, the results for Central Canada indicated that NAFTA led to large increases in trade with the United States. They suggest that total Central Canadian exports to the United States rose by 43 percent because of NAFTA, with all but one U.S. region seeing a large increase. On the import side, NAFTA increased Central Canadian imports from the United States by 18 percent. Although the effects on imports from the Rocky Mountain region and the Far West were small and statistically insignificant, the effects on imports from the other seven regions were all positive.

The mixed region-to-region effects for Western Canada mean that the estimated effect of NAFTA on the region's total trade with the United States was effectively zero. Nonetheless, there were large differences across U.S. regions in the estimated effects of NAFTA on Western Canada's trade. The one positive and statistically significant effect was to the Great Lakes region. The four negative and statistically significant effects were for exports to the Northeast, Mideast, Southeast, and South Central regions. For Western Canadian imports from the United States, only the estimated effect on imports from the Great Lakes region was positive and statistically significant. The four regions with negative and statistically significant effects were the Northeast, South Central, Rocky Mountain, and Far West regions.

Table 3B provides the region-to-region effects aggregated across the three Canadian regions for each of the U.S. regions. From this perspective, it is easy to see that the estimated positive effect of NAFTA on trade between Canada and the United States was fairly general across U.S. regions. Exceptions to this were the Rocky Mountain region, with an estimated 6 percent fall in exports to Canada with no change in imports, and the Far West, with

Table 2A

Coefficients on Region-to-Region NAFTA Dummies

Origin/ destination	Eastern Canada	Central Canada	Western Canada	New England	Mid-east	Great Lakes	Plains	South-east	South Central	South-west	Rocky Mountain	Far West	Mexico	Europe	Asia
Eastern Canada				**–0.115**	–0.034	**–0.051**	**–0.193**	**–0.106**	**–0.157**	**–0.094**	**–0.224**	**–0.164**	**–0.159**	**–0.179**	**–0.174**
				0.022	*0.026*	*0.020*	*0.011*	*0.019*	*0.015*	*0.035*	*0.013*	*0.013*	*0.035*	*0.030*	*0.025*
Central Canada				**0.367**	**0.204**	**0.438**	**0.206**	**0.467**	**0.402**	**0.316**	–0.088	**0.275**	–0.006	**–0.137**	–0.032
				0.050	*0.034*	*0.042*	*0.033*	*0.039*	*0.042*	*0.050*	*0.090*	*0.061*	*0.074*	*0.042*	*0.054*
Western Canada				**–0.121**	**–0.075**	**0.085**	0.015	**–0.092**	**–0.077**	–0.015	0.052	0.017	**0.269**	–0.060	**–0.107**
				0.019	*0.031*	*0.040*	*0.041*	*0.017*	*0.021*	*0.030*	*0.034*	*0.022*	*0.052*	*0.032*	*0.034*
New England	**–0.129**	**0.155**	**–0.144**												
	0.010	*0.017*	*0.012*												
Mideast	**–0.116**	**0.108**	–0.012												
	0.011	*0.043*	*0.018*												
Great Lakes	**–0.139**	**0.229**	**0.110**												
	0.014	*0.040*	*0.024*												
Plains	**–0.158**	**0.082**	0.023												
	0.010	*0.034*	*0.020*												
Southeast	**–0.157**	**0.200**	–0.034												
	0.013	*0.033*	*0.018*												
South Central	**–0.137**	**0.245**	**–0.038**												
	0.011	*0.033*	*0.015*												
Southwest	**–0.161**	**0.093**	–0.007												
	0.014	*0.040*	*0.022*												
Rocky Mountain	**–0.180**	–0.022	**–0.115**												
	0.011	*0.036*	*0.015*												
Far West	**–0.157**	0.023	**–0.062**												
	0.013	*0.030*	*0.016*												
Mexico	**–0.132**	**0.416**	**0.231**											0.001	0.134
	0.056	*0.098*	*0.064*											*0.097*	*0.101*
Europe	**–0.137**	0.067	**–0.137**										–0.079		
	0.035	*0.037*	*0.045*										*0.079*		
Asia	**–0.179**	0.028	**–0.146**										0.106		
	0.030	*0.046*	*0.028*										*0.073*		

NOTE: Numbers in italics are the White-corrected standard errors. Bold indicates significance at the 5 percent level.

Table 2B

Percentage Changes in Region-to-Region Trade Due to NAFTA

Origin/ destination	Eastern Canada	Central Canada	Western Canada	New England	Mid-east	Great Lakes	Plains	South-east	South Central	South-west	Rocky Mountain	Far West	Mexico	Europe	Asia
Eastern Canada				**–10.9**	–3.3	**–5.0**	**–17.6**	**–10.1**	**–14.5**	**–9.0**	**–20.1**	**–15.1**	**–14.7**	**–16.4**	**–16.0**
Central Canada				**44.3**	**22.6**	**55.0**	**22.9**	**59.5**	**49.5**	**37.2**	–8.4	**31.7**	–0.6	**–12.8**	–3.1
Western Canada				**–11.4**	**–7.2**	**8.9**	1.5	**–8.8**	**–7.4**	–1.5	5.3	1.7	**30.9**	–5.8	**–10.1**
New England	**–12.1**	**16.8**	**–13.4**												
Mideast	**–11.0**	**11.4**	–1.2												
Great Lakes	**–13.0**	**25.7**	**11.6**												
Plains	**–14.6**	**8.5**	2.3												
Southeast	**–14.5**	**22.1**	–3.3												
South Central	**–12.8**	**27.8**	**–3.7**												
Southwest	**–14.9**	**9.7**	–0.7												
Rocky Mountain	**–16.5**	–2.2	**–10.9**												
Far West	**–14.5**	2.3	**–6.0**												
Mexico	**–12.4**	**51.6**	**26.0**											0.1	14.3
Europe	**–12.8**	6.9	**–12.8**										–7.6		
Asia	**–16.4**	2.8	**–13.6**										11.2		

NOTE: Bold indicates significance at the 5 percent level.

Table 3

Aggregated Effects of NAFTA on Intra-NAFTA Trade (Percent)

A. Canada–United States by Canadian Regions

Region	Exports to United States	Imports from United States
Eastern Canada	–8.8	–13.1
Central Canada	42.8	18.3
Western Canada	0.9	–0.5
Canada total	29.2	14.3

B. Canada–United States by U.S. Regions

Region	Exports to Canada	Imports from Canada
New England	12.8	25.9
Mideast	9.9	16.7
Great Lakes	23.9	47.1
Plains	6.2	9.3
Southeast	17.1	40.9
South Central	21.7	27.9
Southwest	6.4	18.1
Rocky Mountain	–5.8	0.6
Far West	–1.7	14.4
U.S. total	14.3	29.2

C. Canada-Mexico by Canadian Regions

Region	Exports to Mexico	Imports from Mexico
Eastern Canada	–14.7	–12.4
Central Canada	–0.6	51.6
Western Canada	30.9	26.0
Canada total	11.5	48.2

an even smaller estimated drop in exports to Canada. The results also indicate that the Great Lakes and South Central regions saw the largest increases in exports to Canada, while the largest increases in imports from Canada were for the Great Lakes and Southeast regions.

Canada-Mexico Trade

As reported in Table 3C, according to the model, NAFTA had a large effect on trade between Mexico and Canada, with significant regional variation. It suggests that, for Canada as a whole, NAFTA increased exports to Mexico by 12 percent and imports from Mexico by 48 percent. However, Eastern Canada saw its exports to and imports from Mexico drop by 15 and 12 percent, respectively, whereas Western Canada saw increases of 31 and 26 percent, respectively. For Central Canada, NAFTA had no effect on exports to Mexico, while it increased imports from Mexico by 52 percent.

Trade Creation?

As discussed in the introduction, according to the Vinerian dichotomy, NAFTA should have increased the volume of trade between its members, whether these members are countries or regions. Although my results indicate that trade creation held at the country-to-country level, this was far from universal for region-to-region or region-to-country trade. Of the 60 coefficients on intra-NAFTA region-to-region

Table 4

Aggregated Effects of NAFTA on Extra-NAFTA Trade (Percent)

A. Europe

Region	Exports to Europe	Imports from Europe
Eastern Canada	–16.4	–12.8
Central Canada	–12.8	6.9
Western Canada	–5.8	–12.8
Canada total	–11.7	1.7
Mexico	0.1	–7.6

B. Asia

Region	Exports to Asia	Imports from Asia
Eastern Canada	–16.0	–16.4
Central Canada	–3.1	2.8
Western Canada	–10.1	13.6
Canada total	–8.9	–3.0
Mexico	14.3	11.2

trade, 27 indicate statistically significant decreases in interregional trade because of NAFTA, with 21 of them associated with Eastern Canadian trade. Aggregating the region-to-region effects to the region-to-country level, negative trade effects also arise: The estimated effect on both directions of Eastern Canada's trade with the United States and Mexico are negative and large. Finally, when the regional effects are aggregated to the country-to-country level, all results have NAFTA leading to an increase in intra-NAFTA trade.

TRADE WITH THE REST OF THE WORLD

Canada

As reported in Tables 4A and B, the estimated effects of NAFTA on Canada's regional exports to Europe and Asia were, for the most part, consistent with the Vinerian prediction of trade diversion. NAFTA's estimated effects on total Canadian exports to Europe and Asia were decreases of 12 and 9 percent, respectively. Although the magnitude of these effects differed across Canadian regions, the estimated effect on exports to both Asia and Europe were negative for all regions. For both continents, Eastern Canada experienced the largest drops in exports (greater than 16 percent), whereas Western Canada had the smallest drop in exports to Europe (6 percent), and Central Canada had the smallest drop in exports to Asia (3 percent).

On the import side, the estimated effect of NAFTA on total Canadian imports from Europe was an increase of less than 2 percent, whereas its estimated effect on imports from Asia was a decrease of 3 percent. At the regional level, the results indicate that Eastern and Western Canada both had large decreases in imports from both Europe and Asia, whereas Central Canada saw small and statistically insignificant increases in imports from both continents. So, although the estimated effects of NAFTA on total Canadian imports from each of Asia and Europe were effectively zero, the real story is at the regional level. Consistent with Vinerian trade diversion, the indication is that Eastern and Western Canada both experienced large decreases in imports from Europe and Asia.

Mexico

As reported in Tables 4A and B, the estimated effects of NAFTA on Mexico's exports to the rest of the world were mixed. According to the model, exports to Europe were unaffected by NAFTA, whereas exports to Asia were 14 percent higher. As for Mexican imports, the model suggests that NAFTA led to an 8 percent drop in imports from Europe, whereas it led to an 11 percent increase in imports

from Asia. Note, though, that none of these estimated effects of NAFTA on Mexican trade with Europe and Asia is statistically significant at traditional levels.

Trade Diversion?

At the national and regional levels, the effects of NAFTA on Canada's and Mexico's trade with the non-NAFTA world strongly suggests that there has been more going on than simple trade diversion. Although most of the results for Canadian trade are consistent with trade diversion, the story is different for Mexico. In particular, the results indicate that NAFTA has increased the volume of trade with Asia, although the estimated effect on exports and imports are statistically significant at only the 18 and 15 percent levels, respectively.

SUMMARY AND CONCLUSIONS

According to my results, the effects of NAFTA on the volume and pattern of North American trade have been significant (statistically and otherwise). Specifically, the results indicate that, because of NAFTA, 29 percent more merchandise flowed from Canada to the United States and 14 percent more merchandise flowed from the United States to Canada. Thus, about one-half of the increase in Canadian exports to the United States between 1993 and 1997 is attributed to NAFTA, while about one-quarter of the increase in Canadian imports from the United States over the period is attributed to NAFTA. The results indicate also that NAFTA increased the flow of merchandise from Canada to Mexico by 12 percent and increased the flow from Mexico to Canada by 48 percent. Thus, NAFTA was responsible for about one-quarter of the increase in Canadian exports to Mexico between 1993 and 1997 and roughly 60 percent of the increase in Canadian imports from Mexico over the period.

The volume and pattern of North American trade with Europe and Asia also changed in the wake of NAFTA. Specifically, NAFTA led to large decreases in Canada's exports to Europe and Asia, a decrease in Mexican imports from Europe, and a large increase in Mexican trade with Asia.

The geographical approach reveals interesting regional differences in the effects of NAFTA. For Eastern Canada, NAFTA led to large decreases in trade with the United States, Mexico, Asia, and Europe. For Central Canada, NAFTA led to large increases in trade with the rest of North America and a large decrease in exports to Europe. For Western Canada, NAFTA had no effect on total trade with the United States, but it did lead to large increases in trade with Mexico and decreases in trade with Europe and Asia. For U.S. regions, the increases in trade were spread widely, with the Rocky Mountain and Far West regions as exceptions.

According to the Vinerian dichotomy, NAFTA should have increased trade between North American regions and decreased trade between each North American region and the rest of the world. Although the gravity methodology is not adequate for separating Vinerian effects from geographic effects, it has provided sufficient evidence that there is more to North American integration than trade creation and diversion. The most significant exceptions to the Vinerian dichotomy were as follows: (i) decreased trade between Eastern Canada and all U.S. regions and Mexico, (ii) decreased trade between Western Canada and some U.S. regions, and (iii) increased trade between Mexico and Asia.

REFERENCES

Bayoumi, Tamim and Eichengreen, Barry. "Is Regionalism Simply a Diversion? Evidence from the Evolution of the EC and EFTA," in T. Ito and A.O. Krueger, eds., *Regionalism versus Multilateral Trade Arrangements*. Chicago: University of Chicago Press, 1997.

Bergstrand, Jeffrey H. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Review of Economics and Statistics*, August 1985, *67*(3), pp. 474-81.

_____. "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade." *Review of Economics and Statistics*, February 1989, *71*(1), pp. 143-53.

Busse, Matthias. "NAFTA's Impact on the European Union." *Aussenwirtschaft*, September 1996, *51*(3), pp. 363-82.

Cheng, I-Hui and Wall, Howard J. "Controlling for Heterogeneity in Gravity Models of Trade and Integration." Working Paper 99-010C, Federal Reserve Bank of St. Louis, 2002.

Coughlin, Cletus C. and Wall, Howard J. "NAFTA and the Changing Pattern of State Exports." *Papers in Regional Science*, 2003 (forthcoming).

Deardorff, Alan V. "Testing Trade Theories and Predicting Trade Flows," in R.W. Jones and P.B. Kenen, eds., *Handbook of International Economics*. Vol. I. New York: Elsevier, 1984.

_____. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in J.A. Frankel, ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, 1998.

Devadoss, Stephen; Kropf, Jurgen and Wahl, Thomas. "Trade Creation and Diversion Effects of the North American Free Trade Agreement of U.S. Sugar Imports from Mexico." *Journal of Agricultural and Resource Economics*, December 1995, *20*(2), pp. 215-30.

Egger, Peter. "An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials." *World Economy*, February 2002, *25*(2), pp. 297-312.

Eichengreen, Barry and Irwin, Douglas A. "The Role of History in Bilateral Trade Flows," in J.A. Frankel, ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, 1998.

Espinosa, J. Enrique and Noyola, Pedro. "Emerging Patterns in Mexico-U.S. Trade," in Barry Bosworth, Susan M. Collins, and Nora C. Lustig, eds., *Coming Together? Mexico-U.S. Relations*. Brookings Institution Press, 1997.

Feenstra, Robert C. "World Trade Flows, 1980-1997." Working paper, Center for International Data at University of California–Davis, 2000.

_____; Lipsey, Robert E. and Bowen, Harry P. "World Trade Flows, 1970-1992, with Production and Tariff Data." NBER Working Paper No. w5910, National Bureau of Economic Research, 1997.

Frankel, Jeffrey A. "Introduction," in J.A. Frankel, ed., *The Regionalization of the World Economy*. Chicago: University of Chicago Press, 1998.

Fujita, Masahisa; Krugman, Paul and Venables, Anthony J. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press, 1999.

Glick, Reuven and Rose, Andrew. "Does a Currency Affect Trade? The Time Series Evidence." NBER Working Paper No. w8396, National Bureau of Economic Research, 2001.

Gould, David M. "Has NAFTA Changed North American Trade?" Federal Reserve Bank of Dallas *Economic Review*, First Quarter 1998, *0*(0), pp. 12-23.

Hanson, Gordon H. "Economic Integration, Intraindustry Trade, and Frontier Regions." *European Economic Review*, April 1996, *40*(3-5), pp. 941-49.

_____. "North American Economic Integration and Industry Location." *Oxford Review of Economic Policy*, Summer 1998a, *14*(2), pp. 30-44.

_____. "Regional Adjustment to Trade Liberalization." *Regional Science and Urban Economics*, July 1998b, *28*(4), pp. 419-44.

Helliwell, John F. "Do National Borders Matter for Quebec's Trade?" *Canadian Journal of Economics*, August 1996, *29*(3), pp. 507-22.

Hinojosa-Ojeda, Raul; Dowds, Curt; McCleery, Robert; Robinson, Sherman; Runsten, David; Wolff, Craig and Wolff, Goetz. "North American Integration Three Years after NAFTA." North American Integration and Development Center, UCLA, December 1996.

Karemera, David and Ojah, Kalu, "An Industrial Analysis of Trade Creation and Diversion Effects of NAFTA." *Journal of Economic Integration*, September 1998, *13*(3), pp. 400-25.

Kehoe, Patrick J. and Kehoe, Timothy J., eds. *Modeling North American Economic Integration: Advanced Studies in Theoretical and Applied Econometrics*. Vol. 31. Boston: Kluwer, 1995.

Krueger, Anne O. "Trade Creation and Trade Diversion Under NAFTA." NBER Working Paper No. w7429, National Bureau of Economic Research, December 1999.

_____. "NAFTA's Effects: A Preliminary Assessment." *The World Economy*, June 2000, *23*(6), pp. 761-75.

Krugman, Paul. *Geography and Trade*. Cambridge, MA: MIT Press, 1991a.

_____. "Increasing Returns and Economic Geography." *Journal of Political Economy*, June 1991b, *99*(3), pp. 483-99.

_____. "What's New about the New Economic Geography?" *Oxford Review of Economic Policy*, Summer 1998, *14*(2), pp. 7-17.

_____ and Hanson, Gordon. "Mexico-U.S. Free Trade and the Location of Production," in Peter M. Garber, ed., *The Mexico-U.S. Free Trade Agreement*. Cambridge, MA: MIT Press, 1994, pp. 163-86.

Mátyás, László. "Proper Econometric Specification of the Gravity Model." *The World Economy*, May 1997, *20*(3), pp. 363-68.

McCallum, John. "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review*, June 1995, *85*(3), pp. 615-23.

Pakko, Michael R. and Wall, Howard J. "Reconsidering the Trade-Creating Effects of a Currency Union." Federal Reserve Bank of St. Louis *Review*, September/October 2001, *83*(5), pp. 37-45.

United States International Trade Commission (USITC). "The Impact of the North American Free Trade Agreement on the U.S. Economy and Industries: A Three Year Review." Investigation No. 332-381, 1997.

Viner, Jacob. *The Customs Union Issue*. New York: Carnegie Endowment for International Peace, 1950.

Wylie, Peter J. "Partial Equilibrium Estimates of Manufacturing Trade Creation and Diversion Due to NAFTA." *North American Journal of Economics and Finance*, Spring 1995, *6*(1), pp. 65-84.

Data Appendix

DATA SOURCES

Province-to-state and province-to-country trade data, 1990-98, from Statistics Canada.
Non-Canadian country-to-country trade, 1990-97, from *World Trade Flows, 1980-1997*.
Nominal gross provincial product, 1990-98, from Statistics Canada.
Nominal gross state product, 1990-98, from the Bureau of Economic Analysis.
Nominal gross domestic product, 1990-98, from the World Bank's *World Development Indicators, 1999*.

All variables were converted into real Canadian dollars using the Canadian consumer price index (CPI) and $/C$ market exchange rates from Statistics Canada.

THE COMPOSITION OF THE REGIONS

The nine U.S. regions are based on the eight Bureau of Economic Analysis (BEA) regions, with the BEA Southeast region split into two: Southeast and South Central. The three Canadian regions are according to Statistics Canada. The eight countries assigned to the Asia and Europe regions are taken from Canada's ten most important trading partners, the other two being the United States and Taiwan. Taiwan could not be included because the World Bank does not provide its GDP data.

Eastern Canada:	New Brunswick, Newfoundland, Nova Scotia, and Prince Edward Island
Central Canada:	Ontario and Quebec
Western Canada:	Alberta, British Columbia, Manitoba, and Saskatchewan
New England:	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont
Mideast:	Delaware, District of Columbia, Maryland, New Jersey, New York, and Pennsylvania
Great Lakes:	Illinois, Indiana, Michigan, Ohio, and Wisconsin
Plains:	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota
Southeast:	Florida, Georgia, North Carolina, South Carolina, Virginia, and West Virginia
South Central:	Alabama, Arkansas, Kentucky, Louisiana, Mississippi, and Tennessee
Southwest:	Arizona, New Mexico, Oklahoma, and Texas
Rocky Mountain:	Colorado, Idaho, Montana, Utah, and Wyoming
Far West:	Alaska, California, Hawaii, Nevada, Oregon, and Washington
Mexico:	Mexico
Asia:	China, Hong Kong, Japan, and Korea
Europe:	France, Germany, the Netherlands, and the United Kingdom

Appendix A

Table A1

Percentage Changes in Real Region-to-Region International Trade, 1993-97

Origin/ destination	Eastern Canada	Central Canada	Western Canada	New England	Mid-east	Great Lakes	Plains	South-east	South Central	South-west	Rocky Mountain	Far West	Mexico	Europe	Asia
Eastern Canada				37.7	62.0	71.0	33.3	130.8	12.0	104.2	–4.8	3.6	0.9	11.5	38.0
Central Canada				69.2	41.0	48.6	79.6	86.4	80.2	110.5	160.3	101.2	21.7	9.7	69.4
Western Canada				81.2	36.6	48.7	59.7	53.3	71.0	125.8	82.9	42.6	90.7	35.3	27.6
New England	40.7	45.8	50.8												
Mideast	14.3	50.7	48.3												
Great Lakes	–15.7	41.6	33.2												
Plains	–47.9	84.0	60.5												
Southeast	–4.2	63.8	63.3												
South Central	78.7	87.6	62.8												
Southwest	39.4	95.9	87.2												
Rocky Mountain	–39.9	46.2	39.1												
Far West	22.8	53.3	51.2												
Mexico	26.0	69.2	219.1											58.0	70.0
Europe	45.7	49.4	109.5										22.3		
Asia	23.0	28.9	20.2										15.1		

NOTE: Values are the percentage differences in trade between 1997 and 1993, measured in 1992 Canadian dollars at market exchange rates.

Table A2

Canada–United States by Canadian Regions

Region	Exports to United States	Imports from United States
Eastern Canada	56.7	11.4
Central Canada	58.4	53.7
Western Canada	55.0	51.9
Canada total	57.6	52.7

Table A3

Canada–United States by U.S. Regions

Region	Exports to Canada	Imports from Canada
New England	45.9	63.0
Mideast	49.9	41.3
Great Lakes	40.3	48.8
Plains	74.0	67.2
Southeast	61.2	83.5
South Central	83.8	75.4
Southwest	92.5	114.0
Rocky Mountain	43.2	99.0
Far West	52.0	66.9
U.S. total	52.7	57.6

Table A4

Canada–Mexico, Europe, and Asia

	Canadian Exports	Canadian Imports
Mexico	45.6	77.7
Europe	14.6	55.7
Asia	36.4	25.9

On the Pervasive Effects of Federal Reserve Settlement Regulations

Ken B. Cyree, Mark D. Griffiths, and Drew B. Winters

The purpose of this paper is to determine whether Federal Reserve settlement effects have also appeared in the overnight London interbank offer rate (LIBOR) since the Federal Reserve removed the reserve requirements on Eurocurrency liabilities.[1] We begin by explaining why this is an important issue and by describing the characteristics of these settlement effects.

The primary reason to examine this change in reserve requirements is to determine the reach of U.S. Federal Reserve regulatory changes. Markets are becoming increasingly global and, accordingly, large bank operations are becoming increasingly global. Finding a settlement effect in LIBOR would suggest that banks receiving U.S. deposits actively use the London interbank loan market to manage their domestic reserve accounts. A regular pattern or effect in LIBOR created by a Federal Reserve rule change would show that the impact of Fed regulations is not limited to the national boundaries of the United States. Finding a settlement effect in LIBOR also would show that the mechanics of one market can spill over into other markets thought to be independent.

Since implementation of the Monetary Control Act in 1980, most depository institutions in the United States have been subject to the Federal Reserve's statutory reserve requirements. These requirements establish the percentage of each liability category for which a bank must maintain reserves, either as vault cash or as deposits in the bank's Federal Reserve account.[2] (See the boxed insert: "The Federal Reserve's Statutory Reserve Requirements.") A settlement effect, referred to above, is a regular pattern of interest rate changes associated with the days that banks must settle their reserve accounts with the Fed. Under rules in place since February 1984, the primary reserve management process calls for a biweekly settlement of reserve accounts.[3] This two-week cycle is known as the reserve maintenance period, which begins on a Thursday and ends two weeks later on Wednesday. The last day of each reserve maintenance period, called settlement Wednesday, is when banks must settle their reserve account with the Fed. The settlement rules require that, on settlement Wednesday, a bank's total actual reserves over the two-week period equal or exceed its total *required* reserves for that two-week period. Successfully so doing is referred to as *settling* with the Fed. Banks manage their actual reserves by trading deposits at Federal

[1] The British Bankers Association (BBA) publishes daily reference rates at various short maturities based on a survey of the major London banks. These survey rates are referred to as LIBOR. The BBA survey rates serve as commonly accepted benchmark rates and were crucial to the development of the LIBOR and Eurodollar futures markets. In fact, the BBA is described as "fixing" the benchmark rate when it provides its daily LIBOR data. The BBA has been fixing LIBOR reference rates at various short maturities since the late 1980s but did not begin providing an overnight LIBOR reference rate until 2001. Our data are not BBA reference rates. Our data are the closing overnight dollar-based cash market rates in the London interbank market, and our data source appropriately refers to this rate series as overnight LIBOR, which is how we refer to it in this paper. However, we remind the reader that our data are not the rates fixed by the BBA for reference rates.

Ken B. Cyree is an assistant professor of finance at Texas Tech University. Mark D. Griffiths is an associate professor of finance at the American School of International Management. Drew B. Winters is an associate professor of finance at the University of Central Florida, who conducted a portion of this work while visiting the Federal Reserve Bank of St. Louis. The authors thank Bruce Frost of the Chicago Mercantile Exchange for assistance in obtaining data, Harold Rose of the London Business School and Rohan Christie-David for information about the British banking system, and Tom Lindley for his comments.

[2] Regulation D states that the reserves that depository institutions are required to maintain are to facilitate the implementation of monetary policy by the Federal Reserve System. However, Goodfriend and Hargraves (1983) discuss that banks have been required to hold reserves since 1863 and the rationale for the reserves has changed over time. Reserves initially were required for liquidity, and this rationale was maintained until 1931. In 1931, the liquidity rationale was replaced by the idea that required reserves play a role in the execution of the Fed's credit policies. In the 1950s, Fed policy statements began shifting toward money stock issues; in the late 1970s, M1 became the primary intermediate policy target. The Monetary Control Act of 1980 imposed universal reserve requirements based on the argument by the Fed that its ability to control monetary aggregates was being weakened by deposits moving outside the Fed's jurisdiction; this process officially brought us to the current rationale that required reserves are for implementing monetary policy.

[3] Many small banks settle their reserve position each week based on a reserve requirement amount that is set once each quarter. However, the vast majority of reservable deposits are held in banks that are subject to biweekly settlement. Accordingly, studies of market pressures created by reserve account management, like this study, focus on the biweekly settlement process.

Table 1

Spreads Over 3-Month T-bills

Time period	Federal funds (basis points) spread relative to 3-month T-bills	Overnight government repo (basis points) spread relative to 3-month T-bills	LIBOR spread (basis points) relative to 3-month T-bills
8/4/86 through 1/31/95	45 (43)	47 (43)	55 (51)
2/7/91 through 1/31/95	6 (4)	16 (14)	15 (14)

NOTE: Numbers provided here are mean spreads; numbers in parentheses are median spreads.

Reserve Banks among themselves in the federal funds market. This trading over the two-week period may create pressure in the federal funds market and cause spikes in interest rate changes and volatility on settlement Wednesdays, which are the settlement effects.

At first glance, it seems unlikely that a change in reserve requirements on Eurocurrency liabilities would create settlement effects in LIBOR; after all, settlement effects are the result of banks reconciling their reserves requirements for their domestic bank deposits. Banks generate funds, generally in the form of deposits, and the reserve requirements mandate that a percentage of these funds be held in reserve. However, not all of a bank's funding sources are in the form of traditional deposits and the regulations exempt certain sources of funds from reserve requirements. The most common source of exempt funds for managing reserve accounts are funds purchased through the federal funds market. Banks are allowed to bring in funds through the federal funds market to increase their actual reserves, and the exemption from reserve requirements on federal funds allows this to be done without an accompanying increase in their required reserves. That is, the reserve requirement on the liability created by the purchase of federal funds is 0 percent. In contrast, from 1980 until the end of 1990, Eurocurrency liabilities had a reserve requirement of 3 percent[4]; by 1991, that requirement had become 0 percent. This eventual exemption from reserve requirements for Eurocurrency liabilities is what has raised the possibility of a settlement effect in LIBOR: Without the 3 percent reserve requirement, reserve deposits that are borrowed in the Eurocurrency markets for settlement purposes are now essentially equivalent to deposits borrowed in the federal funds market. Accordingly, then, the law of one price should apply and drive the rates in these markets together.[5] (We show this in Table 1.) However, for settlement effects to appear in LIBOR, banks subject to U.S. Federal Reserve settlement regulations must be using Eurocurrency liabilities to manage their reserve accounts actively enough to affect LIBOR on settlement Wednesdays. The federal funds trading desk manager of a major U.S. bank confirms that banks began managing their reserve accounts with Eurocurrency liabilities after the reduction in their reserve requirement. So, if banks are now using these liabilities to manage their reserve accounts, our question is: Are settlement pressures pervasive enough to reach overseas to create settlement effects in LIBOR?

We use closing overnight dollar-based LIBOR to test for a settlement Wednesday effect in the Eurodollar market because LIBOR is the rate that major London banks offer for Eurocurrency liabilities to other banks. For the pre-1991 period, when Eurocurrency liabilities were subject to a 3 percent reserve requirement, we do not find any settlement effects in overnight LIBOR. We do find settlement effects concurrent with the change to the 0 percent

[4] An amendment to Regulation D changed the reserve requirement on Eurocurrency liabilities: From August 1980 through the reserve maintenance period ending December 12, 1990, the requirement was 3 percent; for the one reserve maintenance period from December 13 through December 26, 1990, the requirement was 1.5 percent; as of December 27, 1990, the requirement was changed to its current 0 percent.

[5] We expect minor risk differences between the Eurocurrency markets and the federal funds market, so the law of one price will not hold exactly. For example, even with a 0 percent reserve requirement, Eurocurrency liabilities are a reservable liability of the bank and hence may affect the marginal reserve requirement on other liabilities as well as the required frequency of deposit reporting and reserve settlement. However, without a reserve requirement on Euro-market U.S. dollars, the rates in the two markets should be close enough to allow Euro-market U.S. dollars to be a viable source of funds for bank settlement.

reserve requirement. In related work, Griffiths and Winters (1997) found settlement effects in closing federal funds rates and closing overnight government repurchase agreements (repo) rates. Thus, we test closing federal funds rates and closing overnight government repo rates for settlement effects during both periods—when Eurocurrency liabilities carried a 3 percent reserve requirement and after the reserve requirement was reduced to zero. We find settlement effects in federal funds rates during both periods, which suggests that a settlement effect in LIBOR only when the reserve requirement is 0 percent is a direct result of the reduction in that reserve requirement. Our results suggest the following: (i) Federal Reserve policies are sufficiently pervasive to have global effects and (ii) the effects of the federal funds market microstructure for U.S. depository institutions spill over into other markets.

In the next section, we discuss the existing theoretical and empirical literature on the rate change and variance patterns unique to reserve account settlement with the U.S. Federal Reserve, as well as the relevant institutional details for U.S. reserve account management related to federal funds trading and the relationship between LIBOR and British bank settlement. We then present the data and methods and finally our test results.

INSTITUTIONAL DETAILS

In this section, we provide various institutional details as background on the different markets. We begin with a brief history of the Eurodollar market and discuss the bank settlement procedures in the United States and the United Kingdom. We also discuss the existing theoretical and empirical literature on the rate change and variance patterns unique to the reserve account settlement process with the Federal Reserve.

A Brief History of the Eurodollar Market

We begin this section with a brief history of the Eurodollar market, which may be unfamiliar to some readers. We draw our discussion from Stigum (1990, Chap. 7, pp. 207-11).

Prior to World War II, it was not uncommon for banks outside the United States to take dollar deposits, but there was little volume in this market and the market had little economic significance. During the 1950s, things began to change as the cold war between the United States and communist countries intensified. Specifically, Soviet businesses needed U.S. dollars for trade but were concerned about holding their dollar deposits in U.S. banks; so, they moved their dollar deposits to banks outside of the United States. This scenario contributed to the birth of the modern Eurodollar market.

Historically, the British pound sterling was the leading currency for international trade. However, following World War II the British ran large balance of payments deficits, so a constant threat of devaluation of the pound sterling existed. In addition, the British restricted the use of the pound sterling in financing international trade, so international trade moved toward the U.S. dollar.

As the Eurodollar market began to grow, U.S. banks were reluctant participants. In fact, Stigum describes their entry into the market as "defensive." However, the interest rate restrictions under Regulation Q forced U.S. banks to play in the Eurodollar market when depositors could get better rates outside the United States. Also, during the 1960s the United States tried to improve its balance of payments deficits by imposing capital constraints that limited the flow of dollars from U.S. banks to foreign borrowers, which created demand for Eurodollar loans.

Stigum notes that the above factors were significant contributors to the growth of the Eurodollar market, but that dollar depositors both in and out of the United States have the ability to place their deposits both in and out of the United States; so, where the dollar deposits go depends on the relative attractiveness of the deposit. Currently, the relative attractiveness of Eurodollar deposits is that they are, in particular, free of Federal Reserve statutory reserve requirements.

Federal Reserve Bank Settlement and the Federal Funds Market Literature

Since 1980, most depository institutions in the United States have been subject to the Federal Reserve's statutory reserve requirements. To enforce these requirements, since February 1984, the Fed has compared each bank's actual and required reserves during a 14-day reserve maintenance period. (For details, see the boxed insert.) Reserve maintenance periods begin every other Thursday and end on Wednesday 14 days later ("settlement Wednesday"). A maintenance period typically has 10 trading days.

A bank satisfies its statutory reserve requirement by holding an adequate amount of eligible vault cash and/or deposits at Federal Reserve Banks; no

THE FEDERAL RESERVE'S STATUTORY RESERVE REQUIREMENTS

Statutory Reserve Requirements Since 1980

The Monetary Control Act of 1980 governs statutory reserve requirements in the United States. The act, implemented in November 1980 through the Federal Reserve's Regulation D (12 CFR 204), imposes federal statutory reserve requirements on all U.S. chartered federally insured depositories (including their Edge and agreement corporation subsidiaries) and on branches and agencies of foreign banks if the parent firm's consolidated worldwide assets exceed $1 billion or if the branch or agency is eligible to apply for federal deposit insurance.[1] Within broad limits, the act delegates the setting of specific reserve-requirement ratios to the Board of Governors of the Federal Reserve System.[2]

The act imposes statutory reserve requirements on three classes of depository institutions' liabilities: net transaction deposits, nonpersonal time (including savings) deposits, and net Eurocurrency liabilities. If any of the net amounts are negative, the amount for reserve-requirement purposes is zero.

- **Net transaction deposits**: A transaction account is defined as "a deposit or account on which a depositor or account holder is permitted to make withdrawals by negotiable or transferable instrument, payment orders of withdrawal, telephone transfers, or similar devices for the purpose of making payments or transfers to third persons or others." A depository's net transaction deposits were defined, if positive, to be total transaction deposits minus the sum of cash items in process of collection and balances held at other depository institutions that could be immediately withdrawn.
- **Nonpersonal time deposits**: Time deposits (including savings deposits, which are regarded as time deposits without a specific maturity date) are defined, generally speaking, to be all deposits that are not transaction deposits. A nonpersonal deposit is a deposit in which the beneficial interest is held by other than a natural person; a natural person is an individual or sole proprietorship.
- **Eurocurrency liabilities**: Eurocurrency liabilities are the sum of net borrowing by domestic banking offices from foreign offices plus certain assets sold by domestic banking offices to foreign offices. (For reserve requirement purposes, the sale of the asset to a foreign office is treated as a dollar-for-dollar reduction in the reservable deposits of the domestic office.) Specifically, Regulation D specifies that Eurocurrency liabilities include the following:
  a. For a depository institution or an Edge or agreement corporation organized under the laws of the United States, the sum of the following transactions of U.S. offices with related offices outside the United States[3]:
     - Net balances due to a depository's non-U.S. offices and its international banking facilities (IBF) from its U.S. offices;
     - Assets (including participations) acquired from its U.S. offices and held by its non-U.S. offices, by its IBF, or by non-U.S. offices of an affiliated Edge or agreement corporation; and, for Edge and agreement corporations, assets acquired from its U.S. offices and held by non-U.S. offices of its U.S. or foreign parent institution, its IBF, or by non-U.S. offices of an affiliated Edge or agreement corporation; and
     - Credit outstanding from a depository institution's non-U.S. offices to U.S. residents (other than assets acquired and net balances due from its U.S. offices), except credit extended (i) from

Continued on p. 31

[1] This text is a general summary. Various provisions and applications of Regulation D have changed through time. At the time of this writing, the current version of Regulation D was available at < http://www.federalreserve.gov/regulations/title12/sec204/12cfr204_01.htm > . Specific legal definitions are contained in Regulation D and its staff interpretations. The citation 12 CFR 204 refers to section 204 of Chap. 12 of the Code of Federal Regulations.

[2] Limits in the legislation include a reserve requirement range of 0 to 9 percent on time deposits (including savings deposits) and 8 to 14 percent on net transaction deposits.

[3] For definitions and discussion of Edge corporations, agreement corporations, and international banking facilities, see Marcia Stigum's *The Money Market* (1990, 3rd ed., Chaps. 6 and 7).

Continued from p. 30

its non-U.S. offices in the aggregate amount of $100,000 or less to any U.S. resident; (ii) by a non-U.S. office that at no time during the computation period had credit outstanding to U.S. residents exceeding $1 million; (iii) to an international banking facility; or (iv) to an institution that will be maintaining reserves on such credit pursuant. Credit extended from non-U.S. offices or from an IBF to a foreign branch, office, subsidiary, affiliate of other foreign establishment (*foreign affiliate*) controlled by one or more domestic corporations is not regarded as credit extended to a U.S. resident if the proceeds will be used to finance the operations outside the United States of the borrower or of other foreign affiliates of the controlling domestic corporation(s).

b. For a U.S. branch or agency of a foreign bank, the sum of the following:
   - Net balances due to its foreign bank and its IBF after deducting an amount equal to 8 percent of the following: the U.S. branch or agency's total assets less the sum of (i) cash items in process of collection; (ii) unposted debits; (iii) demand balances due from depository institutions organized under the laws of the United States and from other foreign banks; (iv) balances due from foreign central banks; and (v) positive net balances due from its IBF, its foreign bank, and the foreign bank's U. S. and non-U. S. offices; and,
   - Assets (including participations) acquired from the U.S. branch or agency (other than assets required to be sold by federal or state supervisory authorities) and held by its foreign bank (including offices thereof located outside the United States), by its parent holding company, by non-U.S. offices or an IBF of an affiliated Edge or agreement corporation, or by its IBFs.

Reserve Settlement with the Federal Reserve

The accounting rules that govern a depository institution's reserve settlement with the Federal Reserve depend on the specific circumstances of the bank.[4] Generally, settlement rules differ across banks with respect to (i) whether during the previous reserve maintenance period the bank had a deficiency (actual reserves less than required) or surplus (actual reserves more than required) and (ii) whether the bank had a clearing balance contract with the Federal Reserve.[5]

Prior to 1990, relatively few banks had clearing balance contracts with the Federal Reserve and, hence, the settlement rules applicable to most banks were those for depository institutions without clearing balance contracts. Within those rules, one of the more important is that a bank may carry a deficiency or surplus forward only once, to the next maintenance period. If a deficiency is not fully offset by holding additional reserves during the next period, the bank may be subject to a monetary penalty (charged at the discount rate in effect as of the beginning of that month plus 2

Continued on p. 32

[4] The Federal Reserve's settlement rules use the concepts of a "reserve maintenance period" and a "reserve computation period." The reserve maintenance period is the interval (14 days in duration as of February 1984 and 7 days in duration prior to February 1984) during which the institution must hold enough deposits at the Federal Reserve to satisfy its reserve requirement after subtracting from its requirement the amount of its vault cash that is eligible to satisfy the requirement. The amount of a depository institution's reserve requirement is calculated from its liabilities during the reserve computation period. For more specific definitions and examples, see the Federal Reserve's *Reserve Maintenance Manual* at < http://www.frbservices.org/Accounting/CustomerReferenceGuide.cfm > .

[5] Clearing balance contracts are discussed in chapters 8 and 11 of the *Reserve Maintenance Manual*. See also E.J. Stevens, "Required Clearing Balances," Federal Reserve Bank of Cleveland *Economic Review*, December 1993, p. 2-14, and J.N. Feinman, "Reserve Requirements: History, Current Practice, and Potential Reform," *Federal Reserve Bulletin*, June 1993, pp. 569-89. Note that the older term "required clearing balance" has been replaced by the term "clearing balance requirement" in more recent Federal Reserve publications. This seems fully appropriate because, unless a bank has a history of excessive overnight and/or daylight overdrafts at the Federal Reserve, clearing balance requirements are a voluntary commitment by a bank to maintain deposits at the Fed above and beyond those required to satisfy mandatory statutory reserve requirements. In exchange for maintaining the additional deposits, the bank receives earnings credits that can be used to defray the cost of financial services (such as check clearing) purchased from the Fed. Prior to December 1990, few larger banks had clearing balance contracts; this changed sharply after the December 1990 reduction in reserve requirements. (See, for example, Feinman, Figure 9, p. 583.)

Continued from p. 31
percentage points, at an annual rate). Conversely, if a surplus is not fully utilized to satisfy requirements during that next period, it is lost. This rule does not prohibit a bank from carrying forward a deficiency or surplus from one maintenance period to the next—but it does prohibit carrying forward the *same* deficiency or excess.[6] The rules limit the maximum amount (deficiency or surplus) that can be carried forward from a reserve maintenance period into the next to the greater of 4 percent of the bank's required reserves or $50,000.[7] Banks whose actual reserves repeatedly fall short of required reserves may receive, in addition to monetary penalties, admonitions or "counseling" from the Fed.

Beginning in 1991, the somewhat different settlement rules that apply to depository institutions with clearing balance contracts are important. Following the December 1990 and April 1992 reductions in reserve requirements, many larger banks entered, for the first time, into clearing balance contracts with the Fed.[8] Between December 1990 and December 1992, for example, the aggregate amount of clearing balance contracts increased from $2 billion to $6 billion. Settlement rules for such banks are more complex than those for banks without clearing balance contracts because a depository institution might incur a deficiency or surplus with respect to its clearing balance requirement, its statutory reserve requirement, or both. Settlement is somewhat less onerous, however, because the clearing balance requirement provides a cushion for the bank with respect to satisfying its statutory reserve requirement.[9] Under the Federal Reserve's accounting rules, settlement begins, first, by subtracting eligible vault cash from the depository's statutory reserve requirement.[10] Next, the remaining portion of the statutory requirement is subtracted from the amount of deposits held at the Fed. Finally, the remaining amount of deposits held at the Fed is compared with the clearing balance requirement. Because of the sequencing of these operations, banks with clearing balance requirements are highly unlikely to be deficient with respect to their statutory reserve requirement. Further, the clearing balance requirement is said to be satisfied if the bank is within $25,000 or 2 percent (above or below) of the required amount.

As noted before, a deficiency or surplus can be carried forward into the next maintenance period but the same deficiency or surplus cannot be carried forward to a subsequent period. The maximum deficiency or surplus that may be carried forward is equal to the greater of 4 percent of the sum of the bank's statutory reserve requirement plus its clearing balance requirement or $50,000, minus the clearing balance allowance (the greater of $25,000 or 2 percent of the clearing balance requirement).[11]

—Richard G. Anderson

[6] For specific examples on deficiencies, see *Reserve Maintenance Manual*, Table 2, examples "D" and "F," p. XI-5; on surpluses, see *Reserve Maintenance Manual*, Table 1, examples "D" and "E," p. XI-2.

[7] Prior to September 1992, the carryover was the greater of (i) 2 percent of required reserves plus the clearing balance requirement or (ii) $25,000.

[8] In December 1990, the reserve requirement ratios on nonpersonal time deposits and Eurocurrency liabilities were reduced to zero. In April 1992, the marginal reserve requirement ratio on net transaction deposits was reduced to 10 percent from 12 percent. For a detailed discussion and examples of the reserve settlement rules applicable to banks with clearing balance requirements, see *Reserve Maintenance Manual*, chapter XII.

[9] These aspects are compared by Feinman (1993).

[10] The eligibility of vault cash has changed through time. Prior to 1917, all vault cash held during the reserve maintenance period was eligible. From 1917 to 1959, no vault cash was eligible. Between December 1959 and December 1960, vault cash eligibility was phased in on a pro rata monthly scale. Beginning in September 1968, eligible vault cash was the amount held during the 7-day period that ended 14 days prior to the end of a 7-day reserve maintenance period. Beginning February 1984, eligible vault cash was the amount held during the 14-day period ending 31 days prior to the end of a 14-day maintenance period. Beginning November 1992, eligible vault cash was the amount held during the 14-day period ending 17 days prior to the end of a 14-day reserve maintenance period. Since July 1998, eligible vault cash has been the amount held during the 14-day period ending 45 days before the end of the reserve maintenance period.

[11] For details, see chapter IX of the *Reserve Maintenance Manual*.

other assets may be used. The eligibility of vault cash has varied through time. During the first part of our sample period (prior to September 1992), eligible vault cash was the average amount held by the bank during a 14-day period ending 31 days before the end of the maintenance period. During the latter part of our sample (beginning September 1992), it was the average amount held during a 14-day period ending 17 days before the end of the maintenance period. At the close of the maintenance period, eligible vault cash is subtracted from the bank's required reserves. The remainder is subtracted from the average daily amount of deposits held by the bank at the Federal Reserve. If the result is negative, the bank is deficient. If positive, the bank has a surplus. Prior to September 1992, a bank could carry over into the next maintenance period, without penalty, a deficiency or surplus equal to the greater of (i) 2 percent of the sum of its required reserves plus its clearing balance requirement or (ii) $25,000. (Again, see the boxed insert for details.) In September 1992, this was increased to the greater of 4 percent or $50,000.[6] Federal Reserve rules require that a penalty be assessed if a deficiency is not offset by reserve holdings during the subsequent maintenance period. Although the rules prohibit carrying forward the *same* deficiency into a subsequent (third) period, a bank may carry forward a new deficiency. That is, a bank's reserves during the current maintenance period may be sufficient to fully satisfy a previous period's deficiency that has been carried forward but, at the same time, inadequate to avoid carrying forward a new deficiency based on its required reserves for the current maintenance period. A surplus carried forward, but not used to satisfy required reserves, expires unused.

Theoretical and empirical studies have described the unique rate change and variance patterns created by the settlement rules. Table 2 provides a cross-reference between the theoretical predictions and the empirical results from the settlement rules.

Griffiths and Winters (1995) provide a model of federal funds rate pressures based on the Federal Reserve settlement rules. Their model provides daily rate pressure predictions across the two-week reserve maintenance period. The daily predictions (Table 2, column 1 of panel A) are as follows:

- rates are expected to decline on Fridays in advance of the weekend,
- rates are expected to decline on the second Tuesday (the day before settlement), and
- rates are expected to rise on the second (settlement) Wednesday.

The predicted rate pressures should create additional daily empirical rate changes when the predicted daily pressures abate. We present the predicted empirical pattern in column 2 of panel A in Table 2. The additional rate changes not described in Griffiths and Winters are the rebound effects that follow from the abatement of the rate pressures predicted in their model. Specifically, one would expect to find the following rebound effects:

- rates are expected to rise on Mondays following the abatement of lending pressure on Fridays, and
- rates are expected to decline on the first Thursday (the first day of the reserve maintenance period) following the abatement of the borrowing pressure on settlement Wednesday of the previous reserve maintenance period.

The empirical literature shows strong support for the rate pressures predicted by Griffiths and Winters (1995). Specifically, declining rates on Fridays and rising rates on settlement Wednesday appear in all five papers presented in panel A. Also, three of the five papers show declining rates on the second Tuesday. In addition, all five papers show the expected rebound effect on the first Monday and four of five papers show the expected rebound on the second Monday. There is not a consistent rebound effect on the first Thursday.

Griffiths and Winters (1995) do not predict specific rate pressures on the first Tuesday, the first Wednesday, or the second Thursday. However, they do identify a general preference for selling over purchasing federal funds across a reserve maintenance period, which suggests declining rates in the absence of any specific rate pressures. The five papers cited show a tendency for rates to decline on the first Tuesday and the first Wednesday and for no rate change on the second Thursday.

[6] The purpose of the increase in carryover to 4 percent from 2 percent was to make successful settlement easier for banks. Griffiths and Winters (2000) examine the size of the settlement effects in closing federal funds and overnight government repo rates around this regulation change. They find no reduction in the identified settlement effects after the increase in the allowable range.

With strong empirical support for the rate change pattern predicted by Griffiths and Winters (1995) the question becomes: What pattern of rate changes is necessary to identify a settlement effect in a substitute funding source for reserve account management? Griffiths and Winters (1997) suggest that the rate change that is unique to the U.S. settlement process is the rate increase on settlement Wednesdays. They argue that any entity with non-earning cash on Fridays (not just banks with deposits at the Federal Reserve) will have an incentive to lend (invest) on Fridays to avoid leaving the funds idle over the weekend. Thus, rates across money market instruments (domestic and foreign) should decline as non-earning cash is moved into investment vehicles for the weekend.[7] Then, following the weekend, rates should rebound on Monday. Gibbons and Hess (1981) find negative returns on T-bills on Mondays and positive returns on T-bills on Wednesdays, reflecting a T-bill yield increase on Mondays and decrease on Wednesdays. Thus, the rate increases that occur on Mondays, cited in Table 2, panel A, are not unique to the U.S. settlement process. In addition, the general tendency for rates to decline on the first Wednesday of the reserve maintenance period is not unique to the U.S. settlement process. Accordingly, the only consistent daily rate change shown in the cited papers that is unique to the U.S. settlement process is the rate increase on settlement Wednesday.

Up to this point, we have described the daily rate changes created by reserve account management for successful settlement. However, rate changes are only part of the picture because the settlement rules also create a predictable pattern in daily and intraday variances. Panel B of Table 2 summarizes the predicted daily variance pattern and some of the empirical work on variances related to the settlement process.

Spindt and Hoffmeister (1988) provide a model for daily and intraday federal funds rate variance based on the U.S. settlement rules. The predictions from their model (Table 2, column 1 of panel B) are as follows:

- daily variances are expected to increase on Fridays because positions must be taken on Friday to cover reserve requirements for Saturday and Sunday;
- daily variances are expected to increase as settlement approaches, with the largest daily variance on settlement Wednesday; and
- intraday variances are expected to increase as the close of the business day approaches.

Spindt and Hoffmeister focus on variances during a reserve maintenance period. Thus, their variance predictions do not include the effect of the transition from one reserve maintenance period to the next. We expect to see an empirically large variance on the first Thursday of the reserve maintenance period following the abatement of the settlement pressures from the preceding day.

The last five columns of Table 2, panel B, reproduce some empirical results on variances across a reserve maintenance period to highlight the importance of both the pattern and the magnitude of the variances. Columns 3 through 5 are reproduced from Table 4 in Griffiths and Winters (1995). Column 3 shows that daily variances increase on Fridays and as settlement approaches, with the variance on settlement Wednesday being by far the largest. Columns 4 and 5 provide variance estimates for the morning and afternoon on each day of the reserve maintenance period. The day-to-day patterns in the morning and afternoon are generally consistent with the daily pattern. In addition, the afternoon variance is larger than the morning variance for each day of the reserve maintenance period, as predicted by Spindt and Hoffmeister. Columns 6 and 7 of panel B in Table 2 are reproduced from Griffiths and Winters (1997). These daily variances are calculated using closing federal funds and overnight general collateral government repo rates. These variance results show that daily variances increase as settlement approaches, with the addition of a large variance in the first Thursday. In the federal funds market, the variance on settlement Wednesday is the largest daily variance; in repos, settlement Wednesday is large relative to the daily variances from the middle of the reserve maintenance period but is the second-largest daily variance. Accordingly, the empirical results on variances provide support for the predicted variance patterns across the reserve maintenance period and suggest that the general pattern in daily variances created by the

[7] Griffiths and Winters (1997) find that rates decline on Fridays in both government repos and Government National Mortgage Association (GMNA) repos. A review of the literature finds no day-of-the-week studies in commercial paper, negotiable CDs, bankers' acceptances, or Eurodollar deposits, so we are unable to provide additional support for declining rates on Fridays being a common occurrence across private-issue money market instruments. Because the issue of declining rates across private-issue money market instruments is outside the focus of this paper, we leave this issue for further research.

Table 2

Patterns in Rate Changes and Variances Related to Federal Reserve Settlement Rules

A: Daily Rate Changes

Day of reserve maintenance period	Prediction[a]	Expected empirical pattern	Daily high/low rate[a]	Daily average rate[b]	Closing federal funds rate[c]	Closing repo rate[c]	Intraday federal funds rate[d]
1st Thursday	None	–	–	+	NS	NS	NS
1st Friday	–	–	–	–	–	–	–
1st Monday	None	+	+	+	+	+	+
1st Tuesday	None	None	–	–	–	NS	–
1st Wednesday	None	None	–	–	NS	–	–
2nd Thursday	None	None	NS	NS	+	NS	NS
2nd Friday	–	–	–	–	–	–	–
2nd Monday	None	+	NS	+	+	+	+
2nd Tuesday	–	–	NS	–	–	NS	–
2nd Wednesday	+	+	+	+	+	+	+

B: Daily and Intraday Variances

Day of reserve maintendance period	Prediction[e]	Expected empirical pattern	Daily high/low rate[f]	Morning high/low rate[f]	Afternoon high/low rate[f]	Closing federal funds rate[g]	Closing repo rate[g]
1st Thursday	None	+	124	31	94	3.003	2.286
1st Friday	+	+	238	88	136	1.387	1.004
1st Monday	None	None	112	42	57	2.398	1.004
1st Tuesday	None	None	143	15	98	2.422	1.451
1st Wednesday	None	None	164	21	86	1.483	1.059
2nd Thursday	None	None	194	70	75	1.831	1.058
2nd Friday	+	+	310	73	164	1.842	1.225
2nd Monday	+	+	245	71	114	3.992	2.679
2nd Tuesday	+	+	1795	43	1587	3.171	1.792
2nd Wednesday	+	+	4029	249	3031	5.717	2.543

NOTE: This table provides theoretical and empirical evidence for the daily rate change pattern in the overnight federal funds rates and overnight general collateral repo rates. NS is an insignificant parameter estimate; None is either no prediction or no expectation.

[a]Griffiths and Winters (1995) provide a model that predicts specific daily rate changes in the federal funds market and provides empirical support for the predication using daily high and low federal funds rates.

[b]Hamilton (1996) examines daily rate changes and variance in a GARCH model using the daily average (effective) federal funds rate.

[c]Griffiths and Winters (1997) examine daily rate changes in the primary funding source (federal funds) and a substitute funding source (overnight general collateral repos) using daily closing rates.

[d]Cyree and Winters (2001) examine daily rate changes in a GARCH model using hourly federal funds rates.

[e]Spindt and Hoffmeister (1988) model the federal funds market. The model provides predictions for daily and intraday patterns in federal funds rate variances. We provide the daily pattern in this table and note that the intraday prediction is for variance to increase in the afternoon.

[f]Griffiths and Winters (1995) use high and low rates to estimate daily and intraday variances, and their estimates (reproduced here) support the predictions from Spindt and Hoffmeister (1988). We provide the Griffiths and Winters point estimates because the magnitude of the variance estimates is important.

[g]Griffiths and Winters (1997) use closing federal funds rates and closing overnight general collateral government repo rates to estimate the daily variance pattern in the primary (federal funds) and a secondary (government repos) funding source for management of Federal Reserve accounts for settlement.

settlement rules spills over into the variances of substitute funding sources.[8]

We note that Spindt and Hoffmeister (1988) and Griffiths and Winters (1995) model bank reserve account management in the absence of Federal Reserve open market activity to manage interest rates. Bartolini, Bertola, and Prati (2000 and 2001) extend the previous models by incorporating Fed market intervention into a model of bank behavior during reserve maintenance periods. The Bartolini, Bertola, and Prati (2001) model suggests that patterns in interest rate volatility should reflect the market's confidence in the Fed's commitment to rate targeting. They suggest that the Fed's move in 1994 toward more transparency in rate targeting and a tendency to change target rates only at FOMC meetings should give the market more confidence in the Fed's commitment to rate targeting. They suggest that, since 1994, less federal funds rate volatility across maintenance periods and as settlement approaches provides support for their model. We note that the patterns found by Bartolini, Bertola, and Prati (2001) are consistent with the literature cited above. However, it has been widely understood that the Fed has been targeting interest rates since the mid-1980s (see Thornton 1988 and 2002), so the patterns identified in the literature cited in Table 2 occurred during a period when the market understood that the Fed was managing interest rates and occurred despite the Fed's active efforts to manage interest rates.

In summary, the daily rate changes and variances that are unique to the U.S. settlement process are the biweekly rate increase and the variance increase on settlement Wednesdays. Accordingly, we focus on settlement Wednesdays to determine whether the effect of U.S. settlement rules reach overseas when Eurocurrency liabilities become a substitute funding source for reserve account management. However, before we can test for a biweekly settlement Wednesday effect, we must determine whether British banking regulations create any daily rate change or variance patterns that would appear biweekly on U.S. settlement Wednesdays.

[8] The only assets that are acceptable as reserves are vault cash and deposits at the Fed. Many possible sources can provide cash or Fed deposits, and (historically) most of these sources of funds alter a bank's required reserves. Accordingly, a substitute for federal funds is a funding source that does not alter required reserves or, in other words, has a 0 percent reserve requirement.

British Bank Settlement and LIBOR

LIBOR is the rate that major London banks offer to other banks on short-term funds. Thus, in the overnight market, LIBOR-based trades are similar to federal funds trades as banks make one-day trades of funds. Given the similarities in the trades in the overnight federal funds and overnight LIBOR-based markets, when the Fed reduced the reserve requirement on Eurocurrency liabilities to 0 percent, the overnight Eurodollar market became a viable substitute for the federal funds market as a source of deposits at the Fed for bank settlement. If the LIBOR-based market acts as a substitute, we would expect to see the settlement Wednesday effect spillover into LIBOR. However, before we can test LIBOR for U.S. Federal Reserve settlement effects, we must understand the settlement rules under which the London banks operate.

Settlement or clearing banks are required to keep small amounts of cash on deposit with the Bank of England. The deposits are not for monetary policy objectives or for clearing, but, instead, are intended to cover central bank operating costs. Each bank must cover its required reserves on a daily basis. With this daily settlement, the central bank intervenes in the market several times each day to ensure adequate liquidity.[9] Given daily settlement, the U.K. settlement process will not cause day-of-the-week regularities in rate changes and variances. Accordingly, we are able to test overnight LIBOR rates for U.S. settlement effects.

Swanson (1988), Fung and Isberg (1992), and Mougoue and Wagster (1997) examine the causality between three-month domestic CD rates and three-month Eurodollar rates and achieve mixed results on the direction of the causality. In this paper, we have a specific expectation on causality: In the absence of confounding effects in the British banking regulations, we expect that U.S. overnight market behavior resulting from Federal Reserve settlement rules will create the appearance of settlement effects in overnight LIBOR after the change to a 0 percent reserve requirement on Eurocurrency liabilities.

[9] The Bank of England conducts its daily trading at noon and 2:30 p.m. The Bank may also trade at 9:45 a.m. if it forecasts a large shortage for the day. The Bank stands ready to intervene for settlement banks only at 3:50 p.m. to provide the necessary end-of-the-day liquidity. The multiple daily interventions by the Bank of England create interesting intraday research opportunities. However, we have access only to daily closing overnight LIBOR rates and thus leave the intraday questions for further research.

DATA AND METHODS

Data

For this paper, we use daily closing data for (i) federal funds rates, (ii) overnight general-collateral government repo rates, and (iii) overnight LIBOR. LIBOR coincides with the London close while the other rates reflect the New York close. Since London time is typically five hours ahead of New York time, the London close occurs late morning New York time. Accordingly, overnight LIBOR established at the close in London is established during the late morning in New York. We also use three-month T-bill yields as a proxy for the general level of short-term interest rates. To control properly for contemporaneous changes in short-term interest rates and thus isolate the daily settlement effects, we must account for the time difference between London and New York in our application of T-bill annualized yields. Accordingly, we use the U.S. closing T-bill yields in our tests on federal funds rates and on overnight general-collateral government repo rates, and we use 11:00 a.m. (Eastern time) T-bill yields in our tests on overnight LIBOR.

The sample period covers August 4, 1986, through January 31, 1995. The beginning of the sample period coincides with the first available date for the overnight LIBOR rates. Ending the sample on January 31, 1995, stops the sample period before the majority of U.S. banks began actively using retail-deposit sweep programs. Active use of such sweep programs altered the reserve positions of U.S. banks to a point where Anderson and Rasche (2001) describe the reserve requirements for most banks as "voluntary constraints." Thus, the active use of sweep programs likely altered the reserve account management behavior of most banks. We end our sample at January 31, 1995, to avoid any possible change in reserve account management from the active use of retail-deposit sweep programs. Also, ending the sample period at this time provides approximately four years of data before and after the reserve reduction on Eurocurrency liabilities. The entire sample period coincides with the two-week reserve maintenance period used by the Federal Reserve.[10] The data on federal funds, overnight repos, and three-month T-bills were collected from the daily logs of the International Monetary Market (IMM) division of the Chicago Mercantile Exchange. The IMM acquires the data from Telerate, and our overnight LIBOR data were purchased from Knight-Ridder, Inc.

Methods

Griffiths and Winters (1997) test for settlement rate-change effects in overnight government repos in a three-equation SUR model (seemingly unrelated regression) with dependent variables for government repos, federal funds, and GNMA repos. The basic equation in their SUR model was

(1)
$$Spd_t = a_1D_{14} + a_2D_{15} + a_3D_{11} + a_4D_{12} + a_5D_{13} + a_6D_{24} + a_7D_{25} + a_8D_{21} + a_9D_{22} + a_{10}D_{23} + a_{11}TB_t + \varepsilon_t ,$$

where Spd_t is the change in the spread of an overnight rate (government repos, federal funds, GNMA repos) for day t ($Spd_t - Spd_{t-1}$) relative to three-month T-bill yields[11]; D_{ik} is a 0/1 dummy variable with i representing the first or second week of the reserve maintenance period and k representing the specific day of the week; and TB_t is the change in the three-month T-bill yield for day t ($yield_t - yield_{t-1}$). The change in T-bill yields is included in the model to control for changes in the general level of short-term interest rates.

The benefit of the SUR model is that it allows us to test for differences in parameter estimates across equations. The limitation of the SUR model in testing for settlement effects is that it does not allow us to incorporate the known daily heteroskedasticity in the federal funds market. Accordingly, Griffiths and Winters separately test daily variances with the following equation:

(2)
$$Var_t = a_1D_{14} + a_2D_{15} + a_3D_{11} + a_4D_{12} + a_5D_{13} + a_6D_{24} + a_7D_{25} + a_8D_{21} + a_9D_{22} + a_{10}D_{23} + \varepsilon_t ,$$

where Var_t is the square of the daily spread change.

[10] The Federal Reserve switched from a one-week reserve maintenance period to a two-week reserve maintenance period for the maintenance period beginning on February 2, 1984. From that date to the present, the Fed has used a two-week maintenance period.

[11] Griffiths and Winters (1997) note that standard conventions suggest that the change in spread be specified as the log relative [$\ln(Spd_t / Spd_{t-1})$] or as the percent change [$(Spd_t - Spd_{t-1})/Spd_{t-1}$]. However, the spread between overnight instruments and three-month T-bills is negative at several points during their sample period. A negative spread precludes using the standard conventions, and thus they use the first difference in daily spreads to calculate spread changes. We also have negative spreads at various times, and, therefore, we also use the first difference in daily spreads.

The presence of heteroskedastic daily variances suggests the need for a G/ARCH (general/autoregressive conditionally heteroskedastic) model that allows for the estimation of returns and conditional variance simultaneously.[12] Such a model allows for the explicit inclusion of heteroskedasticity in the estimation process, which improves the model's standard errors and thus the t statistics. However, a GARCH estimation precludes the direct testing of differences in parameter estimates between equations that can be done in an SUR model. So, to test for settlement effects, ideally we would like a model that includes features of both SUR models and GARCH models. However, at the present time such a model does not exist, so we must choose which type of model works best in this situation. Since we are most interested in determining whether a settlement effect exists and since we have other methods for comparing the size of the effects, we choose to use a GARCH model because of its benefits for the estimation of standard errors and t statistics.

Specifically, we chose a GARCH-M model, where the M denotes that the conditional variance is included in the mean equation and allows the conditional variance to provide information about returns (rate changes in this setting). Then, from the set of GARCH-M models, we chose the model proposed by Glosten, Jagannathan, and Runkle (GJR) (1993) because of its definition of the asymmetric term in the conditional variance. All GARCH-M models contain an intercept and two prespecified variables in the variance equation: (i) a trend term (the ARCH effect) and (ii) a persistence term (the GARCH effect). The innovation in the GJR-M model is a third prespecified term that is based on the sign of the prediction errors (the asymmetric term). This asymmetric term allows that certain prediction errors are more important to the market than other prediction errors. The GJR-M asymmetric term is a binary dummy variable, which we believe is intuitively appealing for the federal funds/LIBOR markets. In addition, we chose to estimate the GJR-M model with robust errors to calculate properly the t statistics for the model parameter estimates (see Bollerslev and Wooldridge, 1992).

In the reserve maintenance process, banks typically have either excess reserves (actual reserves > required reserves) or short reserves (actual reserves < required reserves) and the period-ending excess or short position carries forward as the starting reserve position in the next reserve maintenance period. Excess reserves are an opportunity cost for the bank because it could have invested the excess but did not. Conversely, short reserves indicate that the bank has invested federal funds using an interest-free loan from the Federal Reserve. Thus, a clear asymmetry exists, which we believe makes the dummy variable definition for the asymmetric term in the GJR-M model appropriate.

We also tested two other specifications of GARCH-M to ensure that our results were not an artifact of our model choice. The other models are the EGARCH-M specification suggested by Nelson (1991), with two different specifications of the error term in the mean equation: (i) a normal distribution of errors and (ii) a generalized error distribution. The results from these specifications are qualitatively similar to the results reported below.[13] They are omitted here for brevity, but are available upon request.

Cyree and Winters (2001) examine closing federal funds rates in a GARCH-M model and find significant ARCH/GARCH effects. They suggest that these ARCH/GARCH effects are consistent with the Spindt and Hoffmeister (1988) model of daily federal funds rate variances. Cyree and Winters do not find a significant mean effect for the conditional variance in federal funds rate changes. However, we include the conditional variance in the mean equation for the spread changes we examine to allow for the possibility that the conditional variance provides information about daily spread changes, thus ensuring that we have properly isolated the daily settlement effects in the spreads.

We start our model by putting equations (1) and (2) in the GARCH-M model developed by GJR

[12] See Bollerslev, Chou, and Kroner (1992) for a review of the finance literature using GARCH models.

[13] We tested the specification of our GJR-M model against the EGARCH-M models using likelihood ratio tests and found no significant difference between the models. Accordingly, we chose to report the GJR-M model results because of the intuitive appeal of the model relative to the incentives created by the Federal Reserve settlement rules. In addition, we performed the Box-Ljung Q-test on the residuals and squared residuals. In all cases, the residuals are negatively serially correlated, as expected, and the inclusion of lagged rate changes in the model does not remove the serial correlation. Also, including lagged rate changes in the model does not alter the interpretation (i.e., the qualitative values or significance) of the daily dummy variables. We note that Hamilton (1996) concludes that banks do not view reserves from different days as perfect substitutes, so we chose to use a model specification without lagged rate changes because we believe this best fits the economic realities of the federal funds market and its substitute markets for overnight money.

(1993).[14] We extend equations (1) and (2) to include controls for quarter-ends (see Hamilton, 1996), and we remove the change in T-bill yields from the mean equation.[15] We estimate our GARCH-M model with all the daily dummy variables in equations (1) and (2) and find results that are generally consistent with previous empirical results and with our expectations for LIBOR. However, we note that the model becomes difficult to converge and the results are very sensitive to starting point inputs. We believe the difficulties lie in the large number of dummy variable parameter estimates required when we use all the dummy variables. Furthermore, we have argued that settlement Wednesday is the one day that is unique to the U.S. bank settlement process, so we modify our model to reduce the number of dummy variables in the model. The new model converges easily and is not as sensitive to starting point inputs, yet the results remain consistent with the full-model empirical results, with previous empirical results in the literature, and with expectations. The model used to generate the results reported in the tables is as follows:

The GJR-M mean equation is:

$$\Delta Spd_t = a_0 + a_1 1stFriday + a_2 2ndFriday + a_3 SetWednesday + a_4 QTR + \lambda\sigma_t^2 + \varepsilon_t, \quad (3)$$

where ΔSpd_t is the first difference in the daily spread on day t and the spread is defined as the rate on the overnight instrument minus the yield on three-month T-bills[16]; *1stFriday* is a 0/1 dummy variable that equals 1 on the first Friday of each reserve maintenance period and 0 otherwise; *2ndFriday* is a 0/1 dummy variable that equals 1 on the second Friday of each reserve maintenance period and 0 otherwise; *SetWednesday* is a 0/1 dummy variable that equals 1 on settlement Wednesday and 0 otherwise; *QTR* is a 0/1 dummy variable that equals 1 on the last trading day of each quarter and 0 otherwise; σ_t^2 is the conditional variance (volatility-in-mean term).

Note that in dummy variable regression models with omitted dummy variables, the parameter estimates of the remaining dummy variables are relative to the average of the omitted dummy variables. Fridays have falling rates, so we chose to keep the two Friday dummy variables in the model so that any model bias from omitting some dummy variables would be against finding a settlement Wednesday effect in the mean equation. The conditional variance equation for the GJR-M model is

$$\sigma_t^2 = b_0 + b_1\varepsilon_{t-1}^2 + b_2\sigma_{t-1}^2 + b_3\varepsilon_{t-1}^2 I_t + c_1 1stFriday + c_2 2ndFriday + c_3 SetWednesday + c_4 QTR, \quad (4)$$

where I_t is a 0/1 dummy variable that equals 1 if the lagged error is negative and 0 otherwise. The indicator variable accounts for asymmetry in the conditional variance equation as the addition to the variance based on the sign of last period's error.

Note that in equation (4) the daily dummy variables and quarter-end dummy variable are contemporaneous with the conditional variance (which is not common in a GARCH model). This is important (and appropriate) in the daily dummy variables because the daily variances created by the settlement rules are conditioned by the day of the reserve maintenance period. Similarly, the quarter-end dummy variable is contemporaneous because, as Hamilton (1996) shows, there are significant quarter-end conditioning effects on the daily variance.

Summary Statistics on Daily Spreads

Table 1 shows the mean (median) spreads for the overnight rates relative to three-month T-bill yields for both the entire sample period and the subsample period after the reserve requirement reduction. Note that only after the reserve requirement change are Eurocurrency liabilities a substitute funding source for federal funds in the settlement process. In addition to these summary statistics, we conducted various mean difference tests. We find that each average spread is significantly lower (at the 1 percent level) after the reserve reduction. In

[14] GJR uses non-standard GARCH-M notation. We chose to present our model using standard GARCH(1,1)-M notation as in Chou, Engle, and Kane (1992), with the addition of the GJR binary asymmetric term in the conditional variance.

[15] Griffiths and Winters (1997) include the change in T-bill yields as a proxy for the change in the general level of short-term interest rates. We attempted to include the change in T-bill yields as an explanatory variable in the mean equation, but very inconsistent parameter estimates on T-bill yields across federal funds, repos, and LIBOR lead us to remove T-bill yields from the explanatory variables in the mean equation. Note, T-bills still appear as the base rate in the spreads used as the dependent variable. Also note that equation (4) cannot include a T-bill variable as an explanatory variable because T-bill yields are part of the dependent variable in equation (3); thus, including T-bill yields in equation (4) would create a contemporaneous endogenous variable, which is not allowed in the conditional variance of a GARCH model.

[16] While we are directly interested in the rate change behavior of our test rates, we define the dependent variable in terms of a spread to prevent any general trends in the levels of rates from affecting our results. An alternative to using T-bill yields as the base rate would be to use the daily average rate for federal funds, repos, and LIBOR as the base rate. This could be done for federal funds by using the effective federal funds rate. However, we know of no source for daily average repo rates or for daily average LIBOR, so we chose to stay with T-bill yields as our base rate.

the period after the reserve reduction, we find that the average repo and LIBOR spreads are not different from each other, but that both are significantly higher (at the 1 percent level) than the average federal funds spread. Thus, after the implicit reserve tax was removed, the spread on LIBOR is similar to the spread on government repos—making the LIBOR-based liabilities a viable alternative to repos as a substitute for federal funds, as would be expected from the law of one price. Based on the average spreads, the substitute funding sources are not an attractive alternative to federal funds. However, the rate comparison may be very different on settlement Wednesdays.

RESULTS

This section presents our empirical results. The focus of our tests is to investigate overnight LIBOR spreads for U.S. settlement effects. However, before we can investigate LIBOR spreads we must first test for the presence of a settlement effect in federal funds spreads (the primary funding source) and overnight government repo spreads (the domestic substitute funding source) across our sample period. Accordingly, we begin our settlement effect tests on federal funds spreads. We follow federal funds with tests on overnight repos, and we conclude with tests on overnight LIBOR.

Federal Funds

We begin our analysis by estimating our GARCH-M model on the daily change in federal funds spreads. We divide our sample into two subsamples: before and after the lifting of the reserve requirement on Eurocurrency liabilities. The 3 percent reserve regime ended with the maintenance period that closed December 12, 1990, and the 0 percent regime began with the maintenance period that opened December 27, 1990. The purpose of this analysis of federal funds rates is to determine whether the previously identified settlement effects in rate changes and variance appear in federal funds rates during both subperiods of the our sample. The results on federal funds are reported in Table 3.

The first column of Table 3 reports results for the period from August 4, 1986, through December 12, 1990. The mean (spread-change) equation has, in general, the predicted pattern. That is, spreads fall on Fridays and rise on settlement Wednesday relative to average spread. The increase on settlement Wednesday supports a settlement effect in rate changes.

The variance equation shows significant ARCH/GARCH effects, which suggests that daily variances are conditional. The significant and positive ARCH effect (ε^2_{t-1}) suggests trends in the daily variances, while the significant and positive GARCH effect (σ^2_{t-1}) suggests that shocks persist in the daily variances. Both effects are consistent with the model of daily federal funds variances by Spindt and Hoffmeister (1988). The settlement Wednesday parameter estimate is positive and significant at the 1 percent level, and it is much larger in magnitude than either of the Friday parameter estimates. This finding is consistent with a settlement effect in the conditional variance.

These results suggest that the predicted rate change and variance regularities are present in federal funds during the subperiod when Eurocurrency liabilities carried a 3 percent reserve requirement. With these results, it is reasonable to expect a settlement Wednesday effect in rates on substitute funding sources. A substitute funding source for federal funds is a U.S. dollar source of deposits at the Fed that carries a 0 percent reserve requirement and, thus, through the law of one price should cost a bank approximately the same interest rate as federal funds. During the period with the 3 percent reserve requirement on Eurocurrency liabilities, government repos were a substitute for federal funds but Eurocurrency liabilities were not. So, during this time period we would expect to find a settlement Wednesday effect in overnight government repo rates, but not in LIBOR.

The third column of Table 3 presents the parameter estimates from our GARCH-M model for federal funds during the subperiod from February 7, 1991, through January 31, 1995. Beginning this subperiod on February 7, 1991, removes the first three reserve maintenance periods under the new rules from our analysis. This allows the market a little time to adjust to the new rules. Feinman (1993) notes, in fact, that substantial volatility occurred with the change to the new rules. Also, the manager of the federal funds trading desk of a large regional bank has said that it took banks several weeks to adjust to the new rules. Accordingly, we believe that removing the first three reserve maintenance periods, while clearly arbitrary, is reasonable.[17]

[17] Including these periods prevents our GARCH-M model from converging for federal funds due to the unusual volatility during this six-week period. Using other methods that allow us to include the six-week period, we find qualitatively similar results to our GARCH results excluding the six weeks. These results are omitted for brevity, but are available upon request. Because we anticipate daily heteroskedasticity, we believe a GARCH model is the best method for our tests.

Table 3

Federal Funds GARCH Model Results

	8/4/86–12/12/90		2/7/91–1/31/95	
Variable	Estimate	*t* statistic	Estimate	*t* statistic
Mean equation				
Intercept	–0.0015	–0.31	0.0219***	3.74
1stFriday	–0.0382**	–2.45	–0.1263***	–9.34
2ndFriday	–0.0689***	–4.59	–0.0803***	–8.17
SetWednesday	0.1315***	5.34	0.0840**	2.24
QTR	0.0957	1.30	0.1281**	2.05
σ_t^2	0.0600**	2.13	–0.1058***	–3.26
Variance equation				
Intercept	0.0284***	43.16	0.2088***	4.98
ε_{t-1}^2	0.0194**	2.22	0.4357***	6.28
σ_{t-1}^2	0.8980***	35.03	0.6459***	11.51
$\varepsilon_{t-1}^2 I_t$	–0.7211***	–26.89	0.1877	1.41
1stFriday	–0.0187***	–10.47	–0.0356***	–2.68
2ndFriday	–0.0031	–1.26	–0.0170**	–1.95
SetWednesday	0.3609***	6.74	0.3686***	3.70
QTR	0.1145**	2.23	0.7416*	1.85

NOTE: This table reports the results for estimating the GARCH-M model on the spread of the overnight federal funds rate relative to the closing three-month T-bill rates. The model is estimated twice. The first estimation is for the period from 8/4/86 through 12/12/90 when the reserve requirement for Eurocurrency liabilities is 3 percent. The second estimation is for the period from 2/7/91 through 1/31/95 when the reserve requirement for Eurocurrency liabilities is 0 percent. The mean equation for the GARCH-M model is $\Delta Spd_t = a_0 + a_1 1stFriday + a_2 2ndFriday + a_3 SetWednesday + a_4 QTR + \lambda\sigma_t^2 + \varepsilon_t$.
The conditional variance equation is $\sigma_t^2 = b_0 + b_1\varepsilon_{t-1}^2 + b_2\sigma_{t-1}^2 + b_3\varepsilon_{t-1}^2 I_t + c_1 1stFriday + c_2 2ndFriday + c_3 SetWednesday + c_4 QTR$.
*/**/*** denote significance at the 10/5/1 percent levels.

The spread change results on Fridays and settlement Wednesday are consistent with the predicted pattern and consistent with the results from the earlier subperiod. The variance equation shows significant ARCH/GARCH effects.[18] The variance equation also shows a settlement Wednesday parameter estimate that is positive and significant at the 1 percent level and substantially larger than the Friday parameter estimates. This finding is consistent with expectations and the results from the previous subperiod and is consistent with Feinman's (1993) observation of increased variance after the reserve reduction.

These results suggest that the predicted pattern is present in the later subperiod. Again, this means we would expect to find a settlement Wednesday effect in rates of substitute funding sources.

Overnight Government Repos

Table 4 presents the results from estimating our GARCH-M model on overnight government repos. Government repos are a substitute funding source for federal funds in the settlement process across our entire sample period. Accordingly, we expect to find settlement effects in overnight government repo rate changes and variances across our entire sample period. It is important to identify a settlement effect in a substitute funding source

[18] Note that the sum of the ARCH and GARCH parameter estimates is greater than 1, which is usually considered a sign of an estimation error because it suggests that the conditional variance grows rapidly. However, during this time period, federal funds rates are much more volatile than during the earlier sample, with settlement Wednesday volatility being extremely high. Accordingly, we believe these estimates are reasonable.

Table 4

Overnight Government Repo GARCH Model Results

	8/4/86–12/12/90		2/7/91–1/31/95	
Variable	Estimate	*t* statistic	Estimate	*t* statistic
Mean equation				
Intercept	–0.0082*	–1.75	–0.0011	–0.19
1stFriday	–0.0368***	–3.01	–0.0876***	–5.89
2ndFriday	–0.0224	–1.28	–0.0622***	–5.53
SetWednesday	0.0540**	2.00	0.0603***	2.73
QTR	0.0385	0.58	0.1589	1.64
σ_t^2	0.4502***	5.25	0.0094	0.11
Variance equation				
Intercept	0.0102***	25.62	0.0264***	32.26
ε_{t-1}^2	0.2835***	1.51	0.0433***	2.68
σ_{t-1}^2	0.3471***	15.28	0.4732***	18.62
$\varepsilon_{t-1}^2 I_t$	–0.2591***	–10.46	–0.3917***	–14.14
1stFriday	–0.0100***	–7.05	–0.0143***	–6.88
2ndFriday	0.0128***	3.47	–0.0159***	–8.51
SetWednesday	0.0627***	13.53	0.0453***	7.27
QTR	0.0698***	3.73	0.1805**	2.52

NOTE: This table reports the results for estimating the GARCH-M model on the spread of the overnight government repo rate relative to the closing three-month T-bill rates. The model is estimated twice. The first estimation is for the period from 8/4/86 through 12/12/90 when the reserve requirement for Eurocurrency liabilities is 3 percent. The second estimation is for the period from 2/7/91 through 1/31/95 when the reserve requirement for Eurocurrency liabilities is 0 percent. The mean equation for the GARCH-M model is $\Delta Spd_t = a_0 + a_1 1stFriday + a_2 2ndFriday + a_3 SetWednesday + a_4 QTR + \lambda\sigma_t^2 + \varepsilon_t$.
The conditional variance equation is $\sigma_t^2 = b_0 + b_1\varepsilon_{t-1}^2 + b_2\sigma_{t-1}^2 + b_3\varepsilon_{t-1}^2 I_t + c_1 1stFriday + c_2 2ndFriday + c_3 SetWednesday + c_4 QTR$.
*/**/*** denote significance at the 10/5/1 percent levels.

across the entire period for comparison with Eurocurrency liabilities, which only became a substitute in the latter part of the sample period. The first column of Table 4 reports the results from the subperiod when Eurocurrency liabilities carried a 3 percent reserve requirement, and the third column of Table 4 reports the results from the subperiod with no reserve requirement.

The first set of results shows a negative and significant estimate at the 1 percent level for the first Friday and a positive and significant estimate at the 5 percent level for settlement Wednesday. The variance results indicate significant and positive ARCH/GARCH effects, which are consistent with the results for federal funds. Both Fridays and settlement Wednesday contribute significantly to the conditional variance. The settlement Wednesday parameter estimate is positive and about five times larger than the Friday parameters. These results are consistent with the existence of settlement effects in a substitute funding source during the subperiod with the 3 percent reserve requirement.

The second set of results on overnight government repos shows strong evidence of the predicted spread change effects with significant (1 percent level) and negative spread changes on Fridays and a significant (1 percent level) and positive spread change on settlement Wednesday. The variance equation shows significant and positive ARCH/GARCH effects. As with the results on repos for the earlier subperiod, settlement Wednesday provides a significant (1 percent level) and positive contribution to the conditional variance and its parameter estimate is substantially larger (about three times)

Table 5

LIBOR GARCH Model Results

	8/4/86–12/12/90		2/7/91–1/31/95	
Variable	Estimate	*t* statistic	Estimate	*t* statistic
Mean equation				
Intercept	–0.0001	–0.03	0.0177***	4.96
1stFriday	0.0028	0.25	–0.0791***	–8.01
2ndFriday	0.0148	1.03	–0.0217***	–2.61
SetWednesday	–0.0163	–1.50	0.0368**	2.54
QTR	0.0303	0.56	0.0876	1.24
σ_t^2	0.0693	0.98	–0.5458***	–5.61
Variance equation				
Intercept	0.0030***	22.27	0.0132***	54.60
ε_{t-1}^2	0.5360***	96.37	0.0821***	8.26
σ_{t-1}^2	0.4442***	27.51	0.4504***	17.39
$\varepsilon_{t-1}^2 I_t$	–0.0514**	–2.18	–0.2533***	–6.20
1stFriday	0.0004	0.27	–0.0025***	–2.67
2ndFriday	0.0235***	12.79	–0.0087***	–12.18
SetWednesday	0.0045***	2.79	0.0251***	9.30
QTR	0.0465***	3.04	0.0954**	2.16

NOTE: This table reports the results for estimating the GARCH-M model on the spread of the overnight LIBOR rate relative to the closing three-month T-bill rates. The model is estimated twice. The first estimation is for the period from 8/4/86 through 12/12/90 when the reserve requirement for Eurocurrency liabilities is 3 percent. The second estimation is for the period from 2/7/91 through 1/31/95 when the reserve requirement for Eurocurrency liabilities is 0 percent. The mean equation for the GARCH-M model is $\Delta Spd_t = a_0 + a_1 1stFriday + a_2 2ndFriday + a_3 SetWednesday + a_4 QTR + \lambda\sigma_t^2 + \varepsilon_t$.
The conditional variance equation is $\sigma_t^2 = b_0 + b_1\varepsilon_{t-1}^2 + b_2\sigma_{t-1}^2 + b_3\varepsilon_{t-1}^2 I_t + c_1 1stFriday + c_2 2ndFriday + c_3 SetWednesday + c_4 QTR$.
*/**/*** denote significance at the 10/5/1 percent levels.

than the parameter estimates on Fridays. These results show settlement effects in a substitute funding source during the subperiod with the 0 percent reserve requirement.

LIBOR

To this point, we have shown that the Fed's settlement regulations influence spread changes and variance in the primary funding source (federal funds) and in the domestic substitute funding source (government repos). In this subsection, we explore the hypothesis that LIBOR spreads become influenced by Fed settlement regulations when Eurocurrency liabilities become a substitute funding source for federal funds through the reduction in the reserve requirement to 0 percent. Again, we remind the reader that only vault cash and deposits at the Fed can be used for settlement; therefore, a comparable substitute for federal funds would need to be a source of overnight U.S. dollar deposits at the Fed with a 0 percent reserve requirement, which, through the law of one price, should carry approximately the same interest rate as federal funds.

Table 5 presents the results from estimating our GARCH-M model on overnight LIBOR spreads. The first column reports the results from the subperiod with a 3 percent reserve requirement, and the third column reports the results from the subperiod with the 0 percent reserve requirement on Eurocurrency liabilities. In this formulation, we switch to the 11:00 a.m. quote for three-month Treasury bill rates. This allows for a time-consistent spread for (closing) overnight LIBOR rate quotes from the U.K. market relative to the relevant T-bill quote.

The first set of results shows no evidence of the settlement effects found for federal funds and overnight government repos during this subperiod when Eurocurrency liabilities were not a substitute funding source. Specifically, there is no significant spread increase on settlement Wednesday and settlement Wednesday is not the largest daily contribution to the conditional variance. These results suggest that, during the time period when Eurocurrency liabilities carried a reserve tax, banks did not use Eurocurrency liabilities in managing their reserve accounts.

The second set of results shows strong evidence of settlement effects. Specifically, the spread changes in the mean equation show a significantly positive settlement Wednesday effect. The variance equation shows significant and positive ARCH/GARCH effects. Settlement Wednesday provides a positive and significant (1 percent level) contribution to the conditional variance, with its parameter estimate being about three times larger than the parameter estimates on Fridays. These results are consistent with the results from federal funds and overnight government repos from the same subperiod, which suggests that during this period U.S. banks did become the marginal borrower in dollar-based Eurocurrency liabilities as they managed their reserve accounts on settlement Wednesdays.

The results for LIBOR in the first subperiod are clearly different from those in the second subperiod. With significant settlement effects in the latter period, the combination of LIBOR results suggests that the change in the reserve requirements on Eurocurrency liabilities had a significant impact on the overnight money markets in London.

The Size of Settlement Wednesday

At this point, we have shown that after the reserve reduction on Eurocurrency liabilities the average federal funds spread is significantly smaller than the average spreads on the substitute funding sources and that settlement Wednesday effects in spread changes exist in each asset. In this section, we discuss the size of the settlement Wednesday spread change in each instrument after the reserve reduction and the appropriate implications. Recall that the timing of the LIBOR results aligns with the late morning in the U.S. and that Griffiths and Winters (1995) show that the majority of the settlement Wednesday effect appears in the afternoon. So, since the London market closes in late morning, one would not anticipate the same magnitude of spread changes in the London market as in the domestic markets.

To address the size of the settlement Wednesday spread change effect, we calculate the average spread change on each funding source on settlement Wednesday.[19] We find the following average spread changes on settlement Wednesdays after the reserve reduction: (i) a 24-basis-point increase for federal funds, (ii) a 15-basis-point increase for overnight government repos, and (iii) an 8-basis-point increase in overnight LIBOR.

These findings are consistent with our expectations for the following reasons. First, the primary funding source should have the largest increase because it is, on average, the cheapest and most convenient source of funds (because banks are already regular and active participants in federal funds trading) and thus should be the focus of the settlement Wednesday rate pressure. Second, the rate pressure should be greater in overnight repo rates than overnight LIBOR because of the timing of the two markets: The brokered repo market is active during the afternoon in the United States, while the London LIBOR-based Eurocurrency market is closed. Thus, overnight repo rates are open to the significant rate pressures of settlement Wednesday afternoons while overnight LIBOR is not.

Control Variables

There are results from some of the control variables that deserve a brief discussion. We have delayed the discussion until this point to avoid detracting from the primary focus of the paper, which is the pervasiveness of the Federal Reserve settlement regulations.

Quarter-ends exhibit significant and positive contributions to the conditional variances across all instruments in this study and across both sample subperiods. However, only one of six quarter-end parameter estimates in the mean equations is significant. This combination of results suggests uncertainty in these markets at quarter-ends without a consistent increase or decrease in rate pressure. This finding suggests a great deal of funds movement at quarter-end without squeezes or shortages.

The asymmetric term in the conditional variance is significant and negative in five of six estimates. The one exception is for federal funds in the latter

[19] We cannot directly use the parameter estimates reported from the GARCH-M model because the spread change equation includes a conditional variance effect.

period, which was unusually volatile. A significant parameter estimate suggests that some errors are more important to the market than other errors, and a negative parameter estimate suggests a reduction in the conditional variance. A negative parameter estimate from our specification of the asymmetric term suggests that, when the actual rate is less than the rate the market expected, conditional variance declines.

CONCLUSION

In this paper, we examine overnight LIBOR around the change by the Federal Reserve in the reserve requirement on Eurocurrency liabilities from 3 percent to 0 percent. The change in the reserve requirement eliminates the reserve tax on Eurocurrency liabilities and thus allows the law of one price to make Eurocurrency liabilities a viable alternative funding source for federal funds for banks in managing their reserve accounts. We find no evidence of settlement Wednesday effects in LIBOR during the period of the 3 percent reserve requirement, but we find strong evidence of settlement Wednesday effects in LIBOR during the period of the 0 percent reserve requirement. Our results suggest that, when Eurocurrency liabilities are a substitute funding source for federal funds, banks use them in managing their reserve accounts—becoming the marginal borrower on settlement Wednesdays. In contrast, when Eurocurrency liabilities carry a reserve tax, banks use other funding sources in managing their reserve accounts.

Our results suggest the following: Overnight money markets are global and the micromechanics of one market can spillover into another market and, in this case, carry the effects of Federal Reserve settlement rules into off-shore markets. That is, we demonstrate that changes in Fed policy allowing Eurocurrency liabilities to be a substitute funding source in the settlement process results in LIBOR being affected by the Fed's bank settlement procedures.

REFERENCES

Anderson, Richard G. and Rasche, Robert H. "Retail Sweep Programs and Bank Reserves, 1994-1999." Federal Reserve Bank of St. Louis *Review*, January/February 2001, *83*(1), pp. 51-72.

Bartolini, Leonardo; Bertola, Giuseppe and Prati, Allessandro. "Banks' Reserve Management, Transaction Costs, and the Timing of Federal Reserve Intervention." *Journal of Banking and Finance*, July 2001, *25*(7), pp. 1287-317.

_____; _____ and _____. "Day-to-Day Monetary Policy and the Volatility of the Federal Funds Interest Rate." *Journal of Money, Credit, and Banking*, February 2002, *34*(1), pp. 137-59.

Bollersev, Tim; Chou, Ray Y. and Kroner, Kenneth F. "ARCH Modeling in Finance." *Journal of Econometrics*, April-May 1992, *52*(1-2), pp. 5-59.

_____ and Wooldridge, Jeffrey M. "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances." *Econometric Review*, 1992, *11*, pp. 143-72.

Chou, Ray Y.; Engle, Robert F. and Kane, Alex. "Measuring Risk Aversion from Excess Returns on a Stock Index." *Journal of Econometrics*, April-May 1992, *52*(1-2), pp. 201-24.

Cyree, Ken B. and Winters, Drew B. "Analysis of Federal Funds Rate Changes and Variance Patterns." *Journal of Financial Research*, Fall 2001, *24*(3), pp. 403-18.

Feinman, Joshua N. "Reserve Requirements: History, Current Practice, and Potential Reform." *Federal Reserve Bulletin*, June 1993, *79*(6), pp. 569-89.

Fung, Hung-Gay and Isberg, Steven C. "The International Transmission of Eurodollar and U.S. Interest Rates: A Cointegration Analysis." *Journal of Banking and Finance*, August 1992, *16*(4), pp. 757-69.

Glosten, Lawrence R.; Jagannathan, Ravi and Runkle, David E. "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance*, December 1993, *48*(5), pp. 1779-801.

Goodfriend, Marvin and Hargraves, Monica. "A Historical Assessment of the Rationales and Functions of Reserve Requirements." Federal Reserve Bank of Richmond *Economic Review*, 1983, *69*, pp. 3-21.

Griffiths, Mark D. and Winters, Drew B. "Day-of-the-Week Effects in Federal Funds Rates: Further Empirical Findings." *Journal of Banking and Finance*, October 1995, *19*(7), pp. 1265-84.

_____ and _____. "The Effect of Federal

Reserve Accounting Rules on the Equilibrium Level of Overnight Repo Rates." *Journal of Business Finance and Accounting*, July 1997, *24*(6), pp. 815-32.

_____ and _____. "An Examination of the 1992 Increase in the Allowable Carryover of Reserves in the Bank Settlement Process." *Financial Review*, February 2000, *41*, pp. 67-84.

Hamilton, James D. "The Daily Market for Federal Funds." *Journal of Political Economy*, February 1996, *104*, pp. 26-56.

Mougoue, Mbodja and Wagster, John. "The Causality Effects of the Federal Reserve's Monetary Policy on U.S. and Eurodollar Interest Rates." *Financial Review*, November 1997, *32*(4), pp. 821-44.

Nelson, Daniel B. "Conditional Hetersokedasticity in Asset Returns: A New Approach." *Econometrica*, March 1991, *59*(2), pp. 347-70.

Spindt, Paul A. and Hoffmeister, J. Ronald. "The Micromechanics of the Federal Funds Market: Implications for Day-of-the-Week Effects in Funds Rate Variability." *Journal of Financial and Quantitative Analysis*, December 1988, *23*(4), pp. 401-16.

Swanson, Peggy E. "The International Transmission of Interest Rates: A Note on Causal Relationships Between Short-term External and Domestic U.S. Dollar Returns." *Journal of Banking and Finance*, December 1988, *12*(4), pp. 563-73.

Thornton, Daniel L. "The Borrowed-Reserves Operating Procedure: Theory and Evidence." Federal Reserve Bank of St. Louis *Review*, January/February 1988, *70*(1), pp. 30-54.

_____. "The Fed and Short-term Rates: Is It Open Market Operations, Open Mouth Operations, or Interest Rate Smoothing." *Journal of Banking and Finance*, 2003 (forthcoming).

Identifying Business Cycle Turning Points in Real Time

Marcelle Chauvet and Jeremy M. Piger

A common feature of industrialized economies is that economic activity moves between periods of expansion, in which there is broad economic growth, and periods of recession, in which there is broad economic contraction. Understanding these phases, collectively called the business cycle, has been the focus of much macroeconomic research over the past century. In the United States, the National Bureau of Economic Research (NBER), a private, nonprofit research organization, serves a very useful role in cataloging stylized facts about business cycles and providing a historical accounting of the dates at which regime shifts occur. This task began soon after the founding of the NBER in 1920 and has continued to the present day.[1] Since 1980, the specific task of dating "turning points" in U.S. business cycles, or those dates at which the economy switches from the expansion regime to the contraction regime and vice versa, has fallen to the NBER's Business Cycle Dating Committee.[2]

The NBER dates a turning point in the business cycle when the committee reaches a consensus that a turning point has occurred. Although each committee member likely brings different techniques to bear on this question, the decision is framed by the working definition of a business cycle provided by Arthur Burns and Wesley Mitchell (1946, p. 3):

> Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions and revivals which merge into the expansion phase of the next cycle.

A fundamental element of this definition is the idea that business cycles can be divided into distinct phases, with the phase shifts characterized by changes in the dynamics of the economy. In particular, expansion phases are periods when economic activity tends to trend up, whereas recession phases are periods when economic activity tends to trend down. In practice, to date the transition from an expansion phase to a recession phase, or a business cycle peak, the NBER looks for clustering in the shifts of a broad range of series from a regime of upward trend to a regime of downward trend. The converse exercise is performed to date the shift back to an expansion phase, or a business cycle trough.

The NBER's announcements garner considerable publicity. Given this prominence, it is not surprising that the business cycle dating methodology of the NBER has come under some criticism. Two criticisms that are of primary interest in this paper are as follows: First, because the NBER's decisions represent the consensus of individuals who bring differing techniques to bear on the question of when turning points occur, the dating methodology is neither transparent nor reproducible. Second, the NBER business cycle peaks and troughs are often determined well after the fact. This practice appears to be largely the result of the NBER's desire to avoid calling false turning points.

Of course, the NBER is not the only source of information regarding business cycle turning points. Economists and statisticians have developed many statistical methods that automate the dating of business cycle peaks and troughs (see Boldin, 1994, for a summary). One such technique is the Markov-switching model. This model, popularized by Hamilton (1989) in the economics profession, is capable of statistically identifying shifts in the

---

[1] For an interesting history of the NBER's role in defining and dating the business cycle, see Moore and Zarnowitz (1986).

[2] There are currently six members of the committee: Robert Hall of Stanford University, Martin Feldstein of Harvard University, Jeffrey Frankel of the University of California at Berkeley, Robert Gordon of Northwestern University, N. Gregory Mankiw of Harvard University, and Victor Zarnowitz of Columbia University.

Marcelle Chauvet is an associate professor at the University of California, Riverside, and an associate policy advisor at the Federal Reserve Bank of Atlanta. Jeremy M. Piger is an economist at the Federal Reserve Bank of St. Louis. The authors thank James Bullard, Michael Dueker, James Hamilton, James Morley, and Michael Owyang for helpful comments. Mrinalini Lhila and John Zhu provided research assistance.

parameters of a statistical process driving a time series of interest. These models are quite simple, making them transparent and reproducible. Also, Layton (1996) provides some evidence that Markov-switching models provide timely identification of business cycle turning points.[3]

In this paper we take it as given that the NBER correctly identifies the dates of business cycle turning points. We then evaluate the real-time performance of the Markov-switching model in replicating the NBER's business cycle dates. We apply the model to two datasets, growth in quarterly real gross domestic product (GDP) and growth in monthly economy-wide employment. We first confirm the result found elsewhere that the model is able to replicate the historical NBER business cycle dates fairly closely when estimated using all available data. Second, we evaluate the real-time performance of the model at dating business cycles over the past 40 years; this is accomplished by estimating the model on recursively increasing samples of data and evaluating the evidence for a new turning point at the end of each sample.

This approach builds on the exercise undertaken in Layton (1996), extending it in two main ways. First, while Layton used fully revised data in his recursive estimations, here we use "real-time" data. That is, for each recursive sample we use only data that would have been available at the end of the sample period being considered. This method provides a more realistic assessment of how the model would have performed, as it does not assume knowledge of data revisions that were not available at the time the model would have been used. Second, we extend Layton's sample to include the 2001 recession, in order to investigate the properties of the model in the most recent business cycle.

The results of this exercise suggest that the model chooses turning points in real time that are fairly close to the NBER dates. In addition, we find evidence that the model would have identified business cycle turning points faster than the NBER Business Cycle Dating Committee, with a larger lead time in the case of troughs. The switching model achieves this performance with few incidences of false positives. Overall, these results suggest that the Markov-switching model is a potentially very useful tool to use alongside the traditional NBER analysis.

[3] The Markov-switching approach to dating business cycle turning points has recently received media attention: Kevin Hassett advocated its use on the editorial page of the March 28, 2002, edition of the *Wall Street Journal*.

Of course, this line of research is predicated on the assumption that turning point dates are interesting concepts, and some might question whether these dates have any worthwhile intrinsic meaning. We argue that they do. There is much evidence that the two regimes defined by the NBER turning point dates are quite different, beyond the distinction of expansion versus contraction. First, knowledge of which regime the economy is in can improve forecasts of economic activity (see, for example, Hamilton, 1989). Second, there is evidence that the relationship between economic variables changes over NBER-identified phases. For example, McConnell (1998) and Gavin and Kliesen (2002) have shown that the relationship between initial claims for unemployment insurance and employment growth is stronger during NBER-dated recessions. Third, there is growing evidence that fluctuations in output during NBER recession episodes are purely temporary, whereas those during NBER expansion episodes are permanent (see, for example, Beaudry and Koop, 1993, and Kim, Morley, and Piger, 2002). This finding is suggestive of a "plucking" model for U.S. output, in which the business cycle is characterized more by negative deviations from trend output than by positive deviations.[4] Such a pattern is not generally implied by linear macroeconomic models of the business cycle, suggesting that the NBER dates define interesting economic episodes from a modeling perspective. Finally, the NBER dates, regardless of whether they have intrinsic meaning, garner considerable attention from the media and politicians. Thus, if the economics community is going to produce estimates of turning points, we should be interested in developing accurate, timely, and transparent methods for doing so.

In the next section we provide a review of the Markov-switching model used in this paper. The third section discusses the full sample and "real-time" performance of the model for dating turning points in the business cycle.

THE MARKOV-SWITCHING MODEL OF BUSINESS CYCLE DYNAMICS

As discussed in the introduction, the NBER definition of a business cycle places heavy emphasis on regime shifts in the process driving economic

[4] The "plucking" terminology was coined by Milton Friedman (1964, 1993).

activity. In the past 15 years there have been enormous advances in formally modeling regime shifts in a rigorous statistical framework. In a paper published in 1989, James Hamilton developed an extremely useful tool for statistically modeling regime shifts in autoregressive time series models. To understand this model, it is useful to begin with a simple linear time-series framework for the growth rate of some measure of economic activity, y_t:

$$(1) \qquad \begin{aligned} (y_t - \mu) &= \rho(y_{t-1} - \mu) + \varepsilon_t \\ \varepsilon_t &\sim N(0, \sigma^2). \end{aligned}$$

In this model, the growth rate of economic activity has a mean denoted by μ. Deviations from this mean growth rate are created by the stochastic disturbance, ε_t. These deviations are serially correlated, modeled as an AR(1) time-series process with parameter ρ.

Hamilton's innovation was to allow the parameters of the model in (1) to switch between two regimes, where the switching is governed by a state variable, $S_t = \{0,1\}$. When $S_t = 0$ the parameters of the model are different from those when $S_t = 1$. Clearly, if S_t were an observed variable, this model could simply be estimated using dummy variable methods. However, Hamilton showed that even if the state is *unobserved*, the parameters of the model in each state could be estimated if one is willing to place restrictions on the probability process governing S_t. Hamilton derives an estimation technique that could be used to estimate the model when the probability process governing S_t is a first-order Markov chain. This stipulation simply means that any persistence in the state is completely summarized by the value of the state in the previous period. Under this assumption, the probability process driving S_t is captured by the following four transition probabilities:

$$(2) \qquad \begin{aligned} P(S_t = 1 | S_{t-1} = 1) &= p \\ P(S_t = 0 | S_{t-1} = 1) &= 1 - p \\ P(S_t = 0 | S_{t-1} = 0) &= q \\ P(S_t = 1 | S_{t-1} = 0) &= 1 - q \end{aligned}$$

Clearly, conclusions regarding when S_t changes may depend on which parameters of the model are allowed to change. For example, the instances when the data may support regime shifts in the variance of the disturbance, σ^2, may be at different times from those in the autoregressive parameter, ρ. Thus, if we are interested in using this model for identifying the NBER's turning point dates, we should allow regime switching in those parameters of the model that seem to change from expansion to recession. Hamilton showed that allowing the mean growth-rate parameter, μ, to vary with S_t seems to be adequate for this task. In particular, Hamilton specified the following augmented version of (1):

$$(3) \qquad \begin{aligned} (y_t - \mu_{s_t}) &= \rho(y_{t-1} - \mu_{s_{t-1}}) + \varepsilon_t \\ \varepsilon_t &\sim N(0, \sigma^2) \\ \mu_{s_t} &= \mu_0 + \mu_1 S_t \\ \mu_1 &< 0, \end{aligned}$$

where S_t depends on the transition probabilities in (2). Here, when S_t switches from 0 to 1, the growth rate of economic activity switches from μ_0 to $\mu_0 + \mu_1$. Since $\mu_1 < 0$, the model will estimate these switches at times when economic activity switches from high-growth to low-growth states. Hamilton applied this technique to the growth rate of U.S. gross national product and found the best fit when $\mu_0 > 0$ and $\mu_0 + \mu_1 < 0$, suggesting the model was capturing regimes when the economy was expanding versus regimes when the economy was contracting. The estimated probability that S_t was equal to 1 conditional on all the data in the sample, denoted $P(S_t = 1 | T)$, corresponded very closely to NBER recession dates. This finding was particularly striking in that Hamilton estimated his model with only one variable describing economic activity.

Since the publication of Hamilton's paper, a large number of alternative Markov-switching models of the business cycle have been studied. Boldin (1994) fits the Hamilton model to an alternative measure of economic activity, namely, the unemployment rate. Other authors, for example, Hansen (1992), allow for regime-switching in parameters other than the mean growth rate, such as the residual variance or autoregressive parameters. The Hamilton model was modified to allow for additional phases in business cycle dynamics by Sichel (1994), Kim and Nelson (1999), and Kim, Morley, and Piger (2002). Finally, building on work by Diebold and Rudebusch (1996), Chauvet (1998) and Kim and Yoo (1995) extended the Hamilton model to a multivariate framework, estimating a coincident index of economic activity with a regime-switching mean growth rate.

DATING BUSINESS CYCLES WITH THE SWITCHING MODEL

Model Specification and Estimation

In this paper we work with the model given in equations (2) and (3), which is applied to two different measures of economic activity for which rich unrevised "real-time" datasets are available. The first is the growth rate of quarterly real U.S. GDP, yielding a model very similar to that originally estimated by Hamilton. The second is a higher-frequency measure of economic activity, monthly nonfarm payroll employment.

The model is estimated using Hamilton's (1989) nonlinear filter. The algorithm provides the conditional probability of the latent Markov state, which permits evaluation of the conditional likelihood of the observable variable. The filter evaluates this likelihood function, which can be maximized with respect to the model parameters using a nonlinear optimization algorithm. The estimation procedure and derivation of the likelihood function are described in detail in Hamilton (1994). Following Albert and Chib (1993), we set the autoregressive coefficient, ρ, equal to 0, a priori. This specification seems best able to replicate the historical record of NBER business cycle dates.

Full-Sample Business Cycle Dates

Before analyzing the real-time ability of the model to date turning points, we are first interested in their ability to replicate the NBER business cycle chronology using all available data. Thus, we first estimate the model using data on (i) growth in real GDP from the second quarter of 1947 through the second quarter of 2002 and (ii) data on nonfarm payroll employment growth from February 1947 through July 2002. The GDP data are from the July 31, 2002, release from the Bureau of Economic Analysis, and the employment data are from the August 2, 2002, release from the Bureau of Labor Statistics.

As a first step in evaluating the ability of the model to replicate the NBER turning point dates, consider Figures 1A and B, which hold the estimated probability that $S_t = 1$ conditional on all the data in the sample, or $P(S_t = 1|T)$, for both applications of the model. In the graphs, shading indicates NBER-labeled recessions. The graphs suggest that the model captures the NBER chronology fairly closely. During periods that the NBER classifies as expansions, $P(S_t = 1|T)$ is usually close to 0. Near the point where the NBER recession begins, $P(S_t = 1|T)$ spikes upward and remains high until around the time when the NBER dates the end of the recession.

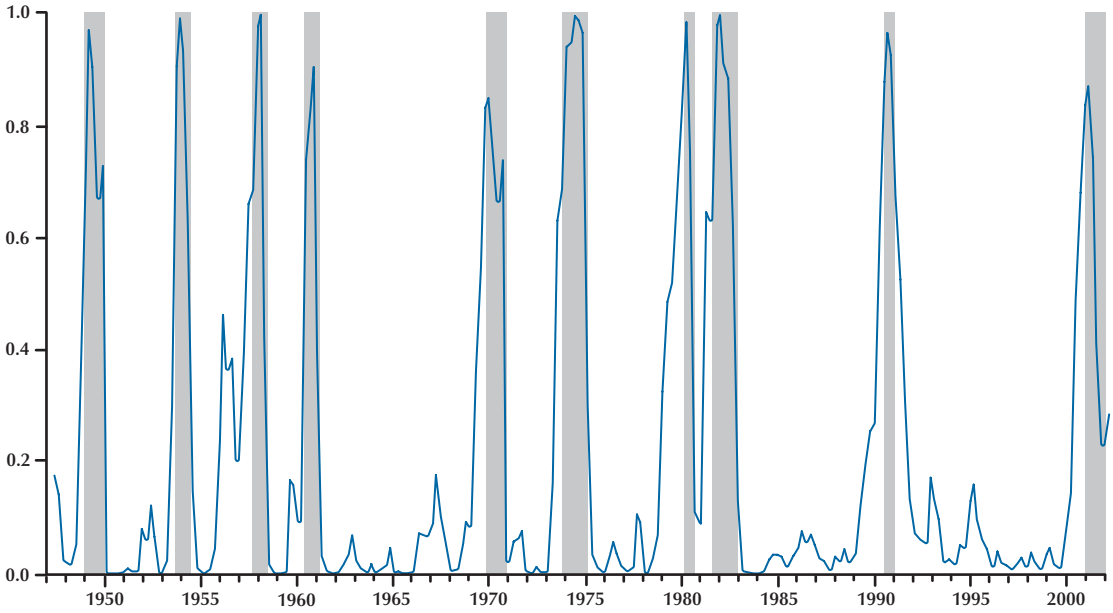
Although visual inspection of the probabilities suggests comparability, it is difficult to tell how close the turning points from the Markov-switching model are to the NBER dates without the tabulation of specific dates based on the probabilities produced by the model. To do this, a formal definition is needed to convert the probabilities produced by the switching model into turning point dates. One approach, used by Hamilton (1989) among others, is to classify a turning point as occurring when $P(S_t = 1|T)$ moves from below 50 percent to above 50 percent or vice versa. This approach has an intuitive appeal as it separates times when an expansion state is more likely from those when a recession state is more likely. This rule would be problematic if $P(S_t = 1|T)$ fluctuated around 50 percent, in which case an excessive number of business cycle peaks and troughs would be called. However, since the Markov-switching model applied to the GDP and employment series produces probabilities that are generally close to 0 or 1, we adopt this simple definition.

We augment this definition with one of two rules specifying how long a phase must persist before a turning point is identified. For example, suppose $P(S_t = 1|T)$ moves from below 50 percent to above 50 percent. Should we immediately declare a business cycle peak has occurred and the economy has entered a recession phase? Or should we require confirmation of the recession phase, by verifying that $P(S_{t+1} = 1|T)$, $P(S_{t+2} = 1|T)$, $\ldots P(S_{t+k} = 1|T)$ are all above 50 percent? A smaller value for k increases the speed at which a turning point might be identified, but increases the chances of calling a false positive. Our first rule is defined for maximum speed, requiring only that a single occurrence of a probability moving from above (below) 50 percent to below (above) 50 percent must be observed before a turning point is determined. Our second rule, consistent with the NBER tradition of not classifying very short downturns or expansions as separate regimes, requires that a recession or expansion last at least 3 months before a new turning point is defined. Note that for real GDP, which is measured quarterly, this requirement is met with only a single occurrence of a probability crossing 50 percent, meaning that rule 1 is identical to rule 2. For employment data, which is measured monthly, rule 2 requires three consecutive probabilities above (below) 50 percent and will thus differ from rule 1.

Formally, our turning point rules for employ-

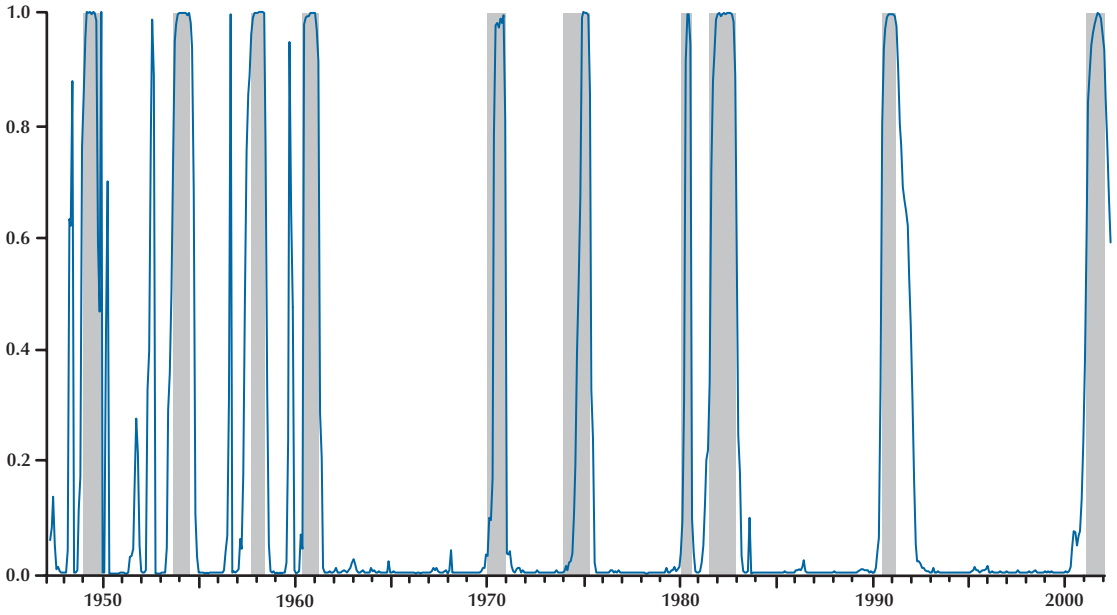
Figure 1

A. Full-Sample Estimated $P(S_t = 1)$ from Markov-Switching Model of Quarterly Real GDP



NOTE: Data vintage July 31, 2002. Shaded areas denote NBER recession dates.

B. Full-Sample Estimated $P(S_t = 1)$ from Markov-Switching Model of Monthly Nonfarm Payroll Employment



NOTE: Data vintage August 2, 2002. Shaded areas denote NBER recession dates.

ment and GDP growth can be specified using the following definitions:

Monthly Employment Growth. Definition 1: A business cycle *peak* is said to occur in month $t+1$ if the economy was in an expansion in month t and

Rule 1: $P(S_{t+1}=1)\geq 0.5$;

Rule 2: $P(S_{t+1}=1)\geq 0.5$ and $P(S_{t+2}=1)\geq 0.5$ and $P(S_{t+3}=1)\geq 0.5$.

Definition 2: A business cycle *trough* is said to occur in month t if the economy was in a recession in month t and

Rule 1: $P(S_{t+1}=1)<0.5$;

Rule 2: $P(S_{t+1}=1)<0.5$ and $P(S_{t+2}=1)<0.5$ and $P(S_{t+3}=1)<0.5$.

GDP Growth. Definition 1: A business cycle *peak* is said to occur in quarter $t+1$ if the economy was in an expansion in quarter t and $P(S_{t+1}=1)\geq 0.5$.

Definition 2: A business cycle *trough* is said to occur in quarter t if the economy was in a recession in quarter t and $P(S_{t+1}=1)<0.5$.

Table 1A contains the NBER turning point dates and the corresponding dates obtained from the Markov-switching model applied to real GDP growth based on the above definitions. The agreement between the two is striking. The Markov-switching model captures each of the NBER business cycle peaks and troughs in the sample. The average discrepancy between the 10 NBER business cycle peaks and the business cycle peaks from the switching model applied to real GDP growth is approximately 2.4 months, with a maximum discrepancy of 6 months and a standard deviation of 1.8 months. Business cycle troughs are dated even closer. There is no discrepancy on average between the nine NBER business cycle troughs and the business cycle troughs from the switching model (the two dates are the same for six of the nine troughs), with a maximum discrepancy of 6 months and a standard deviation of around 2.7 months. Generally the model tends to determine turning points at or before the ones established by the NBER. The only exception is for the 1990-91 recession trough, for which the switching model dates the trough two quarters after the NBER date.

In addition to capturing each of the NBER business cycle dates, the switching model applied to GDP growth generates no false business cycle dates. That is, for the whole sample, the probability of recession only increased (decreased) above (below) 50 percent near the beginning or end of an actual recession. Thus, for the model applied to real GDP, an increase or decrease in the probability of recession above or below 50 percent sends a very strong signal that a turning point has actually occurred.

Table 1B shows the NBER turning point dates and the corresponding dates obtained from the Markov-switching model applied to monthly employment growth under rule 1 defined above. The agreement between the two sets of dates is very close, although somewhat less so than that obtained from GDP. There are two reasons for this. First, we are using employment at the monthly frequency, which is a much more noisy series than quarterly GDP. Second, employment slightly lags the business cycle. Generally employment falls after a recession begins and increases after it ends, as employers are reluctant to fire (or hire) until a recession gains intensity (or there are clear signs of its end). Nevertheless, the switching model applied to monthly employment captures each of the NBER business cycle peaks and troughs in the sample. The average discrepancy between the NBER peaks and the peaks from the switching model is approximately 1 month, with a maximum discrepancy of 9 months and a standard deviation of 3.6 months. Similarly, the average discrepancy between the NBER trough dates and the trough dates from the switching model is 1.8 months, with a maximum discrepancy of 10 months and a standard deviation of 3.2 months.

The trough dates from the switching model applied to employment tend to slightly lag the NBER dates. In particular, all troughs from employment either lag (five of nine) or coincide (four of nine) with the NBER's. The results are mixed for peak dates: Half of the peak dates from the model either coincide or lead the NBER peak dates, whereas half lag the NBER dates.

In addition to the business cycle dates in Table 1B, turning point rule 1 identified three false-positive business cycle dates, all early in the sample. If the minimum number of consecutive months that $P(S_t=1|T)$ is required to be above (below) 50 percent before a turning point is identified were increased to two, only a single false-positive result occurs (February 1948). Under turning point rule 2 defined above, in which $P(S_t=1|T)$ is required to be above (below) 50 percent for three consecutive months before a turning point is defined, there are no false-positive results. This is achieved with no tradeoff in terms of missed NBER dates; rule 2 still captures

Table 1

Business Cycle Dates—NBER and Markov-Switching Models Estimated Over Full Sample

Peak			Trough		
NBER	Switching model	Lead/lag discrepancy	NBER	Switching model	Lead/lag discrepancy
A. Real GDP					
November 1948	1948:Q4	0	October 1949	1949:Q4	0
July 1953	1953:Q3	0	May 1954	1954:Q2	0
August 1957	1957:Q2	1Q	April 1958	1958:Q1	1Q
April 1960	1960:Q2	0	February 1961	1960:Q4	1Q
December 1969	1969:Q3	1Q	November 1970	1970:Q4	0
November 1973	1973:Q3	1Q	March 1975	1975:Q1	0
January 1980	1979:Q3	2Q	July 1980	1980:Q3	0
July 1981	1981:Q2	1Q	November 1982	1982:Q4	0
July 1990	1990:Q2	1Q	March 1991	1991:Q3	–2Q
March 2001	2000:Q4	1Q	Not yet announced	2001:Q3	—
Mean		0.8Q			0.0Q
Median		1.0Q			0.0Q
Standard deviation		0.6Q			0.9Q
B. Nonfarm Payroll Employment					
November 1948	October 1948	1M	October 1949	October 1949	0
July 1953	June 1953	1M	May 1954	August 1954	–3M
August 1957	April 1957	4M	April 1958	May 1958	–1M
April 1960	May 1960	–1M	February 1961	February 1961	0
December 1969	April 1970	–4M	November 1970	November 1970	0
November 1973	August 1974	–9M	March 1975	April 1975	–1M
January 1980	April 1980	–3M	July 1980	July 1980	0
July 1981	August 1981	–1M	November 1982	December 1982	–1M
July 1990	July 1990	0	March 1991	January 1992	–10M
March 2001	February 2001	1M	Not yet announced	Not yet identified	—
Mean		–1.1M			–1.8M
Median		–0.5M			–1.0M
Standard deviation		3.6M			3.2M

NOTE: Leads (lags) are represented by + (–) and indicate how many months the switching model anticipates (lags) the NBER dating, whereas 0 indicates that the two dating systems coincide.

all of the NBER business cycle peaks and troughs in the sample.

Real-Time Business Cycle Dates

In this section we investigate the real-time performance of the switching model for dating business cycles. This will involve an out-of-sample evaluation of the model's performance. Our out-of-sample period will be the past 40 years of data, with prior data used for initial estimation of the model. We are interested in the following question: Had the switching model been used to date business cycles in the past, how would it have performed? We are particularly interested in the ability of the model to accurately and quickly identify in real time the six NBER peaks and five NBER troughs over this period. We are also interested in the incidence of false business cycle dates, that is, business cycle dates identified by the model that do not correspond to an NBER-dated peak or trough.

There are two features of such a real-time exercise. First, only data over the sample period that the business cycle analyst would have had available at that time should be used. We achieve this first requirement by using a recursive estimation routine. This routine works as follows: We begin with data that extends from the second quarter of 1947 to the third quarter of 1965 for real GDP and from February 1947 to October 1964 for employment. The model is estimated and the probability of a new turning point at the end of the sample evaluated. The sample is then extended by one data point, the model reestimated, and the probability of a turning point evaluated. This process is repeated until the final sample is reached, which extends from the second quarter of 1947 to the second quarter of 2002 for real GDP and from February 1947 to July 2002 for employment.

The second feature of the real-time exercise is to assume no more knowledge of data revisions than what would have been known by an econometrician estimating the model at the time. Thus, for each end-of-sample date in the recursive estimation routine, we use the first available release of these data. For example, for our first sample for real GDP data, which extends from the second quarter of 1947 through the third quarter of 1965, we use the first release of data that included the third quarter of 1965. For real GDP these data were available by the beginning of the second month of the fourth quarter of 1965, which we refer to as the *vintage* of this dataset. The monthly employment datasets are similar, except they are more timely than the GDP data. In particular, the first release of employment data for a given month is usually available by the first week of the subsequent month. We obtained the real-time datasets for quarterly real GDP and monthly payroll employment from the Federal Reserve Bank of Philadelphia.[5]

In evaluating the evidence for a turning point, we consider the probability of a recession at the end of the sample for that particular vintage, that is, $P(S_T = 1 | T)$, where T denotes the end of the sample period. This will be referred to as the "real-time recursive probability" throughout the remainder of the paper. Such an estimated probability, which is estimated for time t using time t information, is often called a "filtered" probability. This is, of course, less information than econometricians would have had available to them at the time, as econometricians would also have had the so-called "smoothed" probabilities for prior dates, that is, $P(S_t = 1 | T)$, where $t < T$. Thus, while the model might miss a turning point at time t for the dataset that ends at time t, it might catch this turning point for the dataset that ends at T. We do not allow for this possibility in the following, thus placing the model at a disadvantage for dating turning points. However, as will be shown, the model's performance is still quite good despite this disadvantage.

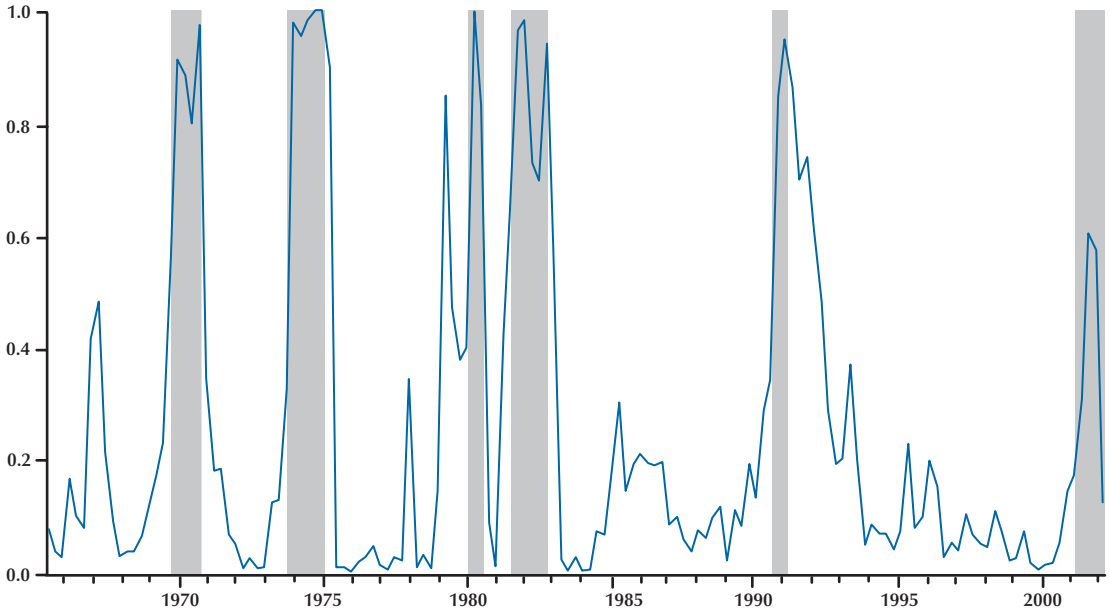
Figures 2A and B plot the real-time recursive probability of a recession at the end of the sample against the NBER business cycle dates. That is, the point on the graph for date t represents the estimated probability of recession at date t for the recursive sample that ended on date t. The probabilities are closely related to the NBER turning points, tending to increase or decrease substantially only around NBER peaks and troughs. The real-time recursive probabilities of recession from the employment data are noisier than those from GDP growth, which is not surprising given the higher frequency of the employment data.

We next move to tabulation of business cycle dates using turning point rule 1 for converting probabilities into business cycle dates defined in the section "Full-Sample Business Cycle Dates." Tables 2A and B contain the NBER business cycle peak and trough dates and the corresponding dates identified in real time by the switching model. The top frame of each table evaluates the performance of the model in capturing business cycle peaks. The bottom frame evaluates business cycle troughs. The first column

[5] See Croushore and Stark (2001) for information regarding this dataset.

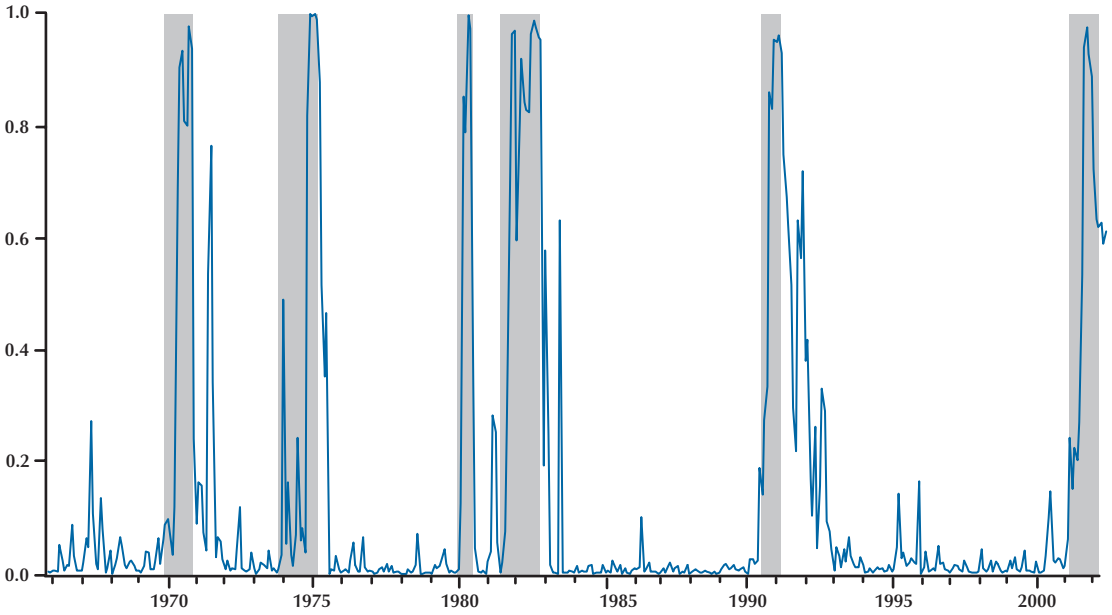
Figure 2

A. Real-Time Recursively Estimated $P(S_t = 1)$ from Markov-Switching Model of Quarterly Real GDP



NOTE: Shaded areas denote NBER recession dates.

B. Real-Time Recursively Estimated $P(S_t = 1)$ from Markov-Switching Model of Monthly Nonfarm Payroll Employment



NOTE: Shaded areas denote NBER recession dates.

Table 2A

Recession Dates Obtained in Real Time—NBER and Markov-Switching Model of Real GDP Estimated Over Recursive Samples

Peak date: switching model	Peak date available: switching model	Peak date: NBER	Peak date announced: NBER	Lead/lag discrepancy	Lead announcement date: switching model
1969:Q4	February 1970	December 1969	—	0	—
1974:Q1	May 1974	November 1973	—	–1Q	—
1980:Q2	August 1980	January 1980	June 3, 1980	–1Q	–2M
1981:Q3	November 1981	July 1981	January 6, 1982	0	2M
1990:Q4	February 1991	July 1990	April 25, 1991	–1Q	2M
2001:Q3	November 2001	March 2001	November 26, 2001	–2Q	0
Mean				–0.8Q	0.5M
Median				–1.0Q	1.0M
Standard deviation				0.8Q	1.9M

Trough date: switching model	Trough date available: switching model	Trough date: NBER	Trough date announced: NBER	Lead/lag discrepancy	Lead announcement date: switching model
1970:Q4	February 1971	November 1970	—	0	—
1975:Q2	August 1975	March 1975	—	–1Q	—
1980:Q3	November 1980	July 1980	July 8, 1981	0	8M
1983:Q1	May 1983	November 1982	July 8, 1983	–1Q	2M
1992:Q1	May 1992	March 1991	December 22, 1992	–4Q	7M
2001:Q4	February 2002	Not yet announced	Not yet announced	—	—
Mean				–1.2Q	5.7M
Median				–1.0Q	7.0M
Standard deviation				1.6Q	3.2M

gives the first date (labeled "Peak date") that a turning point was assigned in real time by the switching model. The second column gives the date this turning point would have first been available. For example, the first entry in the second column of Table 2A is February 1970. This is the date at which the business cycle peak of the fourth quarter of 1969, listed in the first column, would have first been identified using the switching model (as early February is approximately when the first iteration of GDP data for the fourth quarter of 1969 would have been available). The third and fourth columns give the official NBER business cycle dates and when they were announced. Note that the NBER Business Cycle Dating Committee began dating business cycle peaks and troughs in real time with the 1980 recession. Thus, the dates of these announcements are recorded in the table only from this date on. The fifth column records the discrepancy between the peak or trough date first assigned by the switching model and the corresponding date assigned by the NBER, which is the amount of time the date in column 1 precedes that in column 3. The final column shows how far in advance of the NBER date the switching model date would have been available—that is, the amount of time the date in column 2 anticipates that in column 4.

Tables 2A and B demonstrate that the switching

Table 2B

Recession Dates Obtained in Real Time—NBER and Markov-Switching Model of Nonfarm Payroll Employment Estimated Over Recursive Samples

Peak date: switching model	Peak date available: switching model	Peak date: NBER	Peak date announced: NBER	Lead/lag discrepancy	Lead announcement date: switching model
May 1970	June 1970	December 1969	—	–5M	—
Novemer 1974	December 1974	November 1973	—	–12M	—
April 1980	May 1980	January 1980	June 3, 1980	–3M	1M
November 1981	December 1981	July 1981	January 6, 1982	–4M	1M
November 1990	December 1990	July 1990	April 25, 1991	–4M	4M
September 2001	October 2001	March 2001	November 26, 2001	–6M	1M
Mean				–5.7M	1.8M
Median				–4.5M	1.0M
Standard deviation				3.3M	1.5M

Trough date: switching model	Trough date available: switching model	Trough date: NBER	Trough date announced: NBER	Lead/lag discrepancy	Lead announcement date: switching model
November 1970	December 1970	November 1970	—	0	—
May 1975	June 1975	March 1975	—	–2M	—
July 1980	August 1980	July 1980	July 8, 1981	0	11M
December 1982	January 1983	November 1982	July 8, 1983	–1M	6M
Augugust 1991	September 1991	March 1991	December 22, 1992	–5M	15M
July 2002?	Not yet identified	Not yet announced	Not yet announced	—	—
Mean				–1.6M	10.7M
Median				–1.0M	11.0M
Standard deviation				2.1M	4.5M

model calls turning point dates in real time that are fairly close to the NBER dates. Table 2A shows the following: For the six NBER peaks in the past 40 years, the switching model applied to real GDP growth yields business cycle dates in real time that were exactly equal to the NBER's in two cases and one or two quarters away in the other cases. The average discrepancy for peaks is 2.4 months with a standard deviation of 2.4 months. For the five NBER business cycle troughs, the trough dates from the model applied to real GDP growth coincide with the NBER dates in two cases and lag one or four quarters in the other cases. The average discrepancy is 3.6 months with a standard deviation of 4.8 months.

For the model applied to employment, Table 2B shows that the real-time probabilities of recession generally lag the NBER turning points, especially in the case of peak dates. The average discrepancy between the model and the NBER peak dates is 5.7 months with a standard deviation of 3.3 months. For trough dates, the average discrepancy is only 1.6 months with a standard deviation of 2.1 months.

Tables 2A and B demonstrate that the model, when applied to real GDP growth as well as to employment growth, identifies in real time each of the NBER business cycle episodes over the past 40 years. However, this performance would be less impressive if the model also identified numerous

Table 3

Real-Time Turning Point Signal Error—Markov-Switching Models

A. Real GDP Growth

Turning point evaluation (6 recessions: 6 NBER peaks, 5 troughs)	
Correct TP	11
Missed TP	0
False TP	1

B. Employment Growth

Turning point evaluation (6 recessions: 6 NBER peaks, 5 troughs)	Rule 1	Rule 2
Correct TP	11	11
Missed TP	0	0
False TP	2	0

NOTE: Correct TP refers to prediction of a turning point when an NBER turning point occurs. Missed TP refers to prediction of no turning point when an NBER turning point occurs. False TP refers to prediction of a turning point when no NBER turning point occurs.

other "false" business cycle episodes in real time. Tables 3A and B summarize the incidence of such false identifications. From Table 3A, over this 40-year period the dating algorithm applied to real GDP identified only one false business cycle date, in the second quarter of 1979. This increase in the probability of recession signaled an actual slowdown in the U.S. economy in 1979, associated with the second oil shock, and preceded the 1980 recession. From Table 3B, the model applied to employment growth identifies two false business cycle dates using turning point rule 1. The first of these was in June and July 1971, when the probabilities increased above 50 percent with no corresponding NBER recession. The other false turning point for employment occurs immediately following the 1990-91 recession. Using turning point rule 1, the switching model initially dated the trough of this recession as August 1991. However, $P(S_T = 1 | T)$ then increased above 50 percent again from November 1991 to January 1992, thus dating a double-dip recession following the 1990-91 recession. Using turning point rule 2, both of these false turning points would have been ruled out, leaving no false business cycle dates. This would not have come at the expense of any missed NBER business cycle episodes, as turning point rule 2 also captured all 11 NBER turning points.

We now turn to the issue of whether the switching model applied in real time would have identified turning points any faster than the NBER Business Cycle Dating Committee. The sixth column of Tables 2A and B, generated using rule 1, suggests that the answer is yes for both peak and trough dates obtained from the model applied to either real GDP or employment growth. Business cycle peak dates were determined an average of 0.5 months and 1.8 months before the NBER announcement, using the model applied to real GDP growth and employment growth, respectively. The model improves on the timeliness of the NBER even more in determining business cycle trough dates. For the three business cycle troughs in the past 25 years, the model applied to GDP would have determined these dates an average of 5.7 months prior to the NBER, with a maximum of 8 months for the 1980 trough. When applied to employment, the model would have determined trough dates an average of 10.7 months prior to the NBER announcements. The model provides a longer lead time when applied to the employment series partially because the employment series is released more quickly than the GDP series.

What should one conclude from the above analysis? We have focused on two evaluation criteria for the regime-switching model: (i) its ability to identify business cycle turning points that correspond to NBER business cycle peaks or troughs and are in fairly close proximity to the NBER dates and (ii) the speed with which the peaks and troughs are identified. Whether one prefers the switching model or to wait for the NBER announcement depends on the relative weight one attaches to each type of performance. If the primary interest is to quickly

determine whether a recession has begun or ended, the switching model will likely be preferable. If, instead, the primary interest is to obtain the exact NBER date of the business cycle peak or trough, with relatively little weight on the speed with which this is obtained, the switching model will be of less interest. We argue that the switching model presented here provides an improvement in the speed at which NBER business cycle dates are identified, with a reasonable tradeoff in the accuracy of the dates assigned. This performance is impressive given that the model is based on only a single variable.

The 2001 Recession

The most recent U.S. recession merits further discussion for at least two reasons. First, data revisions in recent months have caused significant revisions in the real-time peak date established by the switching model. Indeed, this revision matches or exceeds the largest seen in the sample period considered in Table 2. It is worth exploring further the reasons for these large revisions. Second, the trough date for this recession had not yet been established when this paper was written, providing us with an out-of-sample experiment of the usefulness of the switching model.

In November 2001 the NBER Dating Committee dated the peak of the previous expansion as March 2001. In contrast, the real-time recursive probability of a recession, given by $P(S_T = 1|T)$, first rose above 50 percent in the third quarter of 2001 for the model applied to real GDP and in September 2001 for the model applied to employment growth (Tables 2A and B). A more detailed look at these recession probabilities is given in the first column of Tables 4A and B and shows the real-time recursive probability of a recession at each date over the past several years.

The recent large revisions in GDP and employment data changed the peak date obtained from the switching model. The second column of Tables 4A and B show the smoothed probability of a recession using the most recent data available, which was the July 31, 2002, vintage for real GDP and the August 2, 2002, vintage for employment. Using this data, the switching model dates the recession as beginning much earlier, in the fourth quarter of 2000 for real GDP growth and in February 2001 for employment growth. The large revision in the peak date stems from recent data revisions that indicated significantly slower growth than previously recorded for the first six months of 2001. For example, the release of real GDP data dated June 27, 2002, from

Table 4

Probabilities of Recession from Markov-Switching Models

Period	Recursive in real time (percent)	Full sample using revised data
A. Applied to GDP Growth (percent)		
2000		
Q1	1.6	7.8
Q2	1.9	14.2
Q3	0.5	48.7
Q4	14.6	67.8
2001		
Q1	17.4	83.6
Q2	30.9	86.9
Q3	60.4	74.2
Q4	57.3	41.0
2002		
Q1	12.7	22.9
Q2	28.4	28.4
B. Applied to Employment Growth (percent)		
2001		
Jan	1.1	36.3
Feb	1.5	50.0
Mar	6.4	68.2
Apr	24.4	85.0
May	15.2	89.8
Jun	22.6	94.2
Jul	20.3	96.3
Aug	27.2	97.8
Sep	53.4	99.1
Oct	94.0	99.8
Nov	97.6	99.7
Dec	92.8	98.8
2002		
Jan	88.8	96.4
Feb	72.2	94.2
Mar	63.3	88.0
Apr	61.9	81.6
May	62.8	73.8
Jun	59.0	66.4
Jul	61.2	61.2

the Bureau of Economic Analysis recorded quarterly annualized growth of 1.3 and 0.3 percent for the first and second quarters of 2001, respectively. However, the data released on July 31, 2002, instead recorded declines in GDP of 0.6 and 1.6 percent in these quarters. These data revisions altered the peak date established by the switching model, pushing it much earlier—into late 2000 and early 2001.

Again, at the time this paper was written, the NBER had not yet dated the end of the 2001 recession. However, the switching model applied to real GDP growth has already dated the business cycle trough. The real-time probabilities indicate that the end of the recession occurred in the fourth quarter of 2001. This date would have been available with the initial release of the fourth quarter 2001 GDP data, in February 2002. Using the revised GDP data released in late July, the model dates the trough even earlier, in the third quarter of 2001. Using data up to the August 2, 2002, vintage, the switching model applied to employment growth had not yet dated the end of the recession.

CONCLUSIONS

In this paper we have explored the real-time performance of a Markov-switching model applied to real GDP and employment data for replicating the NBER business cycle chronology over the past 40 years. The model produces business cycle peak and trough dates that are fairly close to the NBER dates, using only information that would have been available at the time the dates were initially established. An important feature of the model is that it generally determines turning-point dates more quickly than the NBER Business Cycle Dating Committee. This timing advantage can be large, especially for business cycle troughs. It accomplishes this performance with a minimum of "false positive" business cycle peak or trough dates over the 40-year period.

Overall, the evidence presented above suggests that a statistical regime-switching model, such as the one used in this paper, could be a useful supplement to the NBER Business Cycle Dating Committee for establishing turning point dates. It appears to capture the features of the NBER chronology accurately and swiftly; furthermore, the method is transparent and consistent. It would be interesting to evaluate the real-time performance of multivariate switching models that incorporate another feature of NBER recessions, comovement across many economic variables over the business cycle, to see whether additional improvements can be made. We leave this for future research.

REFERENCES

Albert, James H. and Chib, Siddhartha. "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts." *Journal of Business and Economic Statistics*, January 1993, *11*(1), pp. 1-15.

Beaudry, Paul and Koop, Gary. "Do Recessions Permanently Change Output?" *Journal of Monetary Economics*, April 1993, *31*(12), pp. 149-63.

Boldin, Michael D. "Dating Turning Points in the Business Cycle." *Journal of Business*, 1994, *67*(1), pp. 97-130.

Burns, Arthur F. and Mitchell, Wesley E. *Measuring Business Cycles*. New York: National Buruea of Economic Research, 1946.

Chauvet, Marcelle. "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching." *International Economic Review*, November 1998, *39*(4), pp. 969-96.

Croushore, Dean and Stark, Tom. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics*, November 2001, *105*(1), pp. 111-30.

Diebold, Francis X. and Rudebusch, Glenn D. "Measuring Business Cycles: A Modern Perspective." *The Review of Economics and Statistics*, February 1996, *78*(1), pp. 67-77.

Friedman, Milton. *Monetary Studies of the National Bureau, the National Bureau Enters its 45th Year, 44th Annual Report*. New York: National Bureau of Economic Research, 1964, pp. 7-25. Reprinted in Friedman, Milton. *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine, 1969.

_____. "The 'Plucking Model' of Business Fluctuations Revisited." *Economic Inquiry*, April 1993, *31*(2), pp. 171-77.

Gavin, William T. and Kliesen, Kevin L. "Unemployment Insurance Claims and Economic Activity." Federal Reserve Bank of St. Louis *Review*, May/June 2002, *84*(3), pp. 15-28.

Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, March 1989, *57*(2), pp. 357-84.

_____. *Time Series Analysis*. Princeton University, 1994.

Hansen, Bruce E. "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov-Switching Model of GNP." *Journal of Applied Econometrics*, October-December 1992, *7*(Suppl 0), pp. S61-S82.

Kim, Chang-Jin; Morley, James and Piger, Jeremy. "Nonlinearity and the Permanent Effects of Recessions." Working Paper 2002-1014, Federal Reserve Bank of St. Louis, October 2002.

_____ and Nelson, Charles R. "Friedman's Plucking Model of Business Fluctuations: Tests and Estimates of Permanent and Transitory Components." *Journal of Money, Credit, and Banking*, August 1999, *31*(3), pp. 317-34.

_____ and Yoo, Ji-Sung. "New Index of Coincident Indicators: A Multivariate Markov Switching Factor Model Approach." *Journal of Monetary Economics*, December 1995, *36*(3), pp. 607-30.

Layton, Allan P. "Dating and Predicting Phase Changes in the U.S. Business Cycle." *International Journal of Forecasting*, September 1996, *12*(3), pp. 417-28.

McConnell, Margaret M. "Rethinking the Value of Initial Claims as a Forecasting Tool." Federal Reserve Bank of New York *Current Issues in Economics and Finance*, November 1998, *4*(11), pp. 1-6.

Moore, Geoffrey H. and Zarnowitz, Victor. "The Development and Role of the National Bureau of Economic Research's Business Cycle Chronologies," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*. Chicago: University of Chicago Press, 1986.

Sichel, Daniel. E. "Inventories and the Three Phases of the Business Cycle." *Journal of Business and Economic Statistics*, July 1994, *12*(3), pp. 269-77.