

The Phases of U.S. Monetary Policy: 1987 to 2001

Marvin Goodfriend

Inflation was relatively well behaved in the 1990s in comparison with preceding decades, yet Federal Reserve monetary policy was no less challenging. The Fed took painful actions in the late 1970s and early 1980s to reverse rising inflation and bring it down, and inflation fell from over 10 percent to around 4 percent by the mid-1980s. The worst economic ills stemming from high and unstable inflation were put behind us. Yet central bankers and monetary economists recognized that more disinflation was needed to achieve price stability. The transition to price stability was expected to be comparatively straightforward. Monetary policy promised to become more routine. Although the 1990s saw the longest cyclical expansion in U.S. history, the promised tranquility did not materialize. In many ways the period to be chronicled here proved to be about as difficult for monetary policy as the preceding inflationary period.

My account of Fed monetary policy divides the period since 1987 into six distinct phases. This division is natural because in each phase the Fed was confronted with a different policy problem. Phase 1 begins with rising inflation in the aftermath of the October 1987 stock market crash and ends with the start of the Gulf War in August 1990. Phase 2 covers the 1990–1991 recession, the slow recovery, and the disinflation to the end of 1993. Phase 3 tells the story of the Fed’s preemptive tightening against inflation in 1994–1995. Phase 4 deals with the long boom to 1999, the near full credibility

The author is Senior Vice President and Policy Advisor. This article was originally produced for a Festschrift conference in honor of Charles Goodhart at the Bank of England, 15–16 November 2001, and is to be published as a chapter “The phases of U.S. monetary policy: 1987 to 2001” in *Central Banking, Monetary Theory and Practice: Essays in Honour of Charles Goodhart*, Volume 1, edited by Paul Mizen (University of Nottingham, England) and published by Edward Elgar Publishing Ltd., Cheltenham, England. It was also presented at the National Bank of Hungary, Budapest, Hungary, and at the National Conference for Teachers of Advanced Placement Economics, Federal Reserve Bank of Richmond, October 2002. The author would like to thank discussant Charles Bean for excellent comments and Sandra Baker for excellent research assistance. The views expressed are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

for low inflation, and rising trend productivity growth. Phase 5 addresses the tightening of monetary policy to slow the growth of aggregate demand in 1999 and 2000. The sixth phase chronicles the collapse of investment in late 2000 and the recession in 2001.

The article presents a relatively compact account of the interaction between interest rate policy and the economy since 1987. It provides the minimum of descriptive detail needed to understand monetary policy during the period. The situations that confronted the Fed were remarkably varied. Nevertheless, the Fed's policy actions can be understood and interpreted as supporting the primary objectives of monetary policy, which were the same throughout. First of all, the Fed aimed to achieve and maintain credibility for low inflation. Second, the Fed managed interest rate policy so that the economy could attain the full benefits of rising trend productivity growth. Third, the alleviation of financial market distress dictated interest rate policy actions on occasion. Fourth, the Fed steered real short-term interest rates sharply lower when economic stimulus was needed. The story of how monetary policy pursued these objectives follows.

1. OCTOBER 1987–JULY 1990: RISING INFLATION AND THE STOCK MARKET CRASH

From Wednesday, 14 October 1987, through the close of trading on Monday, 19 October, the Dow Jones Industrial Average lost about 30 percent of its value. On Monday alone, the Dow lost 23 percent. Not since October 1929, when the Dow lost around 25 percent in two consecutive days, had a sudden collapse of equity values been so great.¹

The Fed responded to the October 1987 stock market crash in a number of ways. For our purposes, its most important responses were these. The Fed accommodated the increased demand for currency and bank reserves with extensive open market purchases. It also dropped its federal funds rate target from around 7.5 percent to about 6.75 percent.

Central bankers now know that sufficiently stimulative monetary policy might well have averted the deflation and depression of the 1930s. The Fed made sure that monetary policy was sufficiently stimulative to avert another catastrophe. The Fed was concerned about the resulting risks to price stability, noting that its actions should not be seen as inflationary.²

As it turned out, inflation rose in 1988, 1989, and 1990 in spite of the fact that the Fed had put the economy through a severe recession in the early 1980s to restore price stability. Core CPI inflation rose from around 3.8 percent in 1986 to 5.3 percent in 1990. Employment cost inflation rose from around

¹ This paragraph is heavily paraphrased from the Brady Report (1988, 1).

² Greenspan (1988, 218).

3 percent in 1986 to over 5 percent in 1989, even as productivity growth averaged less than 1 percent from 1986 to 1990. The unemployment rate fell from around 7 percent in 1986 to 5.3 percent in 1989. Annual average unemployment below 5.5 percent had not been seen since 1973.

Part of the problem was that inflationary pressures began to build well before October 1987. Rising inflation expectations were already evident in the 30-year bond rate, which rose by 2 full percentage points from around 7.5 percent to 9.6 percent between March and October of 1987.³ Surprisingly, the Fed reacted relatively little to the 1987 inflation scare. The Fed's failure to respond created doubts that it would hold the line on inflation, much less push on to price stability. The bond rate did not fall back to the 7.5 percent range until late 1992, reflecting the slow restoration of credibility for low inflation that was lost in the second half of the 1980s.

In short, by mid-1987 there was sufficient reason for the Fed to tighten policy preemptively against inflation. And the Fed raised the discount rate from 5.5 percent to 6 percent in September soon after Alan Greenspan replaced Paul Volcker as Fed Chairman.⁴ But the October stock market crash intervened before policy could be tightened further.

All in all, it seems fair to say that monetary policy restraint was delayed by a couple of years because the Fed was reluctant to act against inflation both before and after the crash of October 1987. By the time the Fed felt it was safe to tighten monetary policy further, it needed to counteract inflationary forces that were already well entrenched. As had been the case in the inflationary go/stop era, the restoration of credibility for low inflation after it was compromised required the Fed to raise real short rates higher than otherwise, with a greater risk of recession.

Beginning in the spring of 1988, the Fed began to raise the funds rate from the 6 to 7 percent range to nearly 10 percent in March 1989. With core CPI inflation then running at about 4.5 percent, that sequence of policy actions increased real short rates by over 3 percentage points to more than 5 percent. Real GDP growth slowed from about 4 percent in 1988 to 2.5 percent in 1989. In response the Fed dropped the funds rate to around 7 percent by late 1990. However, by then core CPI inflation was running at 5.3 percent, well above its mid-1980s average.

³ Ireland (1996) shows quantitatively why a significant change in the long bond rate is likely to represent a change in inflation expectations rather than a change in the expected real rate. Goodfriend (1993) gives an account of inflation scares in the bond market during the 1980s.

⁴ The 1987 inflation scare may have reflected doubts about the credibility of Volcker's unknown successor.

2. AUGUST 1990–JANUARY 1994: WAR, RECESSION, AND DISINFLATION

The August 1990 Gulf War dealt a severe blow to the U.S. economy. It would take until March 1991 for U.S. ground forces to eject Iraqi troops from Kuwait and stabilize the region. The ground war went as well as could have been expected, lasting only 100 hours. But the outcome appeared to be in doubt until a few hours before the war was won. Consequently, uncertainty greatly affected the economy for nearly eight months. In August 1990 oil prices quickly spiked up from about \$15 per barrel to over \$35, falling back only gradually by early 1991. Households and businesses showed an inclination to postpone spending until the outcome of the war became clear. These supply and demand shocks caused economic activity to contract in the fall of 1990 through the first quarter of 1991. The National Bureau of Economic Research dates the 1990 recession from July 1990 to the trough in March 1991.

Monetary policy could do little to avert a recession during the Gulf War. Policy actions take time to act on the economy. Moreover, the war occurred at a time when the Fed's credibility for low inflation had been compromised. As mentioned above, core CPI inflation rose from 3.8 percent in 1986 to 5.3 percent in 1990. And the Fed risked an inflation scare in the bond market if it cut the federal funds rate too sharply. Even so, the Fed brought the federal funds rate down from just above 8 percent at the start of the Gulf War to just under 6 percent at its close in the spring of 1991.

As a result of the restrictive policy actions undertaken by the Fed prior to the Gulf War and the war-related recession itself, inflation began to recede. Core CPI inflation decreased to 4.4 percent in 1991. The recovery from the recession trough in March 1991 proved to be slow in part because the recession itself was mild. The unemployment rate rose only a little more than 1 percentage point during the recession itself, from 5.5 percent in July 1990 to 6.8 percent in March 1991. Even though real GDP growth snapped back to 4 percent in 1992 from 0.8 percent growth in 1991, the unemployment rate continued to climb, peaking at 7.8 percent in June 1992. This was known as the "jobless recovery."

The Fed reacted by steadily reducing the federal funds rate from 6 percent in mid-1991, to 4 percent by the end of 1991, to 3 percent by October 1992, where it stayed until February 1994. Inflation fell as well, to around 3 percent by 1992. The nominal federal funds rate cut partly reflected the 1 1/2 percentage point fall in inflation and partly represented a 1 1/2 percentage point cut in the real federal funds rate, bringing the real rate to approximately zero.

Four factors account for the highly stimulative policy stance. First, the high and rising unemployment rate was a concern. Second, the banking system was undercapitalized in many areas of the country. Bank loans were expensive and somewhat more restricted than usual. Third, inflation had been brought down to around 3 percent, 2 percentage points below where it was in 1990, and

about 1 percentage point below where it had been in the mid-1980s. Fourth, the gains against inflation restored the Fed's credibility enough that it could comfortably risk moving to a zero real federal funds rate to stimulate aggregate demand and job growth.

The zero real short rate remained in place for about 18 months, until February 1994. During that time the unemployment rate fell from 7.8 percent to 6.6 percent. The inflation rate fell slightly. The long bond rate fell from around 7.5 percent in October 1992 to around 6 percent at the end of 1993. The lower bond rate may have been the result of a weak economic expansion and progress against the Federal budget deficit made at the time. The bond rate also probably reflected the acquisition of credibility for low inflation won by the Fed as a consequence of disinflationary policy actions taken since 1988.

3. FEBRUARY 1994–FEBRUARY 1995: PREEMPTING RISING INFLATION

The economic expansion gathered strength in late 1993. The zero real federal funds rate was no longer needed and would become inflationary if left in place. The Fed began to raise the federal funds rate in February 1994, taking it in seven steps from 3 percent to 6 percent by February 1995. Inflation showed little tendency to accelerate and remained between 2.5 percent and 3 percent. Thus, the Fed's policy actions took the real federal funds rate from zero to a little more than 3 percent. The move raised real short-term interest rates to a range that could be considered neutral to mildly restrictive. In spite of the policy tightening, real GDP grew by 4 percent in 1994, up from 2.6 percent in 1993, and the unemployment rate fell from 6.6 percent to 5.6 percent from January to December 1994.

The policy tightening in 1994 succeeded in its main purpose: to hold the line on inflation without creating unemployment. The unemployment rate moved up only slightly to 5.8 percent in April 1995 and then began to fall again. The 1994 tightening demonstrated that a well-timed preemptive increase in real short-term interest rates is nothing to be feared. In this case, it was needed to slow the growth of aggregate demand relative to aggregate supply to avert a build up of inflationary pressures. By holding the line on inflation in 1994, preemptive policy actions laid the foundation for the boom that followed.

Preemptive policy in 1994 was motivated in part by the large increase in the bond rate beginning in October 1993. Starting from a low of 5.9 percent, the 30-year bond rate rose through 1994 to peak at 8.2 percent just before election day in November. The nearly 2 1/2 percentage point increase in the bond rate indicated that the Fed's credibility for low inflation was far from

secure in 1994. By January 1996 the bond rate had returned to around 6 percent and journalists were talking about the “death of inflation.”⁵

Talk of the death of inflation was reassuring. It indicated that the Fed’s preemptive actions had anchored inflation and inflation expectations more securely than ever before. This helps to explain why later in the decade the unemployment rate could fall to 4 percent with little inflationary wage and price pressure. However, the death-of-inflation talk was also disappointing because it tended to undervalue the role played by the Fed in “killing” inflation. The actions taken in 1994 were a textbook example of a successful preemptive campaign against inflation. It is discouraging that even then, the public should misunderstand the crucial role played by the central bank in containing inflation. If inflation is to be contained permanently, the idea that inflation doesn’t just “die” but must be periodically vanquished by proactive interest rate policy is one that the public must appreciate more fully.

The preemptive tightening in 1994 was difficult for the Fed even though it was clearly needed. Beginning with the 25 basis point increase in the federal funds rate in February 1994, the Fed started to announce its current intended federal funds rate target immediately after each FOMC meeting. This new practice made Fed policy more visible than ever. Every increase in the federal funds rate target since then has attracted considerable attention.

Transparency of the Fed’s interest rate target is a good thing because it improves the public’s understanding of monetary policy. However, since 1994 the Fed has operated with a transparent federal funds rate target and somewhat opaque medium- and longer-term goals.⁶ The Federal Reserve Act does not specify how the Fed is to balance medium- or longer-term objectives for inflation, economic growth, and employment. And the Fed does not clarify its medium- or long-term objectives as well as it could. Its interest rate policy actions are scrutinized more than they would be if the Fed were more forthcoming about its objectives.

Part of the problem is that the Fed is naturally unwilling to specify its objectives more clearly without direction from Congress. And Congress has been unable to agree on a mandate for the Fed that would result in clarification. The Fed has been operating without a clear mandate from Congress since the collapse of the gold standard and the Bretton Woods fixed exchange rate system in 1973. Under these circumstances, announcing the federal funds rate target increases the potential for counterproductive disputes between Congress and the Fed.

One such dispute broke into the open in 1994 when Congress objected to the Fed’s preemptive increase in interest rates and took the unprecedented step of inviting all 12 Reserve Bank presidents to explain their views before

⁵ See, for instance, Bootle (1996).

⁶ See Broaddus (2001).

the House and Senate banking committees. Legislation that would remove the presidents from the FOMC was considered at the time on the grounds that the presidents were thought to favor excessively tight monetary policy. The net effect of this very public dispute was to create doubt about the Fed's ability and willingness to take the tightening actions necessary to hold the line on inflation. The public dispute between the Fed and Congress probably contributed to the severity of the 1994 inflation scare in the bond market.

4. JANUARY 1996–MAY 1999: THE LONG BOOM

In many ways managing interest rate policy was more difficult in the last half of the 1990s than in the first half. Two major factors complicated interest rate policy in the period from 1996 to 1999. First, the Fed had to learn to operate with near full credibility for low inflation, credibility it had secured with its successful preemptive policy actions in 1994. Second, the Fed had to deal with rising productivity growth. Both complications benefited the economy greatly. The Fed worked for almost two decades to achieve price stability. Economists had long hoped that advances in computer and information technology would bring an end to the productivity slowdown dating from the mid-1970s. Nevertheless, both developments challenged monetary policy in ways that were not anticipated. This section reviews the developments themselves and points out their complications for monetary policy. It concludes with an assessment of interest rate policy actions taken by the Fed during the period.

Near Full Credibility for Low Inflation

When near full credibility for low inflation is newly won, both the central bank and the public tend to overestimate the economy's noninflationary potential output. In other words, both are inclined to be fooled by the central bank's credibility for low inflation in a way that restrains interest rate policy actions that may be necessary to sustain that very credibility. Even if inflation and inflation expectations remain firmly anchored, there is a risk that interest rate policy actions will be insufficient to head off an unsustainable real boom followed by a painful period of adjustment.⁷ The nature of this risk is detailed below with reference to the long boom from 1996 to 1999.

When credibility for low inflation is secure, labor markets can get surprisingly tight without triggering inflationary wage pressures. Workers are less inclined to demand inflationary nominal wage increases because they have confidence that firms will not push product prices up. And firms are more inclined to hold the line on product price increases even if labor costs begin to

⁷ Goodfriend (2001) and Taylor (2000) explore this sort of logic.

rise. Firms and workers have confidence that any excess of aggregate demand over potential output will be temporary, reversed by sufficiently restrictive subsequent interest rate policy actions. Confidence in the central bank can enable the economy to operate above potential output for a while with little or no increase in inflation.

With inflation and inflation expectations firmly anchored, a central bank will be more inclined to delay monetary tightening when the economy moves above its presumed noninflationary potential level of output. It could take more time to discern whether an excess of aggregate demand is temporary or persistent before it responds with tighter monetary policy. When there is evidence of a rising trend in productivity growth, a central bank could explore the possibility that faster growth of aggregate demand might be accommodated without inflation.

The Fed's very success in anchoring inflation and inflation expectations meant that traditional indicators of excessive monetary stimulus became less reliable.⁸ Inflation as measured by the core CPI ranged between 2 percent and 3 percent for the remainder of the decade. Price stability was maintained even though real GDP grew at around 4.4 percent per year from 1996 through 1999, and the unemployment rate fell from 5.6 percent in January 1996 to 4 percent, a rate not seen since 1970.

Clearly, near full credibility for low inflation helped the economy to operate well beyond a level that would have created concerns about inflation in the past. Real indicators of incipient inflation such as the unemployment rate became less useful as guides for interest rate policy. Moreover, the bond market was less inclined to exhibit inflation scares. After having peaked at 8.2 percent in late 1994, the 30-year bond rate returned to levels below 7 percent and moved in a range between 5 percent and 6 percent in the last two years of the decade. This development recalled the bond market of the late 1960s, which was confident that inflation would remain low even after economic activity moved above what was then considered its noninflationary potential.

Nominal money growth also became less reliable as an indicator of inflation. Growth temporarily in excess of historical standards might be needed to accommodate an increased demand for money due to lower nominal interest rates and growing confidence in the stability of the purchasing power of money. Even truly excessive money growth might not cause inflation if the public believed that the Fed would tighten policy to reverse inflationary money creation before too long.⁹

If the public comes to think that the economy has become "structurally" less prone to inflation, i.e., that "inflation is dead," then the risk of an unsustainable boom increases still further. Excessive optimism encourages households

⁸ These informational problems add to the real-time data problems analyzed in Orphanides (2001).

⁹ Taylor (2000) emphasizes this possibility.

and firms to expect unrealistically high future real income prospects, triggering an unsustainable spending binge. Spending is encouraged further if the central bank appears to buy into the optimism by not raising interest rates as aggregate demand accelerates. Excessively optimistic expectations for the economy's productive potential would be reflected in a run-up in equity prices, real estate values, and asset prices in general. The risk of precipitating a collapse of asset prices would in turn make a central bank more cautious than otherwise in tightening interest rate policy.

Rising Productivity Growth

From 1986Q1 until 1990Q4 nonfarm business productivity growth averaged only 0.8 percent per year, reflecting the ongoing slowdown in productivity growth that began in the mid-1970s. In the next five years productivity growth rose to 1.7 percent per year, and from 1996Q1 to 2000Q4 productivity grew on average by 2.4 percent per year. In other words, productivity growth tripled over this 15-year period. In the late 1990s it was possible to argue that the burst of productivity growth was only temporary and would soon fall back to 2 percent or less. But it was just as reasonable to argue that productivity growth would move even higher for a while as the economy continued to find new ways to employ advances in communications and information technology.

The trend productivity growth rate has enormous implications for standards of living, for perceived lifetime income prospects, and for current spending. When productivity grows at 1 percent a year, national per capita product doubles roughly every 70 years. If productivity grows at 2 percent per year, then per capita product doubles in 35 years and quadruples every 70 years. Sustained 3 percent productivity growth would double per capita income in 23 years, quadruple it in 46 years, and result in an eightfold increase in around 70 years. This last possibility seems unlikely; but sustained productivity growth between 2 percent and 2.5 percent per year well into the 21st century would match the 2.3 percent average productivity growth rate that the United States sustained between 1890 and 1970.¹⁰ These figures indicate the tremendous long-term potential that many saw in the U.S. economy in the last half of the 1990s—and still see in spite of the 2001 recession.

Real wages began to rise during the 1990s after stagnating during the productivity slowdown period. Households could count on the fact that throughout U.S. history, per capita productivity growth was transmitted to real wage growth as firms competed for ever more productive labor. Firm profits and equity values would benefit initially from the installation of more productive technology. But as the installation of that technology became widespread,

¹⁰ See Romer (1989, 56).

firms would be forced to pay up for the more productive labor. Thus, the profit share of national income rose during the 1990s, but it could be expected to return to historic norms once real wages caught up. Whether the increase in income took the form of rising profits or wages, its underlying source was the rising trend in productivity growth.

In short, the period from 1996 to 1999 was characterized by an optimism about future income prospects. This optimism gave rise to an expansion in investment and productive capacity by firms matched by an increasing willingness of households to absorb the output that the growth of productive potential made possible.

Rising productivity growth had two critical implications for monetary policy. First, rising productivity growth reinforced the perception that the economy was inflation-proof and provided an argument against more restrictive monetary policy. For a while, rising productivity growth more than offset the rising nominal wage growth associated with tight labor markets. The problem for monetary policy was that trend productivity growth was not likely to rise much above 2.5 percent or 3 percent per year. And productivity was already growing in that range by 1998. There was less reason to think that nominal wage growth would stop rising if labor markets remained as tight as they became during the period. Rising productivity growth might hold unit labor costs and inflation down for a while, but at some point unit labor costs would begin to rise, necessitating tighter monetary policy.

Second, although *rising* productivity growth made the economy more inflation-proof in the short run, *higher* trend productivity growth would require higher real interest rates in the long run. The reason is this. At initial real interest rates, households are inclined to borrow against their improved future-income prospects to spend some of the proceeds today. Also, firms are inclined to invest more in plant and equipment to profit from improved productivity. In the aggregate, however, households and firms cannot bring goods and services from the future into the present because the future productivity growth has not yet arrived.¹¹ In such circumstances, firms accommodate the growth in aggregate demand in excess of current productivity growth by hiring more labor to meet the demand.¹² Labor markets become increasingly tight, and the economy overshoots even its faster sustainable growth path.

To enable the economy to grow faster without inflation, the central bank would have to maintain higher short-term real rates on average over time to make households and firms sufficiently patient to defer their spending to the future. Higher short- and long-term real rates bring aggregate demand down

¹¹ In part, the United States satisfied its demand for goods and services in excess of current output by importing capital from abroad (where growth prospects were not as bright) and running a current account deficit.

¹² See Goodfriend and King (1997) for a discussion of the macromodel underlying the analysis here and elsewhere in the article.

to potential output so that both can grow together and the employment rate is neither expanding nor contracting over time. In short, when an economy enjoys an increase in the rate at which productivity can grow over the long run, it requires permanently higher real interest rates on average to offset the inclination of the public to spend the proceeds prematurely.¹³

The problem for U.S. monetary policy during the period from 1996 to 1999 was to ascertain the timing and magnitude of the increase in real interest rates necessary to allow the economy to transition to a higher growth path without creating imbalances in labor utilization that could lead to an outbreak of inflation or an unsustainable expansion of real activity. This policy problem was particularly formidable because it had to be solved even as near full credibility for low inflation and rising productivity growth made the economy appear to be more inflation-proof than ever.

Interest Rate Policy 1996–1999

The Fed changed its federal funds rate target relatively little from January 1996 through June 1999. The funds rate was held at 5.25 percent for over a year from January 1996 until March 1997, when it was raised to 5.5 percent. The funds rate was then held constant for another 18 months at 5.5 percent until the fall of 1998, when it was cut by 75 basis points in three 25 basis point steps in September, October, and November in the aftermath of the Russian debt default. Core CPI inflation averaged between 2 percent and 2.5 percent during the entire period, so the real short rate was around 3 percent, except when it was lowered by 75 basis points in the fall of 1998.

The single 25 basis point adjustment in March 1997 was made as the economic expansion gathered momentum. By moving in March 1997, the Fed signaled that it was poised to act if necessary to restrain inflationary growth. However, the Fed declined to raise interest rates further for two years because two world financial crises intervened: the 1997 financial crisis in East Asia and the 1998 financial crisis following the Russian default. Alleviating financial market distress became a primary focus of monetary policy in each case.

The Fed did not actually cut its funds rate target in the second half of 1997 in response to the East Asian crisis, but it probably deferred tightening policy. The 75 basis point cut in the funds rate following the Russian default moved short-term interest rates in the opposite direction from that which would ultimately be needed to stabilize the U.S. economy. As was the case in the aftermath of the October 1987 stock market crash, the two financial crises in

¹³ For log utility, the real interest rate must rise by the increase in the productivity growth rate.

1997 and 1998 helped to delay a necessary policy tightening by as much as two years.

However, my reading of the forces acting on monetary policy during the boom—near full credibility and rising productivity growth—suggests that even without the two financial crises, the Federal Reserve would have been reluctant to tighten monetary policy very much between 1996 and 1999. Not only was inflation under control, but there was great uncertainty about the magnitude and timing of the interest rate policy actions needed to enable the economy to transition to a higher growth path without inflation. Under the circumstances, the Fed chose to wait before tightening very much until the need for restrictive policy became more obvious.¹⁴

5. JUNE 1999–DECEMBER 2000: RESTRAINING THE GROWTH OF DEMAND

By the second half of 1999, the pool of available workers—unemployed plus discouraged workers—looked to be approaching an irreducible minimum, and the growth of aggregate demand in excess of plausible potential GDP tightened labor markets further. If real interest rates were kept too low, then the expansion would end in one of two ways. The Fed could lose its credibility for low inflation and the expansion would end as it had so often in previous decades, with rising inflation, an inflation scare in bond markets, and a policy tightening sufficient to restore credibility for low inflation. Alternatively, if the Fed's near full credibility for low inflation held fast, then rising unit labor costs would result in a profit squeeze, lower equity values, a collapse in investment, and slower growth of consumer spending.

Real GDP grew by a spectacular 4.7 percent and 8.3 percent in Q3 and Q4 of 1999, and the unemployment rate drifted down from 4.3 percent in early 1999 to 4 percent by the end of the year. The extraordinary growth of aggregate demand outstripped even the high accompanying productivity growth rates of 3 percent and 7.4 percent, respectively.

Clearly, real short rates needed to move up further. The Fed reversed the 75 basis point easing of policy it had undertaken the previous autumn with three 25 basis point steps in June, August, and November of 1999. It also raised its federal funds rate target by another percentage point between November 1999 and May 2000 to 6.5 percent, where it was held until January 2001.

With core CPI inflation running at about 2.5 percent, real short rates were roughly 4 percent. By comparison with other occasions of concerted monetary

¹⁴ A new literature considers whether asset prices should help guide interest rate policy. See Bernanke and Gertler (2000), Borio and Lowe (2001), Cecchetti et al. (2000), Goodfriend (2002), and Goodhart (1995).

tightening, the real interest rate was not then particularly high. In part, this was due to the fact that the Fed had not yet lost credibility for low inflation and so did not need higher real rates to bring inflation down. The 4 percent real rate seemed to be enough as real GDP growth in 2000Q1 slowed from the previous quarter by 6 percentage points, to 2.3 percent. However, real growth accelerated again to 5.7 percent in 2000Q2 and the Fed stayed with its 6.5 percent funds rate target. Real GDP growth in Q3 again slowed, to 1.3 percent, but the Fed needed another quarter of evidence that the slowdown would be sustained. That confirmation was received in late 2000 and early 2001, when it became clear that real GDP grew by around 2 percent in 2000Q4.

6. JANUARY 2001–PRESENT: THE COLLAPSE OF INVESTMENT AND THE 2001 RECESSION

The problem for monetary policy in 2001 was that real GDP growth failed to find a bottom and continued to fall, from 1.3 percent in Q1, to 0.3 percent in Q2, to –1.3 percent in Q3. Personal consumption expenditure growth held up better, slowing from 3 percent, to 2.5 percent, and to 1 percent, respectively, in the first three quarters of 2001. In part, consumer spending held up reasonably well because the unemployment rate rose relatively slowly from a very low 4 percent at the end of 2000 to 4.6 percent by July 2001. The comparatively tight labor market continued to provide a sense of job security and robust real wage growth that supported consumer confidence.

The primary drag on growth in 2001 came from nonresidential fixed investment and inventory liquidation. Investment in equipment and software grew much faster than GDP during the boom years. Advances in information processing and communication technologies led investment in equipment and software to rise from about 6 percent of real GDP in 1990 to a peak of around 12 percent of real GDP in mid-2000. Real nonresidential (business) fixed investment, which includes nonresidential structures as well as equipment and software, grew at around 10 percent per year from 1995 until 2000. Growth in business investment collapsed to near zero in 2000Q4 and 2001Q1 and then contracted at more than a 10 percent annual rate in Q2 and Q3 of 2001.

The swing in inventory accumulation compounded the growth slowdown in 2001. After accumulating at an annual rate of \$79 billion, \$52 billion, and \$43 billion dollars in Q2, Q3, and Q4 of 2000, inventories were liquidated at an annual rate of \$27 billion, \$38 billion, and \$62 billion in the first three quarters of 2001, respectively.¹⁵

The developments outlined above reflect the fact that the economy overshoot its sustainable growth rate in the late 1990s. Much capacity put in place

¹⁵ GDP is around \$10 trillion, so \$100 billion is about 1 percent of U.S. GDP.

during the boom began to look excessive once the growth rate slowed. Higher trend productivity growth would eventually enable the economy to absorb that capacity, but not as soon as had been believed. Moreover, rising unemployment in the manufacturing sector caused a secondary collapse of demand that threatened to spill over to the services sector. The rising unemployment rate caused consumers throughout the economy to become more cautious, weakening aggregate demand further. This, in turn, gave businesses an additional reason to put investment plans on hold.

Financial factors significantly amplified the overshooting in investment and the painful adjustment thereafter. Excessive equity values cheapened equity finance during the boom years, and the collapse of equity values raised the cost of equity finance during the slowdown. Likewise, high net worth facilitated external debt finance during the boom, and the loss of net worth raised the cost of external debt finance thereafter. Moreover, investment could be financed readily with internally generated funds during the boom, but the decline of profits during the slowdown caused firms to become more reliant on external finance even as it became more costly.

Recognizing the contractionary forces described above, the Fed cut its federal funds rate target in 11 steps from 6.5 percent at the beginning of 2001 to 1.75 percent in December 2001. Core CPI inflation did not change much during the year, so the policy actions translated into a 4 3/4 percentage point cut in real short-term interest rates. This was a relatively large reduction in the real federal funds rate in so short a time by historical standards, though not when one considers that real GDP grew at around 5.25 percent in the year through 2000Q2 and grew at less than 1 percent in 2001. Real short rates were then negative according to the core CPI inflation rate, which was running at about 2.5 percent. The Fed was able to cut the real federal funds rate so far without precipitating an inflation scare because of the near full credibility for low inflation.

The 11 September 2001 destruction of the World Trade Center in New York made matters worse. Data for October indicated a sharp drop in consumer confidence, and a further contraction in the manufacturing sector. Most striking, roughly 800,000 jobs were lost in October and November combined. The rise in the unemployment rate in September, October, and November was the fastest three-month increase since 1982, bringing the cumulative rise since January to about 1 3/4 percentage points. In November 2001 the National Bureau of Economic Research officially declared that the United States had been in a recession since March.

The big jump in the unemployment rate had the potential to undermine consumer confidence. The unemployment rate in the United States rose sharply by at least 2 percentage points on five occasions since 1960: during and following the recessions of 1960–1961, 1969–1970, 1973–1975, 1981–1982, and 1990–1991. The cumulative rise during the 1980 recession was just 1.5

percentage points.¹⁶ The unemployment rate rose by 4.2 percentage points in 1973–1975, and by 3.6 percentage points in 1981–1982.

The big jump in unemployment carried a second risk: historically, sharply rising unemployment has been associated with falling inflation. For instance, when the unemployment rate rose by 3.6 percentage points in 1981–1982, the inflation rate fell by around 6 percentage points. Disinflation was beneficial when inflation was too high. When inflation was too high, the Fed had the leeway to cut its nominal federal funds rate target to keep the real federal funds rate from rising as the disinflation ran its course. In 2001, the Fed had only 1 3/4 percentage points of leeway before the nominal federal funds rate would hit the zero bound.

That said, there were three reasons to think that disinflation would be relatively mild this time. First, the unemployment rate might not rise much more since the Fed had already cut the real funds rate by 4 3/4 percentage points. Second, slower wage growth due to slack in the labor market might be matched by slower productivity growth. If that were the case, then unit labor costs would not fall much and there would be little downward pressure on prices. Third, the earlier recessions were set off in large part by tighter monetary policy aimed at reducing inflation. This time the Fed was not trying to bring the inflation rate down.

No one can say how the latest situation confronting U.S. monetary policy will turn out. The zero bound may yet become a problem. Hopefully, aggressive interest rate actions undertaken in 2001 have laid the foundation for a full recovery in 2002. In that regard, it is worth noting that the federal funds rate futures market believed at the end of 2001 that the funds rate had hit bottom and that the Fed would raise interest rates as the economy recovered in 2002.

7. CONCLUSION

The challenges facing monetary policy since 1987 have been surprisingly varied. Rising inflation was a problem for the Fed only briefly during the period. Restrictive monetary policy in the late 1980s and early 1990s reversed the rising inflation trend, and preemptive policy actions in 1994 secured near full credibility for low inflation in the late 1990s. The Fed dealt with three major financial crises: the October 1987 crash, the 1997 East Asian crisis, and the consequences of the Russian debt default in 1998. Monetary policy reacted to two wars: the 1990–1991 Gulf War and the 2001 War on Terrorism. The Fed became more transparent by regularly announcing its current federal funds rate target beginning in February 1994. Most importantly, monetary policy adapted to an environment in which the Fed acquired near full credibility for

¹⁶ The 1980 recession was associated with the brief imposition of credit controls. See Schreft (1990).

low inflation, and the Fed navigated a difficult transition toward higher trend productivity growth. Because the problems were so varied, it is difficult to draw overall lessons from the period, but one thing is clear. Similar challenges are likely to be encountered in the future and the experience gained in surmounting them should help the Fed improve monetary policy.

REFERENCES

- Bernanke, Ben, and Mark Gertler. 1999. "Monetary Policy and Asset Price Volatility." In *New Challenges for Monetary Policy*. Kansas City: Federal Reserve Bank of Kansas City: 77–128.
- Bootle, Roger. 1996. *The Death of Inflation*. London: Nicholas Brealey Publishing.
- Borio, Caudio, and Philip Lowe. 2001. "Asset Prices, Financial and Monetary Stability: Exploring the Nexus." Bank for International Settlements.
- Brady Report. 1988. Report of the Presidential Task Force on Market Mechanisms. Washington: Government Printing Office.
- Broadus, J. Alfred, Jr. 2001. "Transparency in the Practice of Monetary Policy." Federal Reserve Bank of Richmond *Economic Quarterly*. (Summer): 1–10. Also forthcoming in 26th Annual Economic Policy Conference, Federal Reserve Bank of St. Louis.
- Cecchetti, Stephen, Hans Genberg, John Lipsky, and Sushil Wadhvani. 2000. *Asset Prices and Central Bank Policy*. The Geneva Report on the World Economy, Number 2.
- Goodfriend, Marvin. 1993. "Interest Rate Policy and the Inflation Scare Problem: 1979–1992." Federal Reserve Bank of Richmond *Economic Quarterly* 79 (Winter): 1–24.
- _____. 2001. "Financial Stability, Deflation, and Monetary Policy." *Monetary and Economic Studies* (Special Edition), Japan: Bank of Japan: 143–66.
- _____. 2002. "Interest Rate Policy Should Not React Directly to Asset Prices." In *Asset Price Bubbles: Implications for Monetary, Regulatory, and International Policies*, ed. William C. Hunter, George G. Kaufman, and Michael Pomerleano. Conference volume, Federal Reserve Bank of Chicago and World Bank Group. Cambridge, Mass.: MIT Press (forthcoming).

- Goodfriend, Marvin, and Robert King. 1997. "The New Neoclassical Synthesis and the Role of Monetary Policy." In *NBER Macroeconomics Annual 1997*, ed. Ben Bernanke and Julio Rotemberg. Cambridge, Mass.: MIT Press: 231–82.
- Goodhart, C. A. E. 1995. "Price Stability and Financial Fragility." In *Financial Stability in a Changing Environment*, ed. Kuniho Sawamoto, Zenta Nakajima, and Hiroo Taguchi. Japan: Bank of Japan: 439–97.
- Greenspan, Alan. 1988. "Federal Reserve Response to the October Crisis." *Federal Reserve Bulletin* (April): 217–25.
- Ireland, Peter. 1996. "Long-Term Interest Rates and Inflation: A Fisherian Approach." Federal Reserve Bank of Richmond *Economic Quarterly* (Winter): 21–35.
- Orphanides, Athanasios. 2001. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91 (September): 964–85.
- Romer, Paul. 1989. "Capital Accumulation in the Theory of Long-Run Growth." In *Modern Business Cycle Theory*, ed. Robert Barro. Cambridge, Mass.: Harvard University Press: 51–127.
- Schreft, Stacey L. 1990. "Credit Controls: 1980." Federal Reserve Bank of Richmond *Economic Review* 76 (November/December): 25–55.
- Taylor, John. 2000. "Low Inflation, Pass Through, and the Pricing Power of Firms." *European Economic Review* 44: 1389–408.

Competition Among Bank Regulators

John A. Weinberg

The organization of bank regulation in the United States is somewhat peculiar. Banks answer to an array of regulators, both federal and state. To begin with, a bank can choose a national or a state charter. National banks are regulated by the Office of the Comptroller of the Currency (OCC). State banks are regulated by their home states, as well as by a federal regulator. The Federal Reserve System regulates state-chartered banks that are Federal Reserve members, and the Federal Deposit Insurance Corporation (FDIC) regulates state, nonmember banks. A bank, by its choice of charter and Federal Reserve membership, chooses its regulators. There is a sense, then, in which U.S. federal bank regulators are in competition with each other. How does this competition affect bank regulation in the United States? On the one hand, one might conclude that the need to compete with other agencies would motivate a regulator to perform its tasks as effectively and efficiently as possible. On the other hand, one might argue that the desire to attract more clients could drive a regulatory agency to be loose.

Banking is not the only industry in which alternative regulatory agencies compete with one another. Most other instances, however, involve different geographic jurisdictions. For instance, to the extent that environmental regulations vary from state to state, a manufacturer's decision on plant location entails a choice among potential regulators. The stringency of such regulations then has the potential to become one tool by which states compete to attract businesses. One could ask the same question about this competition as is often asked about the interaction among bank regulators. Does competition lead to effective or excessively loose environmental control?

■ This article has benefited from helpful comments from Kartik Athreya, Jeffrey Lacker, Yash Mehra, and Tom Humphrey. The views expressed herein are the author's and do not represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

When the effects of the regulated activity, polluting for instance, are predominantly local, geographic regulatory competition, as in the case of state-level environmental rules, is analogous to the jurisdictional competition studied by Tiebout (1956). Tiebout's direct concern was the provision of "local public goods" by local governments funded with local taxes in a setting with a mobile population. His conclusion was that competition in the joint setting of taxes and levels of public goods and services would lead to efficient levels of government expenditures. The same logic applies to local regulation of activities with local effects.

Bank regulation, however, does not have the same geographical limits as some environmental regulation. While state banks are regulated locally by state supervisory agencies, all banks have federal regulators. Further, a bank can change its federal regulator without having to relocate or make any other significant change in its activities. In this environment, does the Tiebout logic of beneficial competition still apply?

This article highlights how the effects of alternative regulatory structures depend on assumptions about such underlying factors as the regulators' objectives, and the way in which regulators' costs are financed. This point can best be made in the context of a model that captures important elements of bank and bank regulator activities. Section 2 presents such a model. The model's emphasis is on the role of bank examinations in assessing the quality of bank assets in the presence of deposit insurance. In the context of this model, an efficient regulatory policy is defined. Possible regulatory outcomes are then studied under alternative assumptions about regulators' preferences regarding banking industry performance and the extent to which deposit insurance and bank examination are integrated activities financed under a consolidated budget constraint. In some cases, regulatory competition leads to efficient policy choices, while in others competition results in inefficient outcomes. Notably, when the financing of regulation and deposit insurance is not integrated, competition among regulators can impose excessive costs on deposit insurance.

1. BACKGROUND

In discussions about rivalry among alternative bank regulators, a common concern is that regulators will "race to the bottom." Each regulator, it is argued, will want to attract as many banks into its constituency as possible. Further, this incentive to attract "client" banks will outweigh the regulators' interest in controlling bank risk-taking incentives. This so-called "competition in laxity" will result in excessive costs to the deposit insurance system. The possibility of a race to the bottom, as discussed by Scott (1977), has partly motivated a number of proposals for the consolidation of federal bank regulation.

The notion that competition might result in excessively lax or otherwise inefficient regulation is not unique to banking. In the general area of corporate governance and the market for corporate control, it has been argued that states compete to be corporations' charter locations by passing laws that inhibit corporate takeovers. Since incumbent managers make location decisions, they might be influenced by laws that protect their incumbency. Karpoff and Malatesta, for instance (1989), report evidence that supports this hypothesis. Similar arguments have been made about local environmental controls when the effects of pollution extend beyond the local area. Local governments and their constituents enjoy the economic benefits of a manufacturer's decision to locate in their area but the environmental cost is shared more widely.

These assertions that regulation results in a "race to the bottom" by economic efficiency standards stand in sharp contrast to Tiebout's notion of beneficial competition. The key difference is seen in the example of environmental controls. Tiebout's result applies when both the costs and the benefits of the pollution-generating activity accrue to the constituents of the local governmental decision maker. Inefficient regulatory choices are more likely to arise when the costs spill over between localities.

The clean dichotomy between beneficial and harmful regulatory competition relies on an additional important assumption involving the governmental decision makers' objectives. In the case of environmental regulation, the assumption is essentially that the local government acts to maximize its constituents' welfare. Other objectives are also possible, however. Stigler (1971) and Peltzman (1976), and the extensive literature that follows their seminal work emphasize the political economy of interest groups as a determining factor in regulatory decisions. Along these lines, one idea that is often voiced is that of "regulatory capture." This term expresses the notion that regulatory actions may be driven more by the interests of the firms in the regulated industry than by considerations of general or consumer welfare. In reference to banking in particular, Kane (1996) has suggested that regulators' self-interest can shape the outcomes of regulation. But there are alternative assumptions that one might make about regulators' objectives. One possibility is that individuals who have some discretion in choosing regulatory actions might be motivated by their personal reputations and career concerns. Another possibility, particularly relevant to settings where regulators can compete with one another, is that agencies seek to maximize their influence by regulating a large portion of the industry.

Clearly, the effects of competition among regulators could depend on regulators' motivations. In a setting of regulatory capture, competition could exacerbate the tendency to weigh the interests of the regulated industry above consumer welfare. If regulators are concerned for their personal reputations, their behavior and their response to competition would depend further on how they believe industry outcomes affect their reputations. For instance, a concern

for reputation might cause bank regulators to be conservative, preventing banks from taking actions that might have bad outcomes. Competition could counter this tendency by inducing regulators to loosen their control of risk taking in order to attract more client banks. While empirical evidence on the behavior of bank regulators and the effects of regulatory competition is sparse, Rosen (2001) has recently studied the characteristics and behavior of banks that switch their federal regulator. He finds evidence consistent with the idea that competition can be beneficial, as banks tend to improve their performance following a switch.

In addition to the underlying motives of regulators, another key factor affecting the way regulators behave under competition is the means of financing regulatory costs. A regulator that must cover all of its costs out of fees that it charges to its regulated businesses might behave quite differently from one that draws on general public funding. This distinction has in fact been highlighted in some recent discussions about the organization of bank regulation. The OCC, for instance, covers its expenses from examination fees, while the FDIC bundles regulation with deposit insurance, paying for both out of deposit insurance premiums. The OCC has argued (for example, in Hawke, 2002) that this difference can distort banks choices among their alternative federal regulators.

The following section sets out a model that focuses on the choice of a regulatory mechanism to control the risk-taking incentives of banks with insured deposits. That basic model provides a framework that allows the consideration of a number of alternative assumptions about the organization, financing, and motivation of regulators. An underlying assumption is that regulators have some discretion in choosing the parameters of their regulatory behavior. In the model, the key parameter is the frequency of examinations. While the actual degree of discretion exercised by bank regulators on this dimension is limited by statute, it is clear that, more generally, regulatory agencies have discretion over the intensity and informativeness of examinations, variables that would have the same effect as the simple probability that is chosen in the model.

2. A MODEL OF BANK REGULATION

A bank will be represented as an agent making an investment decision. The bank raises funds by issuing fully insured deposits. Depositors, therefore, are not particularly interesting actors in this model, as they supply funds perfectly elastically at the risk-free rate-of-return, normalized to zero. A bank raises a fixed amount of deposits, D , and can place funds into one of two investment projects, represented as “actions” a_0 and a_1 . Each action results in a probability distribution over the set of possible outcomes, $R = \{-\theta, -1, 1, \theta\}$, where $1 < \theta < D$. The outcome is the bank’s income (or loss), net of payment to

depositors. Let $P(a)$ denote the vector of probabilities if action a is taken. The specification of $P(a_0)$ and $P(a_1)$ is meant to capture the notion that one of the actions, a_0 , results in both higher risk and lower expected return than the other. A simple specification that captures this dominance is $P(a_0) = ((1 - p_0), 0, 0, p_0)$ and $P(a_1) = (0, (1 - p_1), p_1, 0)$, where $p_0 < 1/2 \leq p_1$. Hence, action a_0 represents a negative net-present-value investment, while a_1 has an expected return at least as great as the risk-free return.

Given full deposit insurance and the absence of any other regulation or intervention affecting its choice, the bank will choose the inferior action, a_0 , if $p_0\theta > p_1$, which will be assumed to be true. The bank's choice of action is subject to moral hazard, since the action cannot be observed by an outsider without cost. Hence, the deposit insurer faces the challenge of ensuring that the bank takes the productive action a_1 . The following analysis assumes a large number of banks, so that, if action a_1 is chosen by all banks, the fraction that earns positive income is equal to the probability p_1 (and similarly for action a_0).

The problem facing the deposit insurer here is quite simple if the insurer can impose ex post, state-contingent payments by the bank. Specifically, since the outcome θ is possible only if the risky action is taken, the insurer could ensure the choice of the preferred action, a_1 , by "taxing" the outcome θ sufficiently.¹ The analysis that follows assumes that such state-contingent payments are not feasible unless costly actions are taken. For instance, θ itself might be a random variable that takes a value of one or higher. Realized outcomes can be uncovered by the insurer only at a cost. Then, it is likely that such a tool would be used by the insurer in the event of negative returns in order to give the appropriate compensation to depositors. With positive returns, however, actual returns might remain unmeasured (by outsiders) as long as the bank makes its payments to depositors (plus an insurance "premium" that covers the expected costs of measurement for "failed banks"). This arrangement, however, would not solve the moral hazard problem of inducing the bank to take the preferred action. For any insurance premium π paid by "solvent" banks (banks with positive returns), if $p_0\theta > p_1$, as assumed, then $p_0(\theta - \pi) > p_1(1 - \pi)$. The left-hand side of this inequality would be the bank's net return under a_0 , while the right-hand side gives the return if a_1 is chosen.

An alternative assumption that prevents the regulator from being able to force the bank to choose a_1 using ex post penalties involves the differential observability of different outcomes. For instance, one could assume that losses can be observed without cost but that positive outcomes cannot be distinguished. This amounts to assuming that it is possible to hide profits but

¹ For a discussion of using the regulation of bank capital structure to achieve ex post payments by banks, see Prescott (2001).

not to hide or otherwise falsify losses. Mathematically, this assumption, which is maintained below, is equivalent to assuming that the cost of monitoring realized losses is zero.

In addition to the ability to measure outcomes after the fact, suppose that the insurer has the ability to determine whether the bank has chosen a_0 or a_1 before outcomes are realized and the ability to close down a bank that is found to have chosen the inferior investment strategy. An examination to determine the bank's action choice results in a cost of c_a , and an early closure of a bank results in a loss of $l < (1 - p_0)\theta - p_0\theta$. The loss l can be thought of as the resource cost of closing a bank early, and this cost is less than the expected losses from a bank that has taken action a_0 .

The regulator's problem is to choose a probability of examination ϕ , a course of action where an examination reveals a_0 , and a fee π to charge banks that do not fail.² Any such combination, (π, ϕ) , will be referred to as a policy. The assumptions above imply that it will be optimal to close a bank observed to have chosen a_0 . Accordingly, an efficient policy can be defined as a ϕ and a π that solve the following problem.

$$\max\{p_1 - (1 - p_1) - \phi c_a\}$$

$$s.t. \quad p_1(1 - \pi) \geq (1 - \phi)p_0(\theta - \pi) \quad (1)$$

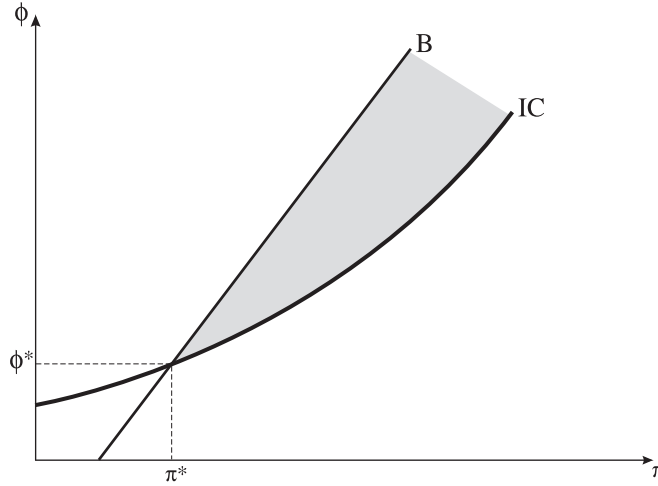
$$\text{and } p_1\pi \geq \phi c_a + (1 - p_1) \quad (2)$$

The objective function here is simply the total net returns from the operations of the typical bank and the regulator – the bank's expected net income minus the regulator's examination costs. Payments from deposit insurance, payments to depositors, and fee payments from the bank to the regulator are simply transfer payments. Hence, the objective function represents the regulated banking industry's net contribution to social welfare. The first constraint is an incentive compatibility constraint, stating that it must be in the bank's interest to choose the productive action a_1 . The left-hand side shows the expected return to the bank if it chooses a_1 , while the right-hand side shows the expected return from a_0 . In both cases, the bank only earns a positive return, out of which it pays the tax π , if it produces positive income. The right-hand side is weighted by $1 - \phi$, the probability of not being monitored. If the bank is monitored and discovered to have taken action a_0 , the regulator closes the bank, and the bank earns nothing. The second constraint is a consolidated budget constraint for bank examination and insurance, stating that fees collected from solvent banks must cover the examination costs and the costs of deposit insurance payouts.

The choice of an efficient arrangement is quite simple. Note first that the objective is equivalent to minimizing examination costs, and therefore

² In principle, one could allow for two distinct fees, depending on whether a surviving bank has or has not been examined. In the analysis below, it is assumed that the regulator must charge a single, nondiscriminating fee to all surviving banks.

Figure 1 The Efficient Policy



Notes: π is the fee charged to successful banks.

ϕ is the probability that a bank is examined.

To satisfy the budget constraint, a policy (π, ϕ) must lie below B . To satisfy the incentive constraint, a policy must lie above IC . Consequently, the shaded area is the set of feasible, self-financing policies. The efficient investment choice a_1 can be achieved at the lowest resource cost (examination cost) at the efficient policy (π^*, ϕ^*) .

the examination probability ϕ , subject to the two constraints. Second, the constraints can be represented by Figure 1 in which the incentive constraint is represented by the curve IC and the budget constraint by the line B .³ On B , which is linear in π and ϕ , the value of π when ϕ is zero is $(1 - p_1)/p_1$. Also along B , when $\phi = 1$, $\pi = (c_a + 1 - p_1)/p_1$. The shape of the incentive constraint can be seen by rewriting it as

$$\phi \geq 1 - \frac{p_1(1 - \pi)}{p_0(\theta - \pi)}.$$

The right-hand side of this inequality is increasing and convex in π . The intercept of IC on the ϕ -axis is $1 - p_1/p_0\theta$, which is greater than zero. Note also that IC goes through the point $(1, 1)$, so that IC and B cross at a point

³ The figure incorporates the additional assumption that $c_a < 2p_1 - 1$. This assumption says that examination costs are less than the average net income under action a_1 , and it is a sufficient condition for a nonempty constraint set. This assumption also ensures that the maximum value of the objective function in the efficient regulation problem is positive. That is, a regulated banking industry yields positive social surplus.

where both ϕ and π are less than one. Incentive compatibility requires that a policy (π, ϕ) lie above IC , while the budget constraint requires that a policy lie below B . The (π, ϕ) pairs that satisfy both constraints (that is, the pairs in the constraint set) are those that lie between IC and B . The efficient policy, which has the lowest ϕ in the constraint set, is denoted (π^*, ϕ^*) , where π^* is found from the consolidated budget constraint at equality, given ϕ^* .

The efficient policy varies with the model's parameters largely in the way that one would suspect. For instance, a worsening of the incentive problem, as would be represented by an increase in θ , leads to an increase in ϕ^* to maintain incentive compatibility. To cover the increase in examination costs, π^* must increase as well. However, one such comparative statics result might seem unexpected. Specifically, an increase in c_a , the cost of an examination, leads to an *increase* in ϕ^* , the frequency of examinations. This counterintuitive result arises from the interaction of the budget and incentive constraints. First, the rising costs need to be met with an increase in the regulator's revenue by increasing π . Next, note that an increase in π causes both the right- and left-hand sides of the incentive constraint to fall. The left-hand side falls faster, however, meaning that the bank may now find it advantageous to take the high-risk, low-return action a_0 . To counter this adverse incentive effect, it is necessary to increase the examination frequency.

3. BEHAVIOR OF A SINGLE REGULATOR WITH A CONSOLIDATED BUDGET CONSTRAINT

Suppose that a single government entity provides deposit insurance and performs bank examinations. This agency chooses a policy (π, ϕ) subject to the incentive and budget constraints in the problem above. Hence, the regulator knows that if it chooses a policy that does not satisfy the incentive constraint, banks will choose the high-risk, low-return investment, a_0 . Under this investment choice, however, the regulator will find it impossible to balance its budget. A balanced budget is impossible because $p_0\theta - (1 - p_0)\theta < 0$, and the most the regulator can charge banks that have positive returns is θ . Hence, the necessity of meeting the budget constraint assures that the regulator will enforce the efficient action, independent of the regulator's objective. A regulator that was willing and able to generate a budget deficit and whose behavior was described by the regulatory capture hypothesis might tolerate action a_0 . This action maximizes the banks' benefits from deposit insurance and limited liability.

While a regulator facing the consolidated budget constraint will always enforce the efficient action, that regulator will not always choose the efficient policy (π^*, ϕ^*) . This choice depends on the regulator's objectives. A regulator that wants to minimize costs will choose (π^*, ϕ^*) . There may be reasons, however, why a self-interested regulator would not seek to minimize costs.

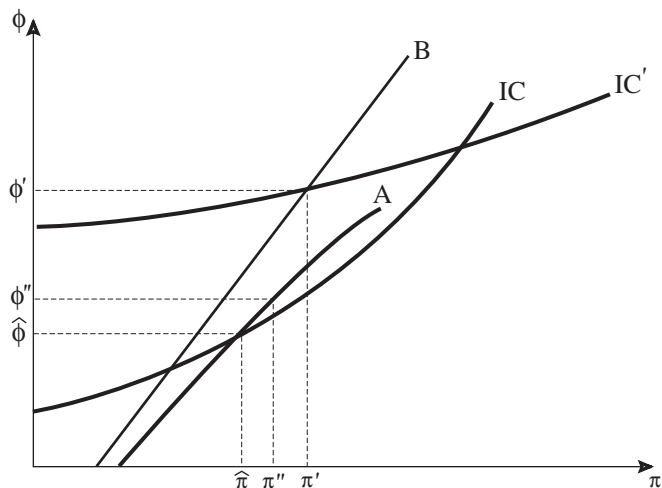
Another of the regulator's objectives could involve their attitude toward bank failures. For example, a "conservative" regulator could be characterized as one who is particularly averse to bank failures that are seen after the fact to have been the result of excessive risk taking. That is, regulators may seek to avoid the eventual revelation that failed banks under their authority took action a_0 . One way to achieve this goal would be for regulators to choose policies that ensure that no banks choose a_0 . In the basic model, with homogeneous banks, such regulators will choose the efficient policy. The following subsection presents an extension of the model with heterogeneous banks in which a conservative regulator could choose too restrictive of a policy.

An Extension Involving Multiple Bank Types

Suppose that there are two types of banks, differentiated only by their high-risk lending opportunities. A fraction λ of the banks will earn returns of θ (with probability p_0) or $-\theta$ (with probability $1 - p_0$) if they take action a_0 , as above. For the remaining banks, a_0 yields θ' (with probability p_0) or $-\theta'$ (with probability $1 - p_0$), where $\theta' > \theta$. The banks with θ' are "high risk," and those with θ are "low risk." If these two types of banks were regulated separately, with a separate (π, ϕ) for each, then the high-risk banks would have both a higher fee (π) and a higher frequency of examination (ϕ). Figure 2 shows the separate incentive constraints for the two types— IC for the low-risk banks and IC' for the high-risk. It takes more frequent examination, and therefore higher fees, to induce the high-risk bank to take the efficient action (a_1). As long as both types are taking the efficient action, then the budget constraint (B) is the same for both types. In this case, the efficient policies with separate treatment for the two types would be at the intersection of B and IC for the low-risk banks and at the intersection of B and IC' , the point labeled (π', ϕ') , for the high-risk banks.

It may not be possible for the regulator to distinguish between the two types of banks. That is, the regulator may have to set a single policy (π, ϕ) that applies to all banks. In this case, the policy (π', ϕ') is the least-cost policy that insures that all banks take action a_1 . However, this might not be the most efficient policy. In particular, if λ is close to 1, so that high-risk banks represent only a small fraction of the population, a policy that prevents only the low-risk banks from taking the high-risk action may be preferable. The best such policy is one that just satisfies the incentive constraint for the low-risk banks, allows high-risk banks to take action a_0 , and satisfies the budget constraint,

$$\begin{aligned} & [\lambda p_1 + (1 - \lambda)(1 - \phi)p_0]\pi \\ \geq & \lambda(1 - p_1) + (1 - \lambda)(1 - \phi)(1 - p_0)\theta' + (1 - \lambda)\phi l + \phi c_a. \end{aligned} \quad (3)$$

Figure 2 Two Bank Types

Notes: The budget constraint if all banks take action a_1 is represented by B . If high-risk banks take a_0 , while low-risk banks take a_1 , then the budget constraint is given by A . The incentive constraints are IC for the low-risk banks and IC' for high-risk banks. The best *conservative* policy (that induces all banks to choose a_1) is (π', ϕ') . But there are relatively few high-risk banks, then the policy $(\hat{\pi}, \hat{\phi})$ is efficient.

This budget constraint is represented by A in Figure 2, and the policy at the intersection of A and IC is denoted $(\hat{\pi}, \hat{\phi})$. When λ is large, A lies very close to B , as in the figure. Compared to (π', ϕ') , $(\hat{\pi}, \hat{\phi})$ involves increased costs associated with the failures and early closures of high-risk banks but a cost savings associated with the reduced examination frequency for all banks. When λ is large enough, the savings will outweigh the costs, making $(\hat{\pi}, \hat{\phi})$ the efficient policy.

In this extension of the model, the chosen policy may depend on the regulator's preferences and objective. As always, a welfare-maximizing regulator will choose the efficient policy. Suppose, however, that the regulator has the conservative preferences outlined above. That is, the regulator is particularly concerned with preventing bank failures that are found after the fact to have been caused by "excessive" risk taking. This concern might arise, for instance, because the regulator is sensitive to how such failures will affect his or her reputation, either with the legislature or with the public at large. Such a conservative regulator might well choose the policy (π', ϕ') , even when the

efficient policy is $(\hat{\pi}, \hat{\phi})$.⁴ This, then, is a case where competitive pressure among alternative regulators might be particularly beneficial.

4. COMPETITION BETWEEN TWO COMBINED INSURANCE AND REGULATION AGENCIES

As seen above, the interaction between the incentive and insurance-regulation budget constraints is the key to determining desirable policies. As an initial step in examining “competition” among regulators, consider the case in which each regulator also has deposit insurance responsibilities for the banks that it regulates. That is, each regulator has its own consolidated budget constraint. The interaction between the regulators is then described as a game in which each regulator chooses a policy, and banks respond by choosing between the regulators. Assume that if the regulators choose identical policies, banks divide evenly between the regulators.

To complete the specification of the game requires a specification of how the regulators’ payoffs respond to the policy choices. These payoff functions would reflect the regulators’ objectives, which might include such goals as cost minimization, or minimization of risk taking by banks (preventing banks from choosing a_0). In a setting with competing regulators, it is likely that whatever other criteria the regulators are considering, they also care about their share of the regulated industry. This objective might, for instance, arise out of a desire by the regulator to maximize its influence on the industry.

In the previous subsection’s extension of the basic model, suppose there are two regulators that care about two things. First, as discussed above, they are conservative, with a dislike for failures or early closures associated with banks taking the action a_0 . Second, each has a preference for regulating as large a share of the industry as possible. One could put more structure on these preferences by, for instance, specifying a function by which the regulators evaluate different possible outcomes. Even without such added structure, however, it is possible to examine the nature of the interaction between regulators’ policy choices. An equilibrium (*Nash equilibrium*) of the game is a pair of policy choices, one by each regulator, such that neither can do better by changing policy, given the policy of the other.

Notice first that given the nature of the game, and assuming the regulators have the same preferences, equilibrium must involve both regulators choosing

⁴A caveat is in order regarding the specification of “efficiency” when the regulator has a preference, whether personal or political, for preventing all banks from choosing a_0 . Strictly speaking, the social welfare function would be the industry’s net income minus examination costs *minus* any utility cost to the regulator that results if some banks choose a_0 . The latter is assumed to be small relative to banks’ income and examination costs. That is, while such a utility cost, even when small, can affect a regulator’s choice of policy, it is assumed that the cost is small enough that it does not affect the determination of an efficient policy.

the same policy. If they have different policies and all banks prefer one of the policies, then the regulator with the less preferred policy will certainly prefer to mimic the other and share the industry.⁵ Two likely candidates for equilibrium policies are the “conservative” policy (π', ϕ') from Figure 2 and the efficient policy $(\widehat{\pi}, \widehat{\phi})$. Recall that the latter policy is efficient under the assumption that λ , the relative number of low-risk banks, is large enough.

Can (π', ϕ') be an equilibrium policy? Suppose one regulator has chosen this policy and consider the other’s optimal response. In particular, consider the second regulator’s choice between (π', ϕ') and (π'', ϕ'') in Figure 2. All banks will prefer (π'', ϕ'') ; low-risk banks prefer it for its lower fee, and high-risk banks also enjoy the potential gains from taking the high-risk action. Note that this policy is also feasible, since it satisfies the consolidated budget constraint (A) that holds when high-risk banks choose a_0 . Given that its counterpart has chosen (π', ϕ') , a regulator will choose (π'', ϕ'') if the perceived benefit of regulating more banks is greater than the perceived cost of allowing a small number of high-risk banks to take action a_0 . Suppose the weights that the regulator places on these criteria are such that (π'', ϕ'') is the preferred of the two policies. Then (π', ϕ') is not an equilibrium policy. Neither, of course, is (π'', ϕ'') , since a rival can attract all banks away with a policy along A with a lower π and a lower ϕ . In this case, bidding by regulators results in an equilibrium policy of $(\widehat{\pi}, \widehat{\phi})$. In contrast, the absence of competition results in the conservative policy of (π', ϕ') ; a sole conservative regulator need not compete for clients and can instead focus only on making sure that no banks have an incentive to take a_0 .

The discussion in this section implicitly involves a regulators’ objective function that exhibits a trade-off between a taste for regulating as large a share of the industry as possible and a distaste for “excessive” risk taking by banks. The preceding paragraph describes a situation in which the former (the desire to increase “turf”) is strong enough that it eliminates the conservative policy (π', ϕ') as a potential equilibrium outcome. Indeed, this is a case in which regulators’ interest in increasing their turf serves a useful social purpose. Of course, it is possible for the other component of regulators’ preferences, their desire to limit risk taking, to be strong enough to support (π', ϕ') as an equilibrium policy. The following example illustrates these points, by taking the assumptions of this section and adding an explicit regulatory objective function.

⁵ It is also possible for the regulators to have different policies and for each type of bank to prefer a different one of the policies. This could only happen, however, if at least one of the policies is not incentive compatible for at least one of the types, since the two bank types’ preferences among incentive compatible policies (policies that induce the action a_1) are identical. A regulator that attracts only high-risk banks with a policy that induces action a_0 cannot satisfy its consolidated budget constraint. Therefore, such a mix of strategies is not an equilibrium outcome.

Example 1 *Label the regulators 1 and 2, and let regulator i 's preferences be represented by*

$$\alpha F^i - \beta D^i$$

where F^i is the fraction of the industry that i regulates, and D^i is the fraction of the banks regulated by i that take action a_0 . The parameters α and β measure the strength of the regulators' preferences for the two objectives. Now consider regulator 2's choice of policy if regulator 1 has chosen (π', ϕ') . In particular, consider regulator 2's choice between (π', ϕ') and a policy along A with lower π and ϕ than at (π', ϕ') . The point (π'', ϕ'') gives one such policy. If (π', ϕ') is preferred, then that is the equilibrium policy. If regulator 2 chooses (π', ϕ') , then the industry is evenly divided between the regulators, and no banks will choose a_0 . That is, $F^2 = 1/2$ and $D^2 = 0$. On the other hand, regulator 2 can capture the entire industry by choosing (π'', ϕ'') at the cost of inducing high-risk banks to take action a_0 . In this case, $F^2 = 1$, and $D^2 = (1 - \lambda)$. Regulator 2 will prefer (π', ϕ') over (π'', ϕ'') if $\alpha/2 \geq \alpha - \beta(1 - \lambda)$, that is if $\alpha/\beta \leq 2(1 - \lambda)$. As suggested above, if the relative distaste for risk taking is strong enough (if β is small enough relative to α), then the conservative policy (π', ϕ') can be an equilibrium. On the other hand, for any given preference specification, if the population of high-risk banks is small enough (λ is big enough), then (π', ϕ') will not be an equilibrium. When this is the case, then the efficient policy $(\hat{\pi}, \hat{\phi})$ is the equilibrium.

The efficient outcome that arises from regulatory competition is similar to the outcome that would arise in this environment if, instead of being determined by regulators, π and ϕ were set by private providers of deposit insurance with the ability to monitor and shut down banks under certain circumstances. A monopolist private insurer in this setting would pick high fees and a high probability of monitoring. In fact a monopolist's profit-maximizing decision, at least under some auxiliary assumptions, is to choose $\pi = \phi = 1$. Competition, on the other hand, would cause rival insurers to bid their insurance and monitoring offers down to the efficient policy.

One key to the efficiency result in this section is the consolidated budget constraints the regulators face. That is, each regulator is both an insurer and an examiner of its banks, and neither can draw on other sources of funds to cover any of its costs. With this assumption, the so-called "race-to-the-bottom" characteristic, by which regulatory competition leads to too little regulation (too little monitoring) cannot be an equilibrium result. From the status quo of $(\hat{\pi}, \hat{\phi})$ with regulators splitting the industry, the only way a regulator can attract more banks is by offering a policy that induces all banks to choose the high-risk action. But no such policy can satisfy the consolidated budget constraint.

This is true by the basic assumptions of the model.⁶ If all banks choose investments with negative expected value, there are not enough resources in successful banks to cover all the costs of insurance, let alone examination costs. Hence, a race to the bottom will not occur. When a regulator is both an examiner and an insurer of banks, the regulator internalizes the effects of examination policy on the deposit insurance fund.

Of course in the United States, the multiple federal bank regulatory agencies do not each face their own consolidated budget constraints. Instead, the FDIC provides deposit insurance to all banks. It finances this insurance with premiums charged to insured institutions (or more generally, by the maintenance of a fund built up by banks' premium payments). The FDIC finances its regulatory and supervisory costs out of the same revenue source as its insurance. At the same time, the FDIC's financial resources are supplemented by the full faith and credit of the federal government. The Federal Reserve pays for its regulatory activities out of its general revenue from central bank operations. The OCC covers its costs out of a fee charged to the banks it regulates. The next section considers how these differences complicate the interaction among regulators.

5. UNCONSOLIDATED BUDGET CONSTRAINTS

When the financing of deposit insurance and bank regulation are not consolidated, there is a possibility that competition among regulators will lead to undesirable results. The simplest way of examining this possibility is to assume that deposit insurance is financed out of general government revenues, while regulatory agencies cover their examination costs, and the costs associated with the early closure of banks, from the fees they charge. In this case, a regulator's budget constraint, assuming incentive compatibility, is

$$p_1\pi \geq \phi c_a.$$

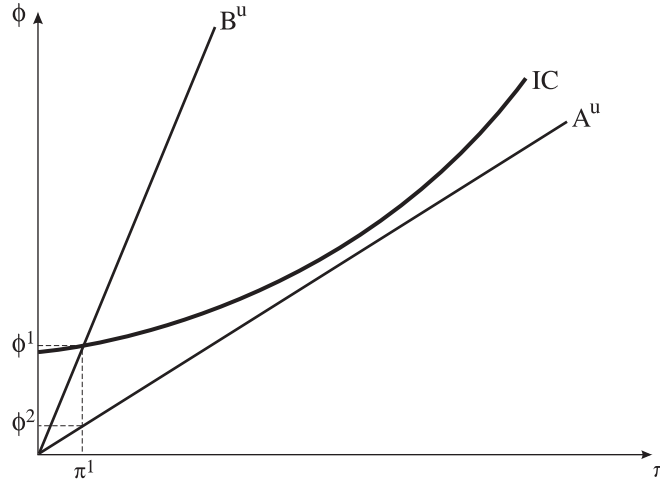
For a policy such that the incentive constraint is not satisfied, the budget constraint is

$$p_0\pi \geq \phi(c_a + l),$$

where l is the resource cost of closing a bank that is examined and found to have taken action a_0 . These two constraints are shown in Figure 3 as B^u and A^u respectively.

⁶The key assumption here is that a_0 represents an investment with a negative net present value. However, if a_0 were a positive net-present-value investment but dominated by a_1 , the efficient policy result would still hold. With all banks taking a_0 , π would have to be large in order to satisfy the consolidated budget constraint, making it impossible to choose a (π, ϕ) that banks prefer to $(\hat{\pi}, \hat{\phi})$.

Figure 3 Unconsolidated Budget Constraints



Notes: This figure shows the budget constraints for a regulator that only covers examination costs out of fees charged. If all banks take action a_1 the budget constraint is B^u . If high-risk banks take a_0 while low-risk banks take a_1 , then the budget constraint is given by A^u . The incentive constraint is IC . If regulators' distaste for risk taking by banks outweighs their desire to attract client banks, then the equilibrium policy is (π^1, ϕ^1) . If regulators desire to compete for clients is stronger, then the policy (π^1, ϕ^2) will bid clients away from a regulator offering (π^1, ϕ^1) .

This section considers the simplest case of a single type of bank (a single θ -type) and regulators whose objective is narrow and parochial. That is, each regulator simply seeks to maximize its turf, or the share of the market it regulates. Recall that under these assumptions, when regulators also faced consolidated budget constraints, competition led to efficient policies. Here that is not the case. Note that the efficient policy (π^*, ϕ^*) from Figure 1, because it satisfies the consolidated budget constraint, yields surplus funding to a regulator that only needs to cover examination costs. That is,

$$p_1\pi^* = \phi^*c_a + (1 - p_1) > \phi^*c_a.$$

Now consider the policy (π^1, ϕ^1) . This is the lowest cost policy that induces banks to choose a_1 and covers examination costs. This policy cannot be an equilibrium when regulators care only about the size of their turf. The policy (π^1, ϕ^2) will be strictly preferred by all banks, because it allows them the opportunity to gamble for the large return, θ . Among policies that induce banks to choose a_0 , however, regulators will continue to bid for banks by reducing π (moving along A^u). Hence, with this specification of objectives

and budget constraints there is a tendency for the regulatory process to unravel altogether, resulting in an equilibrium with no examination ($\phi = 0$) and no fee charged to banks by regulators ($\pi = 0$). In the absence of any external constraint on regulators' discretion, the agencies have no incentive to engage in more than minimal regulatory activities. This case, then, represents the so-called "race to the bottom."

Now suppose that conservativeness, as specified in earlier sections, is also a part of the regulators' objectives. The previous subsection argued that a regulator with such a mix of motives might be willing to loosen regulation in a way that induces only a small number of banks to take action a_0 . With only a single type of bank, however, a regulator is less likely to choose a policy that causes *all* banks to take a_0 , even if doing so attracts many more banks to that regulator. This logic leads to an equilibrium policy of (π^1, ϕ^1) , assuming that there is no separate fee assessment for deposit insurance. While this policy preserves banks' incentives to take the efficient action, it requires a net subsidy to the combined activities of insurance and regulation.

With consolidated budget constraints, regulators directly internalize the effect of regulatory actions on deposit insurance exposure. This automatic connection is lost when regulation and insurance are separately funded. This separation creates a sort of artificial externality that has an effect similar to the externalities that can interfere with the Tiebout result of efficient policies under competition among local governments.

6. SUMMARY AND CONCLUDING REMARKS

The preceding sections presented a model in which the key function of bank regulation is the monitoring of the investment choices made by insured banks. The model predicts policy choices by regulators that depend on the structure of the banking industry (captured by the distribution of bank types), regulators' objectives, and the financing of bank regulation and deposit insurance (captured by the regulators' budget constraints). The key findings of the analysis are: 1) a single regulator facing a consolidated budget constraint and a homogeneous banking industry will typically choose an efficient policy; 2) if there are multiple bank types with a small number of particularly high-risk banks, a single regulator with conservative preferences toward bank risk taking may choose an excessively strict policy; 3) with consolidated budget constraints for all regulators, competition for "turf" among multiple regulators can lead to efficient policies; and 4) competition for turf among regulators whose budget constraints only cover examination costs (and not insurance costs) leads to a "race to the bottom."

The previous section's simple specification of unconsolidated budget constraints still does not match the actual organization and financing of bank deposit insurance and regulation in the United States. Most notably, one of the

agencies—the FDIC—finances both insurance for all banks and regulation for its banks out of the “fees” it charges to all banks. Further, by choosing to be regulated by the OCC or the Fed, however, a bank does not reduce the fees that it pays to the FDIC for insurance. Accordingly, the way in which fees enter into banks’ choices of regulators is more complicated in reality than in this article’s model. Still, since the financing of insurance and regulation is separated for all other banks other than those regulated by the FDIC, the budgetary externality discussed in this article is present.

Many other characteristics of actual bank regulation have also been left out of the analysis. Rather than presenting a richly detailed description of actual regulatory institutions, this article’s intent was to present a simple analytical framework for thinking about the interaction among alternative regulators. In spite of the inherent over simplification, the basic results of this article’s analysis are likely to carry over to more complex environments. Competitive interaction among regulators can have beneficial effects, but the separation of the financing of insurance and regulation can make those benefits less certain.

REFERENCES

- Hawke, John D. 2002. “Remarks by John D. Hawke, Jr., Comptroller of the Currency before the 38th Annual Conference on Bank Structure and Competition.” Chicago, Illinois, 9 May.
- Kane, Edward J. 1996. “Foundations of Financial Regulation.” In *Rethinking Bank Regulation: What Should Regulators Do?* proceedings of the 32nd Annual Conference on Bank Structure and Competition, Federal Reserve Bank of Chicago, 309–33.
- Karpoff, J., and Malatesta E. 1989. “The Wealth Effects of Second Generation State Takeover Legislation.” *Journal of Financial Economics* 25: 295–312.
- Peltzman, S. 1976. “Toward a More General Theory of Regulation.” *Journal of Law and Economics* 19: 211–40.
- Prescott, Edward S. 2001. “Regulating Bank Capital Structure to Control Risk.” Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Summer): 35–52.
- Rosen, R. J. 2001. “Do Regulators Search for the Quiet Life? The Relationship Between Regulators and the Regulated in Banking.” Federal Reserve Bank of Chicago Working Paper 2001-05.

Scott, K. E. 1977. "The Dual Banking System: Model of Competition in Regulation." *Stanford Law Review* 30: 1–49.

Stigler, G. 1971. "The Economic Theory of Regulation." *Bell Journal of Economics* 2: 3–21.

Tiebout, C. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64: 416–24.

Structure from Shocks

Michael Dotsey

Arguments favoring Keynesian models that incorporate sticky prices over real business cycle models are often made on the grounds that the correlations and impulse response patterns found in the latter are inconsistent with the data. Critics further assert that these correlations and patterns are consistent with models that include price stickiness. Gali (1999) constitutes a prominent example of this reasoning. He observes empirically that conditional on a technology shock the contemporaneous correlation between labor effort and labor productivity is negative. He then makes the case that this observation implies that prices are sticky. Basu, Fernald, and Kimball (1998), using different identifying assumptions, also find this correlation in the data and make a similar assertion. Mankiw (1989) provides still another example of this type of reasoning. He argues that RBC models imply, counterfactually, that inflation and real activity are negatively correlated and so are inconsistent with the existence of a Phillips curve, which would not be the case in sticky price models.

But statements like those of Gali, Basu, Fernald, and Kimball, and Mankiw assume a certain characterization of monetary policy. This assumption is best demonstrated by Gali (1999), who uses intuition based on a money supply rule to persuade us that sticky prices are needed to generate a fall in employment in the presence of positive technology shocks. The fall in employment together with an increase in output produces the negative correlation between employment and labor productivity. However, under a monetary policy that employs the interest rate rule estimated in Clarida, Gali, and Gertler (1998), positive technology shocks produce an increase in both employment and labor productivity. Given the correct estimation of the rule, one must question the conclusion drawn by Gali (1999) and the assertions of Basu and Kimball

■ Vice President and Senior Economic Policy Advisor, Research Department, Federal Reserve Bank of Philadelphia. I wish to thank Margarida Duarte, Marvin Goodfriend, Andreas Hornstein, Tom Humphrey, Robert King, Pierre Sarte, and John Walter for a number of useful suggestions. I have also greatly benefitted from many helpful discussions with Alex Wolman. The views expressed herein are the author's and do not represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

(1998).¹ Furthermore, work by Christiano and Todd (1996) is able to generate within the confines of the RBC paradigm the labor-productivity correlation estimated by Gali. Thus, it is clear that discriminating among classes of models based on a few correlations is a perilous enterprise, especially when those correlations are sensitive to the nature of monetary policy.

Within the confines of a model similar to that used by Gali (1999), I show the importance of the specification of monetary policy for the dynamic behavior of the economy. The model includes the more realistic specification of staggered price-setting rather than one-period price rigidity and includes capital accumulation. In all other respects the model is true to Gali's original specification. One can see the effects of the systematic portion of policy by examining how the model economy reacts to a technology shock under different specifications of a monetary policy rule. As in Dotsey (1999a), the experiments show that, in the presence of significant linkages between real and nominal variables, the way shocks propagate through an economy is intimately linked to the systematic behavior of the monetary authority. Thus, even correlations among real variables may be influenced by policy. In particular, the justification put forth by both Gali and Basu and Kimball for favoring a sticky-price model over an RBC model no longer applies.

Also, the correlations between real and nominal variables are sensitive to the specification of the central bank's feedback rule. Depending on the form of the monetary policy rule, the model is capable of producing either positive or negative correlations between output and inflation irrespective of whether prices are sticky or flexible. Therefore, Mankiw's reasoning for favoring a sticky-price model over a flexible-price model is not persuasive.² These latter results are reminiscent of the arguments made by King and Plosser (1984) concerning the correlations between money balances and output. Their article shows that the positive correlation between money and output need not reflect a causal role for money in the behavior of output.

This is not to say that the methodology advocated by Gali or the idea that some form of price stickiness characterizes the economic environment is invalid. Understanding the nature of the price-setting process is of paramount importance for conducting appropriate monetary policy, and comparing model impulse response functions with those found in the data is a potentially valuable tool in helping to discriminate between flexible and sticky price models. Gali's emphasis on conditional correlations is a useful refinement of

¹ One may also question whether labor effort does in fact decline following a technology shock. For a more detailed investigation concerning the robustness of results in the face of varying identifying assumptions, see Sarte (1997). In this article I choose to take as given the correctness of the empirical results cited by Gali and others.

² Similar findings occur with respect to an autonomous shift in aggregate demand. That is, the monetary policy rule is as important in determining the effects of the demand shock as is the underlying model structure. In particular, for the types of rules considered in this article one cannot discriminate between a flexible and sticky price model based on the correlations typically emphasized. For more detail see Dotsey (1999b).

this methodology. However, his conclusions—that the particular impulse response functions and correlations emphasized are helpful in understanding price-setting behavior—are not robust to the specification of monetary policy.

Section 1 sketches the underlying model common to the analysis. A key feature of the model is the presence of price stickiness. Section 2 describes the various monetary policy rules under investigation. One is a simple money growth rule and the others fall into the general category of Taylor-type rules, in which the nominal interest rate responds to inflation and output. Section 3 analyzes the response of the model economy to a technology shock. The responses are quite different and depend on the rule employed by the monetary authority. Section 4 concludes.

1. THE MODEL

For the purpose of this investigation, I use a framework that embeds sticky prices into a dynamic stochastic model of the economy. The underlying model is similar to that of Gali (1999), but it is somewhat less stylized. There are two main differences in the model here, but these do not qualitatively affect the results. The first is that price rigidity is introduced through staggered contracts, and the second is that capital is included. Under flexible prices the underlying economy behaves as a classic real business cycle model. The model is, therefore, of the new neoclassical synthesis variety and displays features that are common to much of the current literature using sticky price models.³ Agents have preferences over consumption, work effort, and leisure, and they own and rent productive factors to firms. For convenience, money is introduced via a demand function rather than entering directly in utility (as in Gali) or through a shopping time technology. Firms are monopolistically competitive and face a fixed schedule for changing prices. Specifically, one-quarter of the firms change their price each period, and each firm can change its price only once a year. This type of staggered time-dependent pricing behavior, referred to as a Taylor contract, is a common methodology for introducing price stickiness into an otherwise neo-classical model.

Consumers

Consumers maximize the following utility function:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t [\ln(C_t) - \chi_n n_t^\zeta - \chi_u U_t^\eta],$$

³ Examples of this literature are Goodfriend and King (1998), Chari, Kehoe, and McGrattan (1998), and Dotsey, King, and Wolman (1999).

where $C = [\int_0^1 c(i)^{(\varepsilon-1)/\varepsilon} di]^{\varepsilon/(\varepsilon-1)}$ is an index of consumption, n is the fraction of time spent in employment, and U is labor effort. This is the preference specification used by Gali (1999), and I use it so that the experiments carried out below are not influenced by an alteration in household behavior.

Consumers also face the intertemporal budget constraint

$$P_t C_t + P_t I_t \leq W_t n_t + V_t U_t + r_t P_t K_t + D_t$$

and the capital accumulation equation

$$K_{t+1} = (1 - \delta)K_t + \phi(I_t/K_t)K_t,$$

where $P = [\int_0^1 p(i)^{1-\varepsilon} di]^{1/(1-\varepsilon)}$ is the price index associated with both the aggregator C and an analogous investment aggregator I , W is the nominal wage for an hour of work, V is the nominal payment for a unit of effort, r is the rental rate on capital, δ is the rate at which capital, K , depreciates, and D is nominal profits remitted by firms to households. The function ϕ is concave and depicts the fact that capital is costly to adjust.⁴

The relevant first order conditions for the consumers' problem are given by

$$(W_t/P_t) = \chi_n \varsigma n_t^{\varsigma-1}, \quad (1a)$$

$$(V_t/P_t) = \chi_u \eta U_t^{\eta-1},$$

and

$$(1/C_t \phi'_t) = \beta E_t (1/C_{t+1} \phi'_{t+1}) [r_{t+1} \phi'_{t+1} + (1-\delta) + \phi_{t+1} - \phi'_{t+1} (\frac{I_{t+1}}{K_{t+1}})]. \quad (1c)$$

Equation (1a) indicates that agents supply the number of labor hours that equate their marginal disutility of labor with the real wage. Similarly, equation (1b) indicates that agents exert a level of effort that equates their marginal disutility of effort with the payment on effort. Equation (1c) employs the shorthand notation ϕ_t and ϕ'_t to indicate the function and its first derivative evaluated at time t investment-to-capital ratios. The intertemporal condition is consistent with optimal capital accumulation. Agents invest up to the point where the marginal utility cost of sacrificing one unit of current consumption equals the marginal benefit of additional future consumption. The derivatives of the adjustment cost scale the utility cost because in this case the marginal utility of investment and consumption are not equal. Adjustment costs also affect the value of next period's capital and thus enter the bracketed expression

⁴ Capital adjustment costs are included primarily for the purpose of making the impulse response functions smoother. As is typical in models with staggered price-setting, the impulse response functions can be rather choppy as firms cycle through the price adjustment process.

on the right-hand side of (1c). With no adjustment costs, $\phi(I/K) = I/K$ and $\phi' = 1$, (1c) would become the standard intertemporal first order condition.

The demand for money, M , posited rather than derived, is given by

$$\ln(M_t/P_t) = \ln Y_t - \eta_R R_t. \quad (2)$$

The nominal interest rate is denoted R , and η_R is the interest semi-elasticity of money demand. One could derive the money demand curve from a shopping time technology without affecting the results in the article.

Firms

There is a continuum of firms indexed by j that produce goods, $y(j)$, using a Cobb-Douglas technology that combines labor and capital according to

$$y(j) = a_t k(j)^\alpha l(j)^{1-\alpha}, \quad (3)$$

where a is a technology shock that is the same for all firms and l is effective labor, which is a function of hours and effort given by $l_t = n_t^\theta U_t^{1-\theta}$. Each firm rents capital and hires labor and labor effort in economywide competitive factor markets. The cost-minimizing demands for each factor are given by

$$\psi_t a_t (1 - \alpha) \theta (k_t(j)/l_t(j))^\alpha (U_t/n_t)^{1-\theta} = W_t/P_t, \quad (4a)$$

$$\psi_t a_t (1 - \alpha) (1 - \theta) (k_t(j)/l_t(j))^\alpha (U_t/n_t)^{-\theta} = V_t/P_t, \quad (4b)$$

and

$$\psi_t a_t \alpha (l_t(j)/k_t(j))^{1-\alpha} = r_t, \quad (4c)$$

where ψ is real marginal cost. Equation (4a) equates the marginal product of labor with the real wage, and (4b) indicates that firms pay for effort until the marginal product on increased effort equals the payment for effort. In equation (4c), cost minimization implies that the marginal product of capital equals the rental rate. The above conditions also imply that capital-labor ratios and employment-effort ratios are equal across firms and that $U/n = ((1 - \theta)/\theta)(W/V)$. Using the latter relationship and equations (1a) and (1b) yields the reduced form production function $y(j) = a_t A k(j)^\alpha n(j)^\varphi$, where $\varphi = \theta(1 - \alpha) + (\psi/\eta)(1 - \theta)(1 - \alpha)$ and A is a function of the parameters θ , χ_n , χ_u , ψ , and η .

Although firms are competitors in factor markets, they possess some monopoly power over their own product and face downward-sloping demand curves of $y(j) = (p(j)/P)^{-\varepsilon} Y$, where $p(j)$ is the price that firm j charges for its product. This demand curve results when individuals minimize the cost of purchasing the consumption and investment indices represented by C and I . Thus $Y = C + I$. Firms are allowed to adjust their price once every four periods, and they may choose a price that will maximize the expected value

of the discounted stream of profits over that period. Specifically, a firm that sets its price in period t has the objective

$$\max_{p_t(j)} E_t \sum_{\tau=t}^{t+3} (\lambda_{\tau}/\lambda_t) \omega_{\tau}(j),$$

where real profits at time τ , $\omega_{\tau}(j)$, are given by $[p_t^*(j)y_{\tau}(j) - \psi_{\tau} P_{\tau} y_{\tau}(j)]/P_{\tau}$, and λ is the multiplier associated with the consumer's budget constraint.

As a result of this maximization, an adjusting firm's price is given by

$$p_t^*(j) = \frac{\varepsilon}{\varepsilon - 1} \frac{\sum_{h=0}^3 \beta^h E_t \{ (\lambda_{t+h}/\lambda_t) \psi_{t+h} (P_{t+h})^{1+\varepsilon} Y_{t+h} \}}{\sum_{h=0}^3 \beta^h E_t \{ (\lambda_{t+h}/\lambda_t) (P_{t+h})^{\varepsilon} Y_{t+h} \}}. \quad (5)$$

Further, the symmetric nature of the economic environment implies that all adjusting firms will choose the same price. One can see from equation (5) that, in a regime of zero inflation and constant marginal costs, firms would set their relative price $p^*(j)/P$ as a constant markup over marginal cost of $\frac{\varepsilon}{\varepsilon-1}$. In general, a firm's pricing decision depends on future marginal costs, the future aggregate price level, future aggregate demand, and future discount rates. For example, if a firm expects marginal costs to rise in the future, or if it expects higher rates of inflation, it will choose a relatively higher current price for its product.

The aggregate price level for the economy will depend on the prices charged by the various firms. Since all adjusting firms choose the same price, there will be four different prices charged for the various individual goods. The aggregate price level is, therefore, given by

$$P_t = \left[\sum_{h=0}^3 (1/4) (p_{t-h}^*)^{1-\varepsilon} \right]^{(1/(1-\varepsilon))}. \quad (6)$$

Steady State and Calibration

An equilibrium in this economy is a vector of prices p_{t-h}^* , wages, rental rates, and quantities that solves the firm's maximization problem and solves the consumer's optimization problem, such that the goods, capital, and labor markets clear. Furthermore, the pricing decisions of firms must be consistent with both the aggregate pricing relationship (6) and the behavior of the monetary authority described in the next section. In an examination of how the economy behaves when the central bank changes its policy rule, the above description of the private sector will remain invariant across policy rules and experiments.

The steady state is solved for the following parametrization. Labor's share, $1 - \alpha$, is set at $2/3$, $\zeta = 9/5$, $\beta = 0.984$, $\varepsilon = 10$, $\delta = 0.025$, $\eta_R = 0$, and agents spend 20 percent of their time working. These parameter values imply a steady state ratio of I/Y of 18 percent, and a value of $\chi = 18.47$. The choice of $\zeta = 9/5$ implies a labor supply elasticity of 1.25, which complies

with recent work by Mulligan (1998). A value of $\varepsilon = 10$ implies a steady state markup of 11 percent, which is consistent with the empirical work in Basu and Fernald (1997) and Basu and Kimball (1997). The interest sensitivity of money demand is set at zero. The demand for money is generally acknowledged to be fairly interest insensitive in the short run, with zero being the extreme case. Since the ensuing analysis concentrates on interest rate rules, the value of this parameter is unimportant. The adjustment cost function is parameterized so that the elasticity of the investment capital ratio with respect to Tobin's q is 0.25. This value is consistent with the estimate provided in Jermann (1998). The remaining parameter of importance is φ . Gali claims that a reasonable value for the parameter lies between 1 and 2, implying increasing returns to employment. Since the general nature of the results presented in Section 3 is not sensitive to this parameter, I set it to 1.5. Finally, the economy is buffeted by a random-walk shock to technology.

2. MONETARY POLICY

To study the effects of the systematic part of monetary policy on the transmission of technology shocks to the economy, I shall investigate the model economy's behavior under three types of policy rules. The first is a simple money growth rule, parameterized so that the economy experiences a steady state inflation rate of 2 percent. This inflation rate is held constant across all three rules.

The other two rules employ an interest rate instrument, thus falling into the category broadly labeled Taylor-type rules (Taylor 1993). The first rule allows the monetary authority to respond both to expected deviations of inflation from target and expected deviations of current output from its steady state or potential level. Because shocks are assumed to be contemporaneously observed in this model, the specification allows policy to respond to current movements in output. This rule is parameterized based on the estimations carried out in Clarida, Gali, and Gertler (1998) for the Volcker-Greenspan period.⁵ Their estimation also implies that the Fed is concerned with smoothing the behavior of the nominal interest rate; that behavior is incorporated into the following specification,

$$R_t = \bar{r} + \pi^* + 0.7R_{t-1} + 0.59(E_t\pi_{t+1} - \pi^*) + 0.04(Y_t - \bar{Y}_t). \quad (7)$$

The second rule is backward looking and allows the Fed to respond to deviations of inflation from target and of output levels from the steady state level of output. Specifically, I use the parameters in Taylor (1993),

$$R_t = \bar{r} + \pi^* + 1.5(\bar{\pi}_t - \pi^*) + 0.5(Y_t - \bar{Y}_t), \quad (8)$$

⁵ This specification is taken from their Table 3b.

where $\bar{\pi}_t$ is the average rate of inflation over the last four quarters, π^* is the inflation target of 2 percent, and \bar{Y}_t is the steady state level of output. Under this rule, when inflation is running above target or output is above trend, monetary policy is tightened and the nominal interest is raised. It is worth noting that because the coefficient on the output gap term is so small in the Clarida, Gali, and Gertler specification (7), there is no perceptible difference between impulse response functions generated in a model that omits this term entirely.

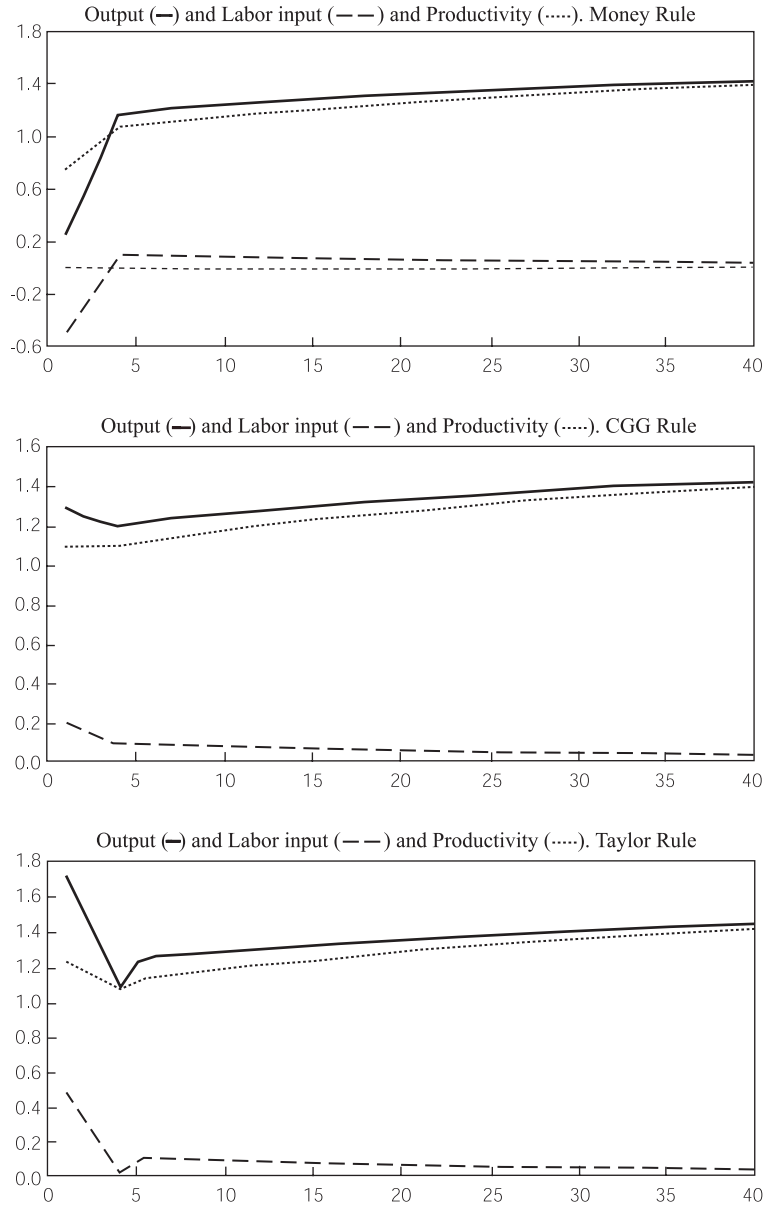
The experiments in the ensuing section show how the model economy's response to a technology shock depends on the specification of the systematic portion of monetary policy. Depending on the monetary rule in place, conditional correlations between output and productivity can vary both in magnitude and sign. In general, one can say nothing about the underlying structure of price setting—sticky or flexible—from these correlations.⁶

3. A COMPARISON OF THE POLICY RULES

I will next demonstrate how the model economy reacts to a technology shock. The underlying specification of the private sector is invariant in all experiments; only the specification of monetary policy is changed. As is conventional in modern macroeconomics, the model's behavioral equations are linearized and the resulting system of expectational difference equations is solved numerically using the procedures outlined in King and Watson (1998).

The response of the model economy to technology shocks is given in Figures 1 and 2. Figure 1 displays the response of hours, output, and average productivity, while Figure 2 examines the relationship between inflation and output. The differences across policy rules are striking. When money growth is held fixed, employment initially falls in response to a permanent change in productivity. With no deviation in money from steady state, there can be no deviation in nominal output from steady state. Because prices are sticky, they do not decline significantly. Therefore, the increase in output is not as great as the increase in productivity, and it takes less labor to produce the necessary output. This mechanism is stressed by Gali (1999). On the other hand, if the central bank follows the rule estimated by either Clarida, Gali, and Gertler (1997) or by Taylor (1993), monetary policy is very accommodative of the technology shock, so much so that the price level increases and output actually overshoots its new steady state level. The large increase in output requires additional labor, implying that labor productivity and labor hours are positively correlated, as they are in a simple RBC model. Thus, under

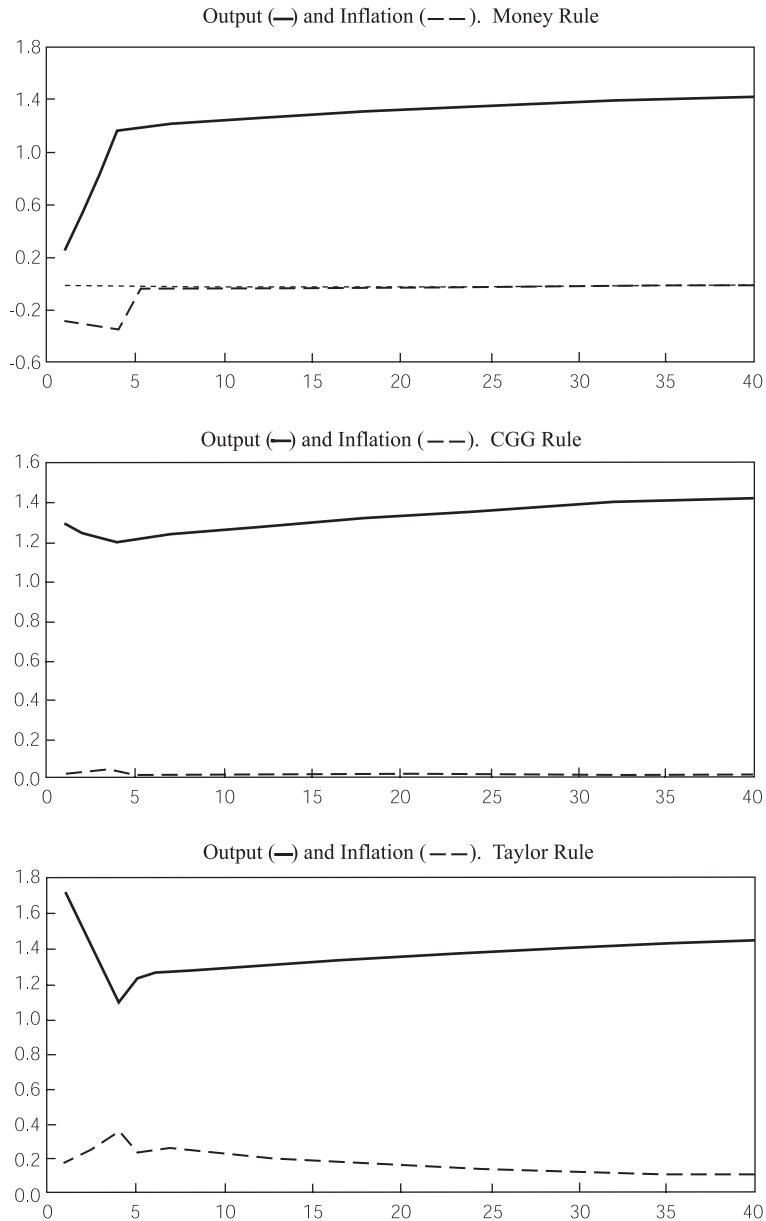
⁶As shown in Dotsey (1999b) a similar message applies to demand shocks. The article's concentration on the sensitivity of the economy's responses to shocks under different policies makes it similar to recent papers by McCallum (1999) and Christiano and Gust (1999).

Figure 1

reasonably specified monetary policy rules, one cannot infer the price-setting behavior of firms from the conditional correlation emphasized in Gali.⁷

⁷ McGrattan (1999) finds in a model with a CGG interest rate rule and two period overlapping Taylor-type contracts, in which prices are set a period in advance, that labor input declines on

Figure 2



impact in response to a technology shock. Her technology shock is stationary and potential output

To muddy the waters further, Christiano and Todd (1996) are able to generate a negative conditional correlation between employment and labor productivity in an RBC model that is augmented with a time-to-plan investment technology. Thus, one must conclude that this particular correlation is not very informative in identifying the feature of the economy that Gali seeks to uncover.

The impulse responses in Figure 2 show that inflation-output correlations are also sensitive to the specification of monetary policy. In both the Clarida, Gali, and Gertler and Taylor specifications, inflation is positively correlated with output. By contrast, in the constant money growth rule inflation is negatively correlated with output. The same relationships hold in a flexible-price model. Therefore, Mankiw's (1989) appeal to Phillips curve relationships as means to identify pricing behavior is problematic.

4. CONCLUSION

There are a number of points established by the analysis presented in this article. First and foremost is that the systematic component of monetary policy is important in determining the economy's reaction to shocks. In fact, the behavior of the model economy can differ so drastically across policies that forming some intuition about the underlying behavior of the private sector, such as whether prices adjust flexibly or are sticky, cannot be divorced from one's assumption about central bank behavior. In the limit, if the central bank were following the optimal policy prescribed in King and Wolman (1999), the bank's policy response to a technology shock would produce real behavior identical to that of the underlying real business cycle model.

Of more relevance to my analysis is the observation that a standard real business cycle model produces a positive correlation between labor productivity and hours, a result that is inconsistent with the data. Yet the same is true for a sticky-price model when the monetary authority follows either the rule estimated by Clarida, Gali, and Gertler (1999) or the rule estimated by Taylor (1993). The apparent inconsistency between model and data is, therefore, a poor reason to favor one type of model over the other, even though under a money stock rule the sticky-price model produces a negative correlation. The fact is, the Fed has probably never followed a money stock rule, so intuition drawn under such a rule may be of little value. In light of the results presented

does not respond to the shock as it does here when the technology shock is permanent. However, it is the presetting of prices that delivers the response of labor in her model. If prices were not preset, then labor would increase on impact as it does in experiments performed above.

above, discriminating among models based on impulse response functions is a subtle exercise that requires an accurate depiction of monetary policy.

REFERENCES

- Basu, Susanto, and John Fernald. 1997. "Returns to Scale in U.S. Production: Estimates and Implications." *Journal of Political Economy* 105 (April): 249–83.
- _____, and Miles Kimball. 1997. "Cyclical Productivity and Unobserved Input Variation." NBER Working Paper 5915.
- _____, John Fernald, and Miles Kimball. 1998. "Are Technology Improvements Contractionary?" Board of Governors of the Federal Reserve System International Finance Discussion Papers Number 625 (September).
- Chari, V. V., Patrick Kehoe, and Ellen McGrattan. 1998. "Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?" Federal Reserve Bank of Minneapolis Staff Report 217 (May).
- Christiano, Lawrence J., and Christopher J. Gust. 1999. "Taylor Rules in a Limited Participation Model." NBER Working Paper 7017 (March).
- _____, and Richard M. Todd. 1996. "Time to Plan and Aggregate Fluctuations," Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 14–27.
- Clarida, Richard, Jordi Gali, and Mark Gertler. 1998. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." CEPR Discussion Papers 1998.
- Dotsey, Michael. 1999a. "The Importance of Systematic Monetary Policy for Economic Activity." Federal Reserve Bank of Richmond *Economic Quarterly* 85 (Summer): 41–60.
- _____. 1999b. "Structure from Shocks." Federal Reserve Bank of Richmond Working Paper 99-6 (August).
- _____, Robert G. King, and Alexander L. Wolman. 1999. "State Dependent Pricing and the General Equilibrium Dynamics of Money and Output" *Quarterly Journal of Economics* CXIV (May): 655–90.
- Gali, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do

- Technology Shocks Explain Aggregate Productivity?" *The American Economic Review* 89 (March): 249–71.
- Goodfriend, Marvin S., and Robert G. King. "The New Neoclassical Synthesis and the Role of Monetary Policy." In *NBER Macroeconomics Annual 1997*, ed. Ben S. Bernanke and Julio J. Rotemberg. Cambridge, Mass.: MIT Press: 231–93.
- Jermann, Urban J. 1998. "Asset Pricing in Production Economies" *Journal of Monetary Economics* 41 (April): 257–76.
- King, Robert G., and Charles I. Plosser. 1984. "Money, Credit, and Prices in a Real Business Cycle Model." *The American Economic Review* 74 (June): 363–80.
- _____, and Mark W. Watson. 1998. "The Solution of Singular Linear Difference Systems Under Rational Expectations." *International Economic Review* 39 (November): 1015–26.
- _____, and Alexander L. Wolman. 1999. "What Should the Monetary Authority Do When Prices are Sticky?" In *Monetary Policy Rules*, ed. John B. Taylor. Chicago: University of Chicago Press: 349–98.
- Mankiw, N. Gregory. 1989. "Real Business Cycles: A New Keynesian Perspective." *Journal of Economic Perspectives* 3 (Summer): 79–90.
- McCallum, Bennett T. 1999. "Analysis of the Monetary Transmission Mechanism: Methodological Issues." Manuscript, Carnegie Mellon University (February).
- McGrattan, Ellen R. 1999. "Predicting the Effects of Federal Reserve Policy in a Sticky-Price Model: An Analytical Approach." Federal Reserve Bank of Minneapolis Working Paper 598 (December).
- Mulligan, Casey B. 1999. "Substitution over Time: Another Look at Life Cycle Labor Supply." In *NBER Macroeconomics Annual 1999*, ed. Ben S. Bernanke and Julio J. Rotemberg. Cambridge, Mass.: MIT Press: 75–134.
- Sarte, Pierre-Daniel G. 1997. "On the Identification of Structural Vector Autoregressions." Federal Reserve Bank of Richmond *Economic Quarterly* 83 (Summer): 45–67.
- Taylor, John B. 1993. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference on Public Policy* 39 (1993): 195–214.

Expectations and the Term Structure of Interest Rates: Evidence and Implications

Robert G. King and André Kurmann

Interest rates on long-term bonds are widely viewed as important for many economic decisions, notably business plant and equipment investment expenditures and household purchases of homes and automobiles. Consequently, macroeconomists have extensively studied the *term structure of interest rates*. For monetary policy analysis this is a crucial topic, as it concerns the link between short-term interest rates, which are heavily affected by central bank decisions, and long-term rates.

The dominant explanation of the relationship between short- and long-term interest rates is the *expectations theory*, which suggests that long rates are entirely governed by the expected future path of short-term interest rates. While this theory has strong implications that have been rejected in many studies, it nonetheless seems to contain important elements of truth. Therefore, many central bankers and other practitioners of monetary policy continue to apply it as an admittedly imperfect yet useful benchmark. In this article, we work to quantify both the dimensions along which the expectations theory succeeds in describing the link between expectations and the term structure and those along which it does not, thus providing a better sense of the utility of this benchmark.

Following Sargent (1979) and Campbell and Shiller (1987), we focus on linear versions of the expectations theory and linear forecasting models of

■ The authors would like to thank Michael Dotsey, Huberto Ennis, Pierre-Daniel Sarte, and Mark Watson for helpful comments. The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. Robert G. King: Professor of Economics, Boston University, and consultant to the Research Department of the Federal Reserve Bank of Richmond. André Kurmann: Department of Economics, University of Québec at Montréal.

future interest rate expectations. In this context, we reach five notable conclusions for the period since the Federal Reserve-Treasury Accord of March 1951.¹

First, cointegration tests confirm that the levels of both long and short interest rates are driven by a common stochastic trend. In other words, there is a permanent component that affects long and short rates equally, which accords with one of the basic predictions of the expectations theory.

Second, while changes in this stochastic trend dominate the month-to-month changes in long-term interest rates, the same changes affect the short-term rate to a much less important degree. We summarize our detailed econometric analysis with a useful rule of thumb for applied researchers: it is optimal to infer that the stochastic trend in interest rates has varied by 97 percent of any change in the long-term interest rate.² In this sense, the long-term interest rate is a good indicator of the stochastic trend in interest rates in general.³

Third, according to cointegration tests, the spread between long and short rates is not affected by the stochastic trend, which is consistent with the expectations theory. Rather, the spread is a reasonably good indicator of changes in the temporary component of short-term interest rates. Developing a similar rule of thumb, we compute that on average, a 1 percent increase in the spread indicates a 0.71 percent decrease in the temporary component of the short rate, i.e., in the difference between the current short rate and the stochastic trend.

Fourth, the expectations theory imposes important rational expectations restrictions on linear time series models in the spread and short-rate changes. Like Campbell and Shiller (1987), who pioneered testing of the expectations theory in a cointegration framework, we find that these restrictions are decisively rejected by the data. But our work strengthens this conclusion by using a longer sample period and a better testing methodology.⁴ We interpret the rejection as arising from predictable time-variations in term premia. Under the strongest form of the expectations theory, term premia should be constant and fluctuations in the spread should be entirely determined by expectations about future short-rate changes. However, our calculations indicate that—as another

¹ See Hetzel and Leach (2001) for an interesting recent account of the events surrounding the Accord.

² The sense in which this measure is optimal is discussed in more detail below, but it is based on minimizing the variance of prediction errors over our sample period of 1951 to 2001.

³ By contrast, a similar calculation indicates that changes in short-term interest rates are a much less strong indicator of changes in the stochastic trend: the comparable adjustment coefficient is 0.17 rather than 0.97. This finding is consistent with other evidence of important temporary variations in short-term interest rates, presented in this article and other studies.

⁴ We impose the cross-equation restrictions on the VAR and calculate a likelihood ratio test that compares the fit of the constrained and unconstrained VAR, while Campbell and Shiller (1987) use a Wald-type test of the restrictions on an estimated unrestricted VAR. It is now understood that Wald tests of nonlinear restrictions are sensitive to the details of how such tests are set up and suffer from much more severe small-sample bias than the method we employ here (see Bekaert and Hodrick [2001]).

rule of thumb—a 1 percent deviation of the spread from its mean signals a 0.69 percent fluctuation of the expectations component with the remainder viewed as arising from shifts in the term premia.

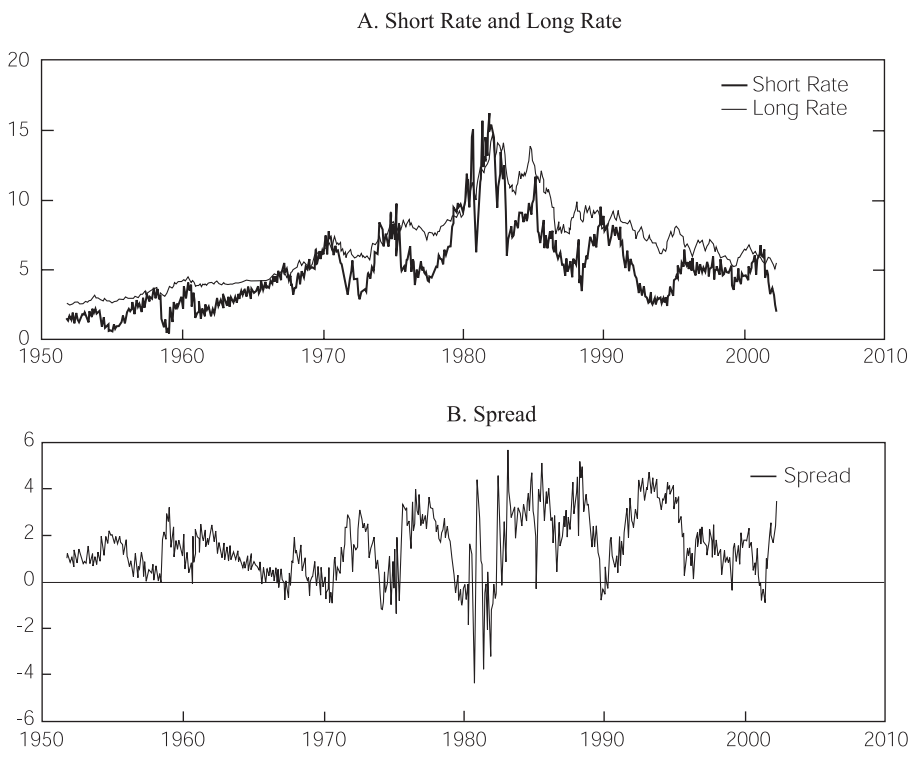
Fifth, based on the work by Sargent (1979), we show how to adapt the restrictions implied by the expectations theory to a situation where term premia are time-varying but unpredictable over some forecasting horizons. Our tests indicate that these modified restrictions continue to be rejected with forecasting horizons of up to a year. Thus, departures from the expectations theory in the form of time-varying term premia are not simply of a high frequency form, although the cointegration results indicate that the term premia are stationary.

Our empirical findings should provide some guidance for macroeconomic modeling, including work on small-scale econometric models and on monetary policy rules. In particular, our results suggest that the presence of a common stochastic trend in short and long nominal rates is a feature of post-Accord history that deserves greater attention. Furthermore, the detailed empirical results and the summary rules-of-thumb can be considered as a useful guide for monetary policy discussions. As an example, we ask whether the general patterns in the 50-year sample hold up over the period 1986–2001. Interestingly, we find a reduced variability in the interest rate stochastic trend: it is only about half as volatile as during the entire sample period. Nevertheless, the appropriate rule of thumb is still to view 85 percent of any change in the long rate as reflecting a shift in the stochastic trend. Our analysis also indicates that the expectations component of the spread (the discounted sum of expected short-rate changes) is of larger importance in the more recent sample, justifying an increase of the relevant rule-of-thumb coefficient from 69 percent to 77 percent. One interpretation of these different results is that they indicate increased credibility of the Federal Reserve System over the last decade and a half, which Goodfriend (1993) describes as the Golden Age of monetary policy because of enhanced credibility.

1. HISTORICAL BEHAVIOR OF INTEREST RATES

The historical behavior of short-term and long-term interest rates during the period April 1951 to November 2001 is shown in Figure 1. The two specific series that we employ have been compiled by Ibbotson (2002) and pertain to the 30-day T-bill yield for the short rate and the long-term yield on a bond of roughly twenty years to maturity for the long rate. One motivation for our use of this sample period is that the research of Mankiw and Miron (1986) suggests that the expectations theory encounters particular difficulties after the founding of the Federal Reserve System, particularly during the post-Accord period, because of the nonstationarity of short-term interest rates.

In this section, we start by discussing some key stylized facts that have previously attracted the attention of many researchers. We then conduct some

Figure 1 The Post-Accord History of Interest Rates

basic statistical tests on these series that provide important background to our subsequent analysis.

Basic Stylized Facts

We begin by discussing three important facts about the levels and comovement of short-term and long-term interest rates and then discuss two additional important facts about the predictability of these series.

Wandering levels: The levels of short-term and long-term interest rates vary substantially through time, as shown in Figure 1. Table 1 reports the very different average values over subsamples: in the 1950s, the short rate averaged 1.85 percent and the long rate averaged 3.02 percent; in the 1970s, the short rate averaged 6.13 percent and the long rate averaged 7.57 percent; and in the 1990s, the short rate averaged 4.80 percent and the long rate averaged 7.10 percent. These varying averages suggest that there are highly persistent factors that affect interest rates.

Table 1 Decade Averages

	Short Rate	Long Rate	Spread
1950s	1.85	3.02	1.17
1960s	3.81	4.63	0.82
1970s	6.13	7.57	1.45
1980s	8.54	10.69	2.15
1990s	4.80	7.10	2.30
Full Sample	5.13	6.67	1.57

Notes: All values are in percent per annum.

Comovement: While the levels of interest rates wander through time, subperiods of high average short rates are also periods of high average long rates. Symmetrically, short-term and long-term interest rates have a tendency to simultaneously display low average values within subperiods. This suggests that there may be common factors affecting long and short rates.

Relative stability of the spread: The spread between long- and short-term interest rates is much more stable over time, with average values of 1.17 percent, 1.45 percent, and 2.30 percent over the three decades discussed above. This again suggests that there is a common source of persistent variation in the two rates.

Predictability of the spread: While apparently returning to a more or less constant value, the spread between long and short rates appears relatively forecastable, even from its own past, because it displays substantial autocorrelation. This predictability has made the spread the focus of many empirical investigations of interest rates.

Changes in short-term and long-term interest rates: Figure 2 shows that changes in short and long rates are much less auto correlated. The two plots also highlight the changing volatility of short-term and long-term interest rates, which has been the subject of a number of recent investigations, including that of Watson (1999).

Basic Statistical Tests

The behavior of short-term and long-term interest rates displayed in Figures 1 and 2 has led many researchers to model the two series as stationary in first differences rather than in levels.

Unit root tests for interest rates: Accordingly, we begin by investigating whether there is evidence against the assumption that each series is stationary

in differences rather than in levels. For this purpose, the first two columns of Table 2 report regressions of the augmented Dickey-Fuller (ADF) form. Specifically, the regression for the short rate R_t takes the form

$$\Delta R_t = a_0 + a_1 \Delta R_{t-1} + a_2 \Delta R_{t-2} + \dots + a_p \Delta R_{t-p} + f R_{t-1} + e_{Rt}.$$

Our null hypothesis is that the short-term interest rate is difference stationary and that there is no deterministic trend in the level of the rate. In particular, stationarity in first differences implies that $f = 0$; if a deterministic trend is also absent, then $a_0 = 0$ as well. The alternative hypothesis is that the interest rate is stationary in levels ($f < 0$); in this case, a constant term is not generally zero because there is a non-zero mean to the level of the interest rate. The relevant test is reported in Table 2 for a lag length of $p = 4$.⁵ It involves a comparison of fit of the constrained regression in the first column and the unconstrained regression in the second column, with the former appropriate under the null hypothesis of a unit root and the latter appropriate under the alternative of stationarity. There is no strong evidence against the null, since the Dickey-Fuller F-statistic of 2.94 is less than the 10 percent critical value of 3.78.⁶ Looking at comparable results for the long rate R_t^L , we find even less evidence against the null hypothesis.⁷ The value of the Dickey-Fuller F-statistic is even smaller.⁸ We therefore model both interest rates as first difference stationary throughout our analysis.

In these regressions, we also find the first evidence of different predictability of short-term and long-term interest rates, a topic that will be a focus of much discussion below. Foreshadowing this discussion, we will find in every case that long-rate changes are less predictable than short-rate changes. In Table 2, the unconstrained regression for changes in the long rate accounts for about 3.5 percent of its variance, and the unconstrained regression for changes in the short rate accounts for about 8 percent of its variance.⁹

A simple cointegration test: Since we take the long-term and short-term rate as containing unit roots, the spread $S_t = R_t^L - R_t$ may either be

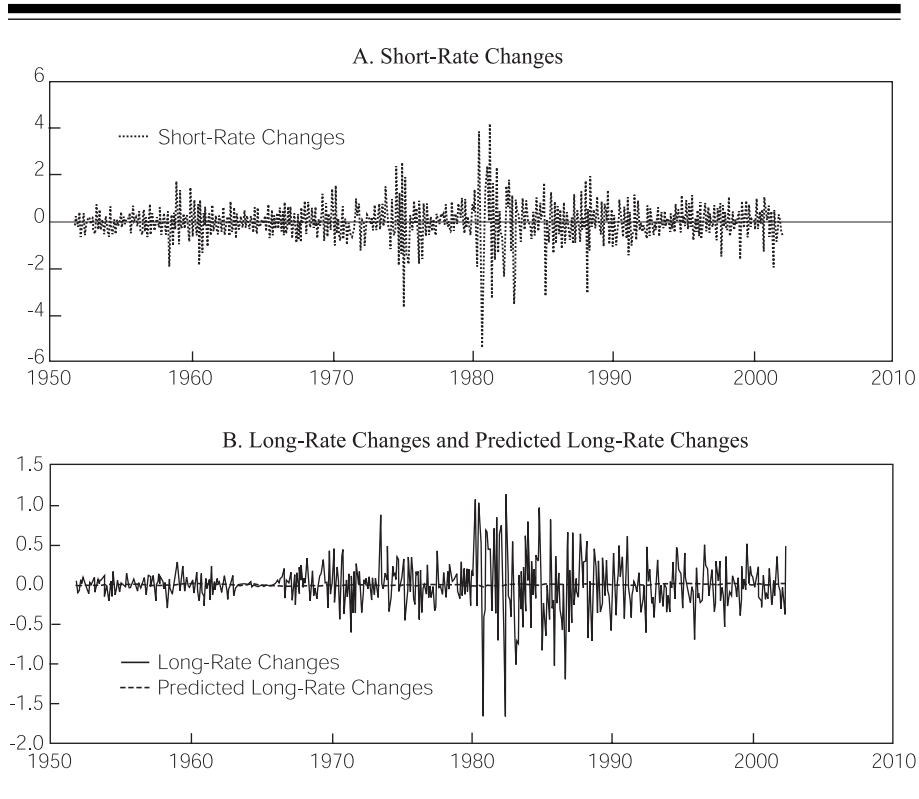
⁵ For the sake of simplicity, we use the same lag length of four months throughout the article. However, we also performed the different econometric tests with a higher lag length of $p = 6$ (as used for example by Watson [1999]) and found our results to be robust to this change.

⁶ See Dickey and Fuller (1981) for a discussion of the nonstandard distribution of this test statistic and a table of critical values.

⁷ A weaker null hypothesis, advocated for example by Hamilton (1994, 511–12), does not require $a_0 = 0$. This allows there to be a deterministic trend in the level of nominal rates, which seems implausible to us. But the second column of Table 2 also shows that there is no strong evidence against this null hypothesis, since $\hat{f} = -0.0283$ with a standard error of 0.0116. More specifically, the value of the Dickey-Fuller t-statistic is -2.43 , which is less than the 10 percent critical level of -2.57 .

⁸ The estimated level coefficient is also smaller and the associated Dickey-Fuller t-statistic takes on a value of -1.62 .

⁹ The constrained regressions display a similar pattern, although there are the familiar difficulties with interpreting R^2 when no constant term is present (see, for example, Judge et al. [1985, 30–31]).

Figure 2 The History of Interest Rate Changes

nonstationary or stationary. If the spread is stationary, then the long-term and short-term interest rates are cointegrated in the terminology of Engle and Granger (1987), since a linear combination of the variables is stationary. One simple test for cointegration when the cointegrating vector is known, discussed for example in Hamilton (1994, 582–86), is based on a Dickey-Fuller regression. In our context, we run the regression

$$\Delta S_t = a_0 + a_1 \Delta S_{t-1} + a_2 \Delta S_{t-2} + \dots + a_p \Delta S_{t-p} + f S_{t-1} + e_{St}.$$

As above, we take the null hypothesis to be that the spread is nonstationary, but that there is no deterministic trend in the level of the spread. The alternative of stationarity (cointegration) is a negative value of f ; the value of a_0 then captures the non-zero mean of the spread. The results in Table 2 show that we can reject the null at a high critical level: the value of the Dickey-Fuller F-statistic is 9.67, which exceeds the 5 percent critical level of 4.59.

Thus, we tentatively take the short-term and long-term interest rate to be cointegrated, but we will later conduct a more powerful test of cointegration. The regression results in Table 2 also highlight the fact that the spread is

Table 2 Unit Root Tests

Full Sample Estimates (1951.4–2001.11)						
	ΔR_t		ΔR_t^L		S_{t-1}	
	con.	uncon.	con.	uncon.	con.	uncon.
constant	0	0.0123 (0.0057)	0	0.0043 (0.0025)	0	0.0154 (0.0043)
lagged level	0	-0.0283 (0.0116)	0	-0.0068 (0.0042)	0	-0.1149 (0.0261)
lag 1	-0.2151 (0.0406)	-0.0198 (0.0411)	0.0896 (0.0409)	0.0918 (0.0409)	-0.3256 (0.0404)	-0.2471 (0.0437)
lag 2	-0.1649 (0.0415)	-0.1499 (0.0419)	-0.0441 (0.0407)	-0.0418 (0.0407)	-0.2610 (0.0425)	-0.1954 (0.0444)
lag 3	-0.0082 (0.0416)	0.0037 (0.0417)	-0.1390 (0.0407)	-0.1369 (0.0407)	-0.0759 (0.0425)	-0.0268 (0.0433)
lag 4	-0.1193 (0.0406)	-0.1094 (0.0407)	0.0384 (0.0409)	0.0398 (0.0409)	-0.1521 (0.0404)	-0.1157 (0.0407)
R-square	0.0721	0.0811	0.0301	0.0348	0.1322	0.1594
F-value	2.9352		1.4415		9.6688	

Notes: Numbers in parentheses represent standard errors. The critical 5 percent (10 percent) value for the Adjusted Dickey-Fuller F-test is 4.59 (3.78).

more predictable from its own past than are either of its components. In the unconstrained regression, 16 percent of month-to-month changes in the spread can be forecast from past values.

Cointegration of short-term and long-term interest rates is a formal version of the second stylized fact above: there is comovement of short and long rates despite their shifting levels. It is based on the third stylized fact: the spread appears relatively stationary although it is variable through time.

2. THE EXPECTATIONS THEORY

The dominant economic theory of the term structure of interest rates is called the expectations theory, as it stresses the role of expectations of future short-term interest rates in the determination of the prices and yields on longer-term bonds. There are a variety of statements of this theory in the literature that differ in terms of the nature of the bond which is priced and the factors that enter into pricing. We make use of a basic version of the theory developed in

Shiller (1972) and used in many subsequent studies.¹⁰ This version is suitable for empirical analyses of yields on long-term coupon bonds such as those that we study, since it delivers a simple linear formula for long-term yields. The derivation of this formula, which is reviewed in Appendix A, is based on the assumption that investors equate the expected holding period yield on long-term bonds to the short-term interest rate R_t , plus a time-varying excess holding period return k_t , which is not described or restricted by the model but could represent variation in risk premia, liquidity premia and so forth. It is based on a linear approximation to this expected holding period condition that neglects higher order terms. More specifically, the theory indicates that

$$R_t^L = \beta E_t R_{t+1}^L + (1 - \beta)(R_t + k_t), \quad (1)$$

where $\beta = 1/(1 + R^L)$ is a parameter based on the mean of the long-term interest rate around which the approximation is taken.¹¹

This expectational difference equation can be solved forward to relate the current long-term interest rate to a discounted value of current and future R and k :

$$R_t^L = (1 - \beta) \sum_{j=0}^{\infty} \beta^j [E_t R_{t+j} + E_t k_{t+j}]. \quad (2)$$

Various popular term-structure theories can be accommodated within this framework. The pure expectations theory occurs when there are no k terms, so that $R_t^L = (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t R_{t+j}$. This is a useful form for discussing various propositions about long-term and short-term interest rates that also arise in richer theories.

Implication for permanent changes in interest rates: Notably, the pure expectations theory predicts that if interest rates increase at date t in a manner which agents expect to be permanent, then there is a one-for-one effect of such a permanent increase on the level of the long rate because the weights sum to one, i.e., $(1 - \beta) \sum_{j=0}^{\infty} \beta^j = (1 - \beta)/(1 - \beta) = 1$. This is a basic and important implication of the expectations theory long stressed by analysts of the term structure and that appears capable of potentially explaining the comovement of short-term and long-term interest rates that we discussed above.

Implications for temporary changes in interest rates: Temporary changes in interest rates have a smaller effect under the pure expectations theory, with the extent of this effect depending on how sustained the temporary changes are assumed to be. Supposing that the short-term interest rate is governed by the simple autoregressive process $R_t = \rho R_{t-1} + e_{Rt}$ with the error term being

¹⁰ See, for example, Campbell, Shiller, and Schoenholz (1983) or Campbell and Shiller (1987).

¹¹ For our full sample, the average of the long rate equals 6.67 percent, or expressed as a monthly fraction: $R^L = 0.0667/12 = 0.00556$.

unforecastable, it is easy to see that $E R_{t+j} = \rho^j R_t$. It follows that a rational expectations solution for the long-term rate is

$$\begin{aligned} R_t^L &= (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t R_{t+j} \\ &= (1 - \beta) \sum_{j=0}^{\infty} \beta^j \rho^j R_t = \frac{1 - \beta}{1 - \beta \rho} R_t = \theta R_t. \end{aligned}$$

This solution can be used to derive implications for temporary changes in short rates. If these are completely transitory, so that $\rho = 0$, there is a minimal effect on the long rate, since $\theta = 1 - \beta \approx 0.005$. On the other hand, as the changes become more permanent (ρ approaches one) the θ coefficient approaches the one-for-one response previously discussed as the implication for fully permanent changes in the level of rates. Accordingly, the response of the long rate under the expectations theory depends on the degree of persistence that agents perceive in short-term interest rates, a property that Mankiw and Miron (1986) and Watson (1999) have exploited to derive interesting implications of the term structure theory that accord with various changes in the patterns of short-term and long-term interest rates in different periods of U.S. history.

The spread as an indicator of future changes: There has been much interest in the idea that the expectations theory implies that the long-short spread is an indicator of future changes in short-term interest rates. With a little bit of algebra, as in Campbell and Shiller (1987), we can rewrite (2) as

$$R_t^L - R_t = (1 - \beta) \sum_{j=0}^{\infty} \beta^j [(E_t R_{t+j} - R_t)] = \sum_{j=1}^{\infty} \beta^j E_t \Delta R_{t+j},$$

when there are no term premia.¹² Hence, the spread is high when short-term interest rates are expected to increase in the future, and it is low when they are expected to decrease. Further, permanent changes in the level of short-term interest rates, such as those considered above, have no effect on the spread because they do not imply any expected future changes in interest rates.

While these three implications can easily be derived under the pure expectations theory, they carry over to other more general theories so long as the changes in interest rates do not affect $(1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t k_{t+j}$ in (2). Further, while the pure expectations theory is a useful expository device, it is simply rejected: one of the stylized facts is that long rates are generally higher than short rates (there is a positive average value to the term spread). For this reason, all empirical studies of the effects of expectations on the long rate

¹² To undertake this derivation, note that $R_{t+j} - R_t = R_{t+j} - R_{t+j-1} + \dots + (R_{t+1} - R_t)$. Hence, each expected change enters many times in the sum, with a total effect of $\sum_{h=j}^{\infty} \beta^h E_t (R_{t+h} - R_{t+h-1}) = \frac{\beta^j}{1 - \beta} E_t (R_{t+j} - R_{t+j-1})$.

minimally use a modified form

$$R_t^L = (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t R_{t+j} + K,$$

where K is a parameter capturing the average value of the term spread that comes from assuming that k_t is constant.¹³

The Efficient Markets Test

As exemplified by the work of Roll (1969), one strategy is to derive testable implications of the expectations theory that (i) do not require making assumptions about the nature of the information set that market participants use to forecast future interest rates and that (ii) impose restrictions on a single linear equation. In the current setting, such an efficient markets test is based on manipulating (1) so as to isolate a pure expectations error, $R_t^L = \frac{1}{\beta} R_{t-1}^L - (\frac{1-\beta}{\beta})(R_{t-1} + K) + \xi_t$, where $\xi_t = R_t^L - E_{t-1} R_t^L$. As in Campbell and Shiller (1987, 1991), this condition may be usefully reorganized to indicate that the long-short spread (and only the spread) should forecast long-rate changes,

$$R_t^L - R_{t-1}^L = (\frac{1}{\beta} - 1)(R_{t-1}^L - R_{t-1} - K) + \xi_t,$$

which is a form that is robust to nonstationarity in the interest rate.

The essence of efficient markets tests is to determine whether any variables that are plausibly in the information set of agents at time $t - 1$ can be used to predict $\xi_t = R_t^L - R_{t-1}^L - (\frac{1}{\beta} - 1)(R_{t-1}^L - R_{t-1} - K)$. The forecasting relevance of any stationary variable can be tested with a standard t-statistic and the relevance of any group of p stationary variables can be tested by a likelihood ratio test, which has an asymptotic χ_p^2 distribution. Table 3 reports a battery of such efficient markets tests. The first regression simply is a benchmark, relating $R_t^L - R_{t-1}^L$ to a constant and to $(\frac{1}{\beta} - 1)S_{t-1}$ in the manner suggested by the efficient markets theory. The second regression frees up the coefficient on S_{t-1} and finds its estimated value to be negative rather than positive. The t-statistic for testing the hypothesis that the coefficient equals $(\frac{1}{\beta} - 1) = 0.005$ takes on a value of 2.345, which exceeds the standard 95 percent critical level. This finding has been much discussed in the context of long-term bonds and some other financial assets, in that financial markets spreads have a “wrong-way” influence on future changes relative to the predictions of basic theory.¹⁴ At the same time, the low R^2 of 0.0051 indicates that the prediction performance of the regression is very modest.

¹³ Below, we use the notation $K_t = (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t k_{t+j}$. But if $k_t = k$, then $K = k$.

¹⁴ See, for example, Campbell and Shiller (1991) for the term structure of interest rates or Bekaert and Hodrick (2001) for foreign exchange rates.

Table 3 Efficient Markets Tests

Full Sample Estimates (1951.4–2001.11)					
	test 1	test 2	ΔR_t^L test 3	test 4	test 5
constant	0.0002 (0.0010)	0.0023 (0.0015)	0.0033 (0.0015)	0.0032 (0.0016)	0.0034 (0.0016)
S_{t-1}	0.0050	-0.0147 (0.0084)	-0.0215 (0.0085)	-0.0214 (0.0095)	-0.0229 (0.0094)
ΔR_{t-1}				-0.0284 (0.0159)	-0.0147 (0.0171)
ΔR_{t-2}				-0.0321 (0.0160)	-0.0164 (0.0175)
ΔR_{t-3}				-0.0258 (0.0157)	-0.0250 (0.0168)
ΔR_{t-4}				0.0295 (0.0152)	0.0301 (0.0152)
ΔR_{t-1}^L			0.1002 (0.0410)		0.1048 (0.0409)
ΔR_{t-2}^L			-0.0496 (0.0406)		-0.0335 (0.0429)
ΔR_{t-3}^L			-0.1523 (0.0409)		-0.1328 (0.0440)
ΔR_{t-4}^L			0.0248 (0.0411)		0.0507 (0.0443)
R-square	-0.0040	0.0051	0.0408	0.0272	0.0550

Notes: Numbers in parentheses represent standard errors. F-stat (Regression 3 vs. Regression 2) = 5.598. F-stat (Regression 4 vs. Regression 2) = 3.436. F-stat (Regression 5 vs. Regression 3) = 2.280. F-stat (Regression 5 vs. Regression 4) = 4.411. The critical 5 percent (1 percent) F(4,400) value is 2.39 (3.36).

Additional evidence against the efficient markets view comes when lags of short-rate changes and lags of long-rate changes or both are added to the above equation. As regressions 3 through 5 in Table 3 show, the estimated coefficient on S_{t-1} remains significantly different from its predicted theoretical value. Furthermore, the prediction performance remains small (the R^2 is less than 10 percent for all the cases) and the F-tests reported at the bottom of the

table indicate that adding lagged variables does not significantly increase the explanatory power compared to the original efficient markets regression.¹⁵

The efficient markets regression again highlights that there is a substantial amount of unpredictable variation in changes in long bond yields, which makes it difficult to draw strong conclusions about the nature of predictable variations in these returns.¹⁶ One measure of the degree of this unpredictable variation is presented in panel B of Figure 2, where there is a very smooth and apparently quite flat line that is labelled as the “predicted changes in long rates.” Those predicted changes are $(\frac{1}{\beta} - 1)(R_{t-1}^L - R_{t-1})$ with a value of β suggested by the average level of long rates over our sample period. Panel B of Figure 2 highlights the fact that the expectations theory would explain only a tiny portion of interest rate variation if it were exactly true. Sargent (1979) refers to this as the “near-martingale property of long-term rates” under the expectations hypothesis. But it would not look very different if the fitted values of the other specifications in Table 3 were employed. Changes in the long rate are quite hard to predict and their predictable components are inconsistent with the efficient markets hypothesis.

Where Do We Go from Here?

Given that the efficient markets restriction is rejected, some academics simply conclude we know nothing about the term structure.¹⁷ However, central bankers and other practitioners actually do seem to employ the expectations theory as a useful yet admittedly imperfect device to interpret current and historical events (examples in this review are Dotsey [1998], Goodfriend [1993], and Owens and Webb [2001]). In the remainder of this analysis, we recognize that the expectations theory is not true but instead of simply rejecting it, we use modern time series methods to understand the dimensions along which it appears to succeed and those along which it does not. Section 3 develops and tests the common stochastic trend/cointegration restrictions that the expectations theory imposes. Consistent with earlier studies, we find that U.S. data

¹⁵ One potential explanation for the failure of the efficient markets tests—highlighted in Fama (1977)—is that there may be time-variation k_t in the equilibrium returns, which investors require to hold an asset. Then the theory predicts that

$$R_t^L - R_{t-1}^L = (\frac{1}{\beta} - 1)(R_{t-1}^L - R_{t-1} - k_{t-1}) + \xi_t.$$

But the researcher conducting the test does not observe time variation in k , which may give rise to a biased estimate on the spread. Fama stresses that efficient markets tests involve a joint hypothesis about the efficient use of information and a model of equilibrium returns, so that a rejection of the theory may arise from either element.

¹⁶ See the discussion of Nelson and Schwert (1977) on testing for a constant real rate.

¹⁷ For example, at a recent macroeconomics conference, one prominent monetary economist argued that the expectations theory of the term structure has been rejected so many times that it should never be built into any model.

do not allow us to reject these restrictions and, thus, that the theory appears to contain an important element of truth as far as the common stochastic trend implication is concerned. Section 4 then follows Sargent (1979) in developing and testing a variety of cross-equation restrictions that the expectations theory implies. These restrictions are rejected in the data. Finally, in Section 5, we build on the approach by Campbell and Shiller (1987) to extract estimates of changes in market expectations, which also allows us to extract estimates of time-variation in term premia.

3. COINTEGRATION AND COMMON TRENDS

A basic implication of the expectations theory is that an unexpected and permanent change in the level of short rates should have a one-for-one effect on the long rate. In other words, the theory implies that there is a common trend for the short and the long rate. This idea can be developed further using the concept of cointegration and related methods can be used to estimate the common trend.

The starting point is Campbell and Shiller's (1987) observation that present value models have cointegration implications, if the underlying series are nonstationary in levels, and that these implications survive the introduction of stationary deviations from the pure expectations theory such as time-varying term premia. In the context of the term structure, we can rewrite the long-rate equation (2) as

$$R_t^L - R_t = (1 - \beta) \sum_{j=0}^{\infty} \beta^j [(E_t R_{t+j} - R_t) + E_t k_{t+j}] \quad (3)$$

$$= \sum_{j=1}^{\infty} \beta^j E_t \Delta R_{t+j} + (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t k_{t+j} \quad (4)$$

so that the expectations theory stipulates that the spread is stationary so long as (i) first differences of short rates are stationary and (ii) the expected deviations from the pure expectations theory are stationary. Thus, cointegration tests are one way of assessing this implication of the theory.

In Section 1, we found evidence against the hypothesis that the spread contains a unit root and suggested that a stationary spread was a better description of the U.S. data. That is, we found some initial evidence consistent with modeling the short rate and the long rate as cointegrated. Here, in Section 3, we confirm that the spread also passes a more rigorous cointegration test. Given this result, we then define and estimate the common stochastic trend for the short rate and the long rate. We also present an easy-to-use rule of thumb that decomposes fluctuations of the short and the long rate into fluctuations in the common trend and fluctuations in the temporary components.

Testing for Cointegration

To develop the intuition behind the more rigorous cointegration tests, consider a vector autoregression (VAR) in the first difference of the short rate and the first difference of the long rate:

$$\Delta R_t = \sum_{j=1}^p a_j \Delta R_{t-j}^L + \sum_{j=1}^p b_j \Delta R_{t-j} + e_{Rt}, \quad (5)$$

$$\Delta R_t^L = \sum_{j=1}^p c_j \Delta R_{t-j}^L + \sum_{j=1}^p d_j \Delta R_{t-j} + e_{Lt}. \quad (6)$$

By virtue of the Wold decomposition theorem, we may be tempted to believe that such a VAR in first differences can approximate the dynamics of short- and long-rate changes arbitrarily well, so long as the vector $\Delta x_t = [\Delta R_t \ \Delta R_t^L]$ is a stationary stochastic process (this last condition being asserted by the Dickey-Fuller tests of the last section). However, if the two variables R_t and R_t^L are also cointegrated, then this argument breaks down. The above VAR represents a poor approximation in such circumstances because the short and long rate only contain one common stochastic trend and first differencing both variables thus deletes useful information.¹⁸

However, as Engle and Granger (1987) demonstrate, if first differences of x_t are stationary and there is cointegration among the variables of the form αx_t , then there always exists an empirical specification relating Δx_t , its lags Δx_{t-p} , and αx_{t-1} that describes the dynamics of Δx_t arbitrarily well. Such a system of equations is called a vector error correction model (VECM). In our context, if R_t and R_t^L are cointegrated, as under the weak form of the expectation theory discussed above, then the following VECM should provide a better description of the dynamics of Δx_t than the VAR in (5) and (6):

$$\Delta R_t = \sum_{j=1}^p a_j \Delta R_{t-j}^L + \sum_{j=1}^p b_j \Delta R_{t-j} + f[S_{t-1} - K] + e_{Rt}, \quad (7)$$

$$\Delta R_t^L = \sum_{j=1}^p c_j \Delta R_{t-j}^L + \sum_{j=1}^p d_j \Delta R_{t-j} + g[S_{t-1} - K] + e_{Lt}. \quad (8)$$

In these equations, f and g capture the effects of the lagged spread on forecastable variations in the short and long rates; K is the mean value of the spread.

¹⁸ In more technical terms, when R_t and R_t^L are cointegrated, then the vector moving average representation of $\Delta x_t = [\Delta R_t \ \Delta R_t^L]$ (which exists by definition of the Wold decomposition theorem) is noninvertible. As a result, no corresponding finite-order VAR approximation can exist. See Hamilton (1994, 574–75) for details.

To test for cointegration, we estimate both the VAR and the VECM and compare their respective fit. A substantial increase in the log likelihood of the VECM over the VAR signals that the cointegration terms aid in the prediction of interest rate changes. More specifically, a large likelihood ratio results in a rejection of the null hypothesis in favor of the alternative of cointegration. In particular, we follow the testing procedure by Horvath and Watson (1995) and assume *a priori* that the cointegrating relationship is given by the spread $S_t = R_t^L - R_t$ rather than estimating the cointegrating vector.¹⁹ Table 4 reports estimates of the VAR and VECM models for the lag length of $p = 4$, which we choose as the reference lag length throughout. Before discussing the cointegration test results in detail, it is worthwhile looking at a few elements that the VAR and VECM regressions have in common. First, changes in short rates are somewhat predictable from past changes in short rates, as was previously found with the Dickey-Fuller regression in Table 2. In addition, past changes in long rates are important for predicting changes in short rates in both the VAR and the VECM.²⁰ Finally, changes in short rates are predicted by the lagged spread: if the long rate is above the short rate, then short rates are predicted to rise. Second, changes in long rates are still fairly hard to predict with either the VAR or the VECM.

Moving to the cointegration test, the likelihood ratio between the VECM and the VAR equals $2 * (L_{VECM} - L_{VAR}) = 27.67$, which exceeds the 5 percent critical level of 6.28 calculated by the methods of Horvath and Watson (1995).²¹ In other words, we can comfortably reject the hypothesis of no cointegration between R_t and R_t^L , which is consistent with earlier studies

¹⁹ This type of test is somewhat more powerful than the unit root test on the spread reported in Table 2, which may be revealed by taking the difference between the two VECM equations and reorganizing the results slightly to obtain

$$\Delta S_t = \sum_{j=1}^p (c_j - a_j) \Delta R_{t-j}^L + \sum_{j=1}^p (d_j - b_j) \Delta R_{t-j} + (g - f) S_{t-1} + (e_{L_t} - e_{R_t}),$$

which can be further rewritten as

$$\Delta S_t = \sum_{j=1}^p (c_j - a_j) \Delta S_{t-j} + \sum_{j=1}^p (c_j + d_j - a_j - b_j) \Delta R_{t-j} + (g - f) S_{t-1} + (e_{L_t} - e_{R_t}).$$

That is, the Horvath-Watson test essentially introduces some additional stationary regressors to the forecasting equation for changes in the spread that was used in the DF test. Adding these regressors can improve the explanatory power of the regression, resulting in a more powerful test.

²⁰ That is, long rates Granger-cause short rates.

²¹ As Horvath and Watson (1995) stress, the relevant critical values for the likelihood ratio must take into account that the spread is nonstationary under the null. Thus, we cannot refer to a standard chi-square table. We estimate the VAR and VECM without constant terms, since we are assuming no deterministic trends in interest rates. However, we allow for a mean value of the spread, which is not zero as shown in (7) and (8). Unfortunately, this combination of assumptions means that we cannot use the tables in Horvath and Watson (1995), but must conduct the Monte Carlo simulations their method suggests to calculate the critical values reported in the text. Details are contained in replication materials available at <http://people.bu.edu/rking>.

Table 4 VAR/VECM Estimates

Full Sample Estimates (1951.4–2001.11)					
	ΔR_t		ΔR_t^L		
	VAR	VECM	VAR	VECM	
S_{t-1}		0.1101 (0.0237)		–0.0229 (0.0094)	
ΔR_{t-1}	–0.3095 (0.0408)	–0.2382 (0.0430)	0.0001 (0.0160)	–0.0147 (0.0171)	
ΔR_{t-2}	–0.1997 (0.0427)	–0.1393 (0.0440)	–0.0038 (0.0168)	–0.0164 (0.0175)	
ΔR_{t-3}	–0.0051 (0.0417)	0.0426 (0.0423)	–0.0151 (0.0164)	–0.0250 (0.0168)	
ΔR_{t-4}	–0.0879 (0.0377)	–0.0466 (0.0382)	0.0387 (0.0148)	0.0301 (0.0152)	
ΔR_{t-1}^L	0.8712 (0.1038)	0.8209 (0.1026)	0.0943 (0.0408)	0.1048 (0.0409)	
ΔR_{t-2}^L	0.6250 (0.1095)	0.5954 (0.1078)	–0.0397 (0.0430)	–0.0335 (0.0429)	
ΔR_{t-3}^L	0.1791 (0.1123)	0.1698 (0.1104)	–0.1347 (0.0441)	–0.1328 (0.0440)	
ΔR_{t-4}^L	0.0430 (0.1133)	0.0520 (0.1114)	0.0526 (0.0445)	0.0507 (0.0443)	
R-square	0.2220	0.2492	0.0459	0.0553	
F-statistic	21.2172	21.9030	3.5785	3.8610	

Notes: Numbers in parentheses represent standard errors. The likelihood ratio statistic of the VECM against the VAR is 27.6704. Comparing this value to the corresponding critical value in Horvath and Watson's tables leads to strong rejection of null of two unit roots (p-value higher than 0.01).

and reinforces the statistical support for the common trend implication of the expectations theory. Therefore, the data is consistent with the basic implication of cointegration of the expectations theory and we thus view the VECM as the preferred specification and assume cointegration for the remainder of our analysis.²²

²² An alternative approach in this section would be to estimate the cointegrating vector and use the well-known testing method of Johansen (1988). Horvath and Watson (1995) establish that their procedure is more powerful if the cointegrating vector is known.

Uncovering the Common Stochastic Trend

A key implication of cointegration in our context is that the short and long rates share a common stochastic trend, which we will now work to uncover.²³ Following Beveridge and Nelson (1981), the stochastic trend of a single series such as the short-term interest rate is defined as the limit forecast $\bar{R}_t = \lim_{k \rightarrow \infty} E_t R_{t+k}$, or equivalently

$$\bar{R}_t = R_{t-1} + \lim_{k \rightarrow \infty} \sum_{j=0}^k E_t \Delta R_{t+j}. \quad (9)$$

However, in order to obtain a series of \bar{R}_t , we need to take a stand on how to compute the $E_t \Delta R_{t+k}$ terms. The VECM suggests a straight-forward way to do so. Specifically, suppose that the system expressed by equations (7) and (8) is written in the form

$$z_t \equiv \begin{bmatrix} \Delta R_t \\ \Delta R_t^L \\ S_t \end{bmatrix} = H x_t$$

$$x_t = M x_{t-1} + G e_t,$$

where e_t is the vector of one-step-ahead forecast errors $e_t = [e_{Rt} \ e_{Lt}]'$ and $x_t = [\Delta R_t \ \Delta R_t^L \ \Delta R_{t-1} \ \Delta R_{t-1}^L \ \dots \ \Delta R_{t-(p-1)} \ \Delta R_{t-(p-1)}^L \ S_t]$ is the vector of information that the VECM identifies as useful for forecasting future spreads and interest rate changes. The matrix H simply selects the elements of x_t , and the elements of M and G depend on the parameter estimates $\{a, b, c, d, f, g\}$ in a manner spelled out in Appendix B.

Given this setup, forecasts of ΔR_{t+k} conditional information on x_t are easily computed as

$$E[\Delta R_{t+k} | x_t] = h_R E[z_{t+k} | x_t] = h_R H E[x_{t+k} | x_t] = h_R H M^k x_t,$$

where $h_R = [1 \ 0 \ 0]$ such that $\Delta R_t = h_R z_t$. Mapping these forecasts of ΔR_{t+k} into (9), we obtain a closed-form solution for the stochastic trend of the short rate:

$$\bar{R}_t = R_{t-1} + \sum_{k=0}^{\infty} h_R H M^k x_t = R_{t-1} + h_R H [I - M]^{-1} x_t.$$

The same procedure for computing multiperiod forecasts also provides a recipe for computing the stochastic trend in the long rate, that is,

$$\bar{R}_t^L = R_{t-1}^L + \lim_{k \rightarrow \infty} \sum_{j=0}^k E_t \Delta R_{t+j}^L$$

²³ The idea that cointegration implies common stochastic trends is developed in Stock and Watson (1988) and King, Plosser, Stock, and Watson (1991).

$$= R_{t-1}^L + \sum_{k=0}^{\infty} h_L H M^k x_t = R_{t-1}^L + h_L H [I - M]^{-1} x_t,$$

where $h_L = [0 \ 1 \ 0]$ such that $\Delta R_t^L = h_L z_t$. Finally, the difference between \bar{R}_t^L and \bar{R}_t is the limit forecast of the spread. By definition of cointegration, the spread is stationary and therefore its limit forecast must be a constant:²⁴

$$K = \lim_{k \rightarrow \infty} E_t S_{t+k} = \lim_{k \rightarrow \infty} E_t R_{t+k}^L - \lim_{k \rightarrow \infty} E_t R_{t+k} = \bar{R}_t^L - \bar{R}_t.$$

Thus, the trends for the long rate and the short rate differ only by the constant K : in other words, the long rate and the short rate have a common stochastic trend component. Since this is sometimes termed the permanent component, deviations from it are described as temporary components. Using this language, the temporary component of the short rate is $R_t - \bar{R}_t$ and that of the long rate is $R_t^L - \bar{R}_t^L$.

A Stochastic Trend Estimate: 1951–2001

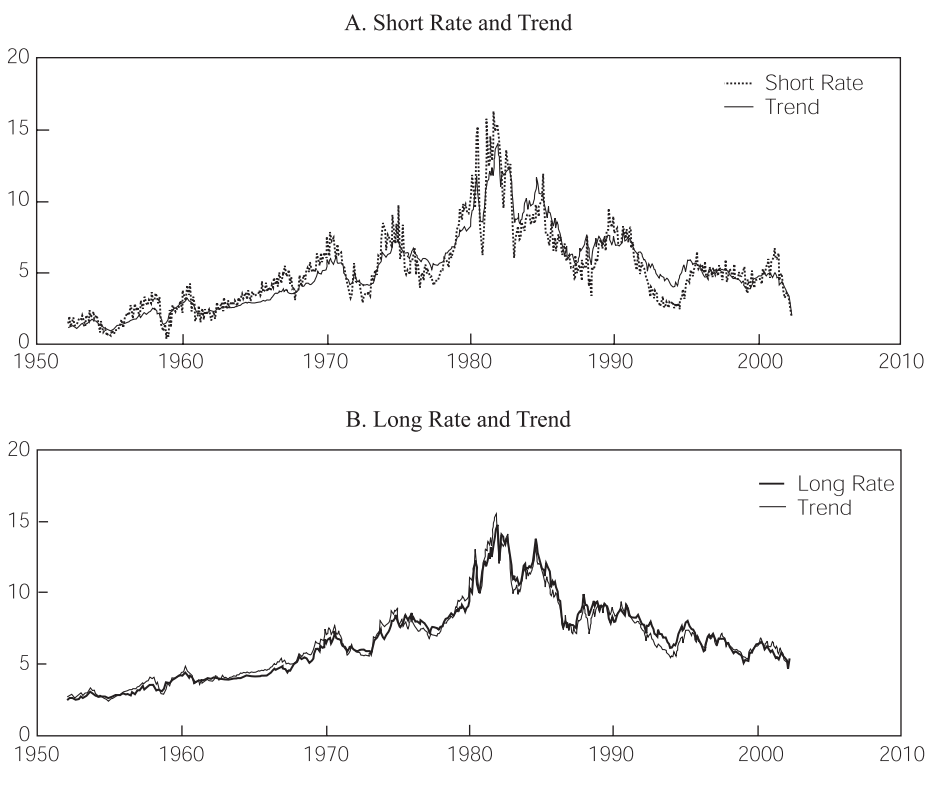
Figure 3 shows the common stochastic trend in long and short rates based on the VECM from Table 3, constructed using the method that we just discussed. In line with the expectations theory, we interpret this stochastic trend as describing permanent changes in the level of the short rate, which are reflected one-for-one in the long rate.

Short rates and the stochastic trend: In panel A, we see that the short rate fluctuates around its stochastic trend. There are some lengthy periods, such as the mid-1960s, where the short rate is above the stochastic trend for a lengthy period and others, such as the mid-1990s, where the short rate is below the stochastic trend. The vertical distance is a measure of the temporary component to short rates, which we will discuss in greater detail further below.

Long rates and the stochastic trend: In panel B, we see that the long rate and the stochastic trend correspond considerably more closely. This result accords with a very basic implication of the expectations theory: long rates should be highly responsive to permanent variations in the short-term interest rate.²⁵

²⁴ Under the expectations theory with a constant term premium, the average value of the spread must be the term premium K . So, to avoid proliferation of symbols, we use that notation here.

²⁵ To understand the sensitivity of the trend to the form of the estimated equation for the long rate, we compared three alternative measures of the trend. The first was the test measure based on the estimated VECM (i.e., the one reported in this section); the second was based on replacing the long-rate equation with the result of a simple regression of long-rate changes on the spread (i.e., the specification that we used for testing the efficient markets restriction above) so that there was a small negative weight on the spread in the long-rate equation; and the third

Figure 3 Interest Rates and the Common Stochastic Trend

Variance Decompositions

It is useful to consider a decomposition of the variance of short-rate and long-rate changes into contributions in terms of changes in the temporary and permanent components. For the short-rate changes, since $var(\Delta R_t) = var(\Delta \bar{R}_t + \Delta(R_t - \bar{R}_t))$, this decomposition takes the form

$$\begin{aligned} var(\Delta R_t) &= var(\Delta \bar{R}_t) + var(\Delta(R_t - \bar{R}_t)) \\ &\quad + 2 * cov(\Delta \bar{R}_t, \Delta(R_t - \bar{R}_t)) \\ 0.656 &= 0.105 + 0.544 + 2 * (0.004) \end{aligned}$$

was based on the efficient markets restriction (i.e., placed a small positive weight on the lagged spread). While there were some differences in these trend estimates on a period-by-period basis, they tell the same basic story in terms of the general pattern of rise and fall in the stochastic trend.

Table 5 Summary Statistics for Permanent-Temporary Decomposition

Full Sample Estimates (1951.4–2001.11)		
A. Short-Rate Changes		
Total	Permanent	Temporary
0.6559	0.1083	0.5477
0.4133	0.1046	0.0036
0.9168	0.0152	0.5440
B. Long-Rate Changes		
Total	Permanent	Temporary
0.0826	0.0802	0.0023
0.8631	0.1046	−0.0244
0.0499	−0.4614	0.0268
C. Long-Short Spread		
Total	Temporary Long Rate	Temporary Short Rate
1.9318	0.5649	−1.3668
−0.9920	−0.3841	0.9827
0.9559	0.1808	−0.9114

Notes: Table 5 is based on the VECM estimates in Table 4. Each panel contains a 3 by 3 matrix. On the diagonal, variances are reported (e.g., the variance of changes in long rates is 0.0826). Above the diagonal, covariances are listed (e.g., the covariance between changes in the long rate and changes in its permanent component is 0.0802). Below the diagonal, the corresponding correlation is reported (e.g., the correlation between changes in the long rate and changes in its permanent component is 0.8631).

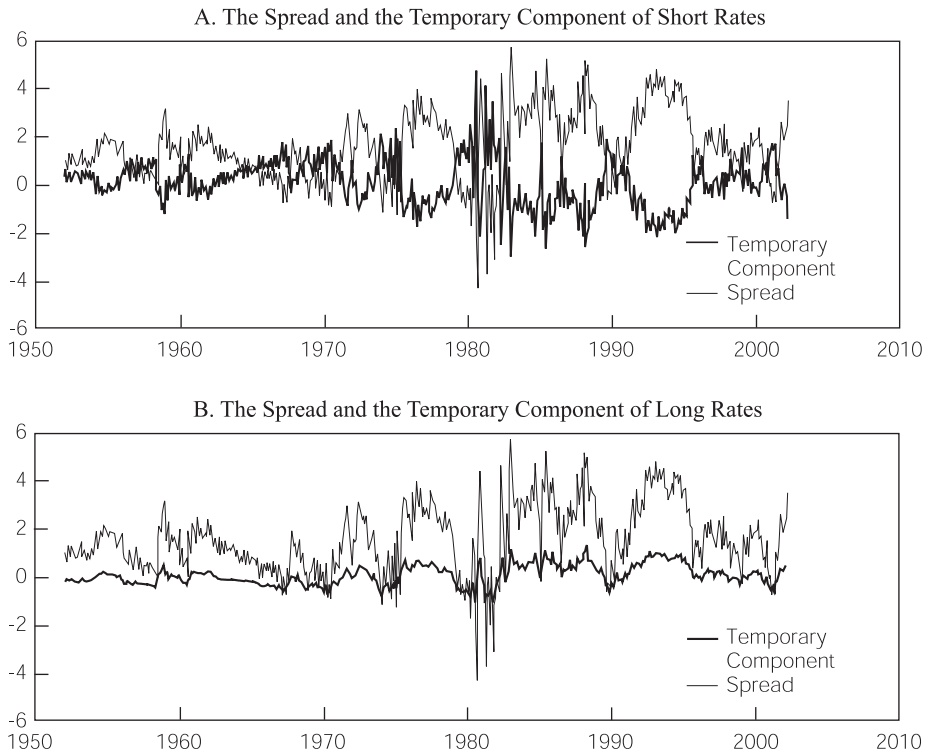
with the last line drawn from the first panel of Table 5.²⁶

The variance of month-to-month changes in interest rates is 0.66. Changes in the temporary component account for the great bulk (82.9 percent) of this variance, while the variance of changes in the permanent component contributes 15.9 percent and the covariance between the two components contributes only about 1.2 percent.

For the long rate, the decomposition takes conceptually the same form, but we find a very different result in terms of relative contributions:

$$\begin{aligned} \text{var}(\Delta R_t^L) &= \text{var}(\Delta \bar{R}_t^L) + \text{var}(\Delta(R_t^L - \bar{R}_t^L)) \\ &\quad + 2 * \text{cov}(\Delta \bar{R}_t^L, \Delta(R_t^L - \bar{R}_t^L)) \end{aligned}$$

²⁶ Here and below, our estimate of the stochastic trend allows us to calculate the variance decomposition, including the variance of changes in the trend and the covariance term. Note that due to rounding errors, the variance decompositions do not add up exactly.

Figure 4 The Spread and Temporary Components

$$0.083 = 0.104 + 0.027 + 2 * (-0.024)$$

First, the overall variance of month-to-month changes in the long rate is much smaller. In contrast to the short rate, this variance is dominated by the variance in its permanent component, which is actually somewhat larger because there is a negative correlation between the permanent and the transitory component.

The permanent-temporary decomposition also permits us to undertake a decomposition of the long-short spread, which is displayed in Figure 4. The spread and the two temporary components are connected via the identity

$$S_t - S = R_t^L - R_t - S = (R_t^L - \bar{R}_t^L) - (R_t - \bar{R}_t).$$

Hence, there is a mechanical inverse relationship between the spread and the temporary component of the short rate, which is clearly evident in panel A of Figure 4: everything else equal, whenever the short-term rate is high relative to its permanent component, the spread is low on this account. We can undertake

a similar decomposition of the variance of the spread to those used above,

$$\begin{aligned} \text{var}(S_t) &= \text{var}(R_t^L - \bar{R}_t^L) + \text{var}(R_t - \bar{R}) \\ &\quad - 2 * \text{cov}((R_t^L - \bar{R}_t^L), (R_t - \bar{R})) \\ 1.93 &= 0.18 + 0.98 - 2 * (-0.38). \end{aligned}$$

According to this expression, there is a variance of 1.93. Of this, 51 percent is attributable to the variability of the temporary component of the short rate, 9 percent is attributable to the temporary component of the long rate, and a substantial amount (39 percent) is attributable to the covariance between these two expressions.²⁷

Simple Rules of Thumb

Suppose that we observe just the change in the long rate and want to know how much of a change has taken place in the permanent component. Our variance decompositions let us provide an answer to this and related questions below. Specifically, we derive a simple rule of thumb as follows. First, define the change in the permanent component as an unobserved zero-mean variable Y_t . This variable is known to be connected to the observed zero-mean variables ΔR_t^L according to the identity $Y_t = \Delta R_t^L + U_t$, where U_t is an error. Then we can ask the question: What is the optimal linear estimate of Y_t given the observed series ΔR_t^L ? To calculate this measure, $\hat{Y}_t = b\Delta R_t^L$, we minimize the expected squared errors, $\text{var}(Y_t - \hat{Y}_t) = \text{var}(Y_t) + b^2\text{var}(\Delta R_t^L) - 2bcov(Y_t, \Delta R_t^L)$. The optimal value of b is the familiar OLS regression coefficient

$$b = \frac{\text{cov}(Y_t, \Delta R_t^L)}{\text{var}(\Delta R_t^L)}.$$

Using our estimates of the common stochastic trend, we compute that the variance of long-rate changes is 0.0826 and that the covariance of long-rate and permanent component changes is 0.0802 (see second panel of Table 5). Thus, the coefficient b takes on a value of 0.97, which leads to the following rule of thumb.

Long-rate rule of thumb: *If a 1 percent rise (fall) in the long rate occurs, then our calculations suggest that an observer should increase (decrease) his or her estimate of the permanent component by 97 percent of this rise (fall).*²⁸

²⁷ There is also substantial serial correlation in the spread, as well as in the temporary components of the short rate and the long rate. The first order autocorrelations of these series are, respectively, 0.81, 0.72, and 0.93.

²⁸ Of course, we could have devised a similar rule of thumb for the short rate by replacing ΔR_t^L by ΔR_t in the formula for the coefficient b . The result would have been a much more

A similar rule of thumb can be derived by linking changes in the unobserved temporary component of the short rate ($R_t - \bar{R}_t$) to the spread.²⁹

Spread rule of thumb #1: *If the spread exceeds its mean by 1 percent, then our estimates suggest that the temporary component of short-term interest rates is low by -0.71 percent ($-0.71 = (-1.37)/(1.93)$).*

Our two rules of thumb indicate that changes in the long rate are dominated by changes in the permanent component and the level of the spread (relative to its mean) is substantially influenced by the temporary component.

4. RATIONAL EXPECTATIONS TESTS

A hallmark of rational expectations models of the term structure, stressed by Sargent (1979), is that they impose testable cross-equation restrictions on linear time series models. In this section, we describe the strategy behind rational expectations tests along the lines of Sargent (1979) and Campbell and Shiller (1987); we also discuss how to extend the tests to accommodate time-varying term premia. We then implement these tests and find that there is a broad rejection of the rational expectations restrictions that we trace to divergent forecastability of the spread and changes in short-term interest rates.

A Simple Reference Model

To illustrate the nature of the cross-equation restrictions that the expectations theory imposes and to motivate the ensuing discussion of rational expectations tests, consider the following simple model. Suppose that the short-term interest rate is governed by

$$R_t = \tau_t + x_t,$$

where τ_t is a relatively persistent permanent component that we model as a unit root process and x_t is a relatively less persistent temporary component. In addition, suppose that agents observe τ_t and x_t separately and also understand that these evolve according to

$$\begin{aligned}\tau_t &= \tau_{t-1} + e_{\tau,t} \\ x_t &= \rho x_{t-1} + e_{x,t},\end{aligned}$$

with $-1 < \rho < 1$ and with $e_{\tau,t}, e_{x,t}$ being white noises. Suppose also that the expectations theory holds true. Using equation (2) and setting (1 –

modest rule of thumb coefficient ($0.1651 = 0.1083/0.6559$). This smaller coefficient reflects the fact that temporary variations are much more important for the short rate.

²⁹ For this purpose, we interpret Y_t as the change in the temporary component of the short rate and replace ΔR_t^L with the spread (less its mean) in the above formula for b . Based on the third panel of Table 5, the covariance between changes in the temporary component of the short rate and the spread equals -1.37 and the variance for changes in the spread is 1.93 .

$\beta) \sum_{j=0}^{\infty} \beta^j E_t k_{t+j} = K = 0$ for all t , the dynamics of the long rate can thus be described as³⁰

$$\begin{aligned} R_t^L &= (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t R_{t+j} \\ &= (1 - \beta) \sum_{j=0}^{\infty} \beta^j E_t [\tau_{t+j} + x_{t+j}] = \tau_t + \theta x_t, \end{aligned} \quad (10)$$

where $\theta = (1 - \beta)/(1 - \beta\rho) < 1$ since $\rho < 1$ as in Section 2 above. Finally, notice that the spread by definition takes the form

$$S_t = R_t^L - R_t = (\theta - 1)x_t,$$

which implies that under the expectations theory, the spread is a perfect negative indicator of the temporary component of short-term interest rates.

Cross-equation restrictions on a stationary VAR system: By assuming a unit root component τ_t in the short rate and the expectations theory being true, we determined above that both the short rate and the long rate in our reference model are stationary in first differences rather than levels. We therefore follow Campbell and Shiller (1987) and study the bivariate system in short-rate changes,

$$\begin{aligned} \Delta R_t &= \Delta \tau_t + \Delta x_t = e_{\tau,t} + e_{x,t} + (\rho - 1)x_{t-1} \\ &= e_{\tau,t} + e_{x,t} + \frac{\rho - 1}{\theta - 1} S_{t-1} = e_{\tau,t} + e_{x,t} + \frac{1 - \beta\rho}{\beta} S_{t-1}, \end{aligned}$$

and in the spread,

$$S_t = (\theta - 1)x_t = \rho S_{t-1} + (\theta - 1)e_{x,t}.$$

Both of these variables are stationary, which has the advantage that testable restrictions are easier to develop in the presence of time-varying, but stationary, term premia.³¹

As stressed by Sargent (1979), the expectations theory imposes cross-equation restrictions. In the case of ΔR_t and S_t , these restrictions become immediately apparent when we compare the two model equations above to an unrestricted bivariate, first order vector autoregression:

$$\begin{aligned} \Delta R_t &= a \Delta R_{t-1} + b S_{t-1} + e_{\Delta R,t}. \\ S_t &= c \Delta R_{t-1} + d S_{t-1} + e_{S,t}. \end{aligned}$$

³⁰ According to the expectations theory, K does not have to equal zero. For the sake of convenience, we set $K = 0$, which can be reconciled with the data if we consider all variables as deviations from their respective means.

³¹ Such a stationary system is sometimes called a VECM in Phillips's triangular form. See Hamilton (1994, 576–78) and Appendix C.

In particular, we see that the expectations theory imposes $a = c = 0$, $b = (1 - \beta\rho)/\beta$, $d = \rho$, and $e_{\Delta R,t} = e_{\tau,t} + e_{x,t}$, $e_{S,t} = (\theta - 1)e_{x,t}$.³² In our econometric analysis below, we will focus on deriving and testing similar restrictions for a more general rational expectations framework that contains the assumption of agents having more information than the econometrician.³³

Restrictions on a VAR Model

For the purpose of testing the cross-equation restrictions in the data, we adopt a general strategy initially put forth by Sargent (1979). Following Campbell and Shiller (1987), we consider a bivariate VAR in the short-rate change and the spread:³⁴

$$\Delta R_t = \sum_{i=1}^p a_i \Delta R_{t-i} + \sum_{i=1}^p b_i S_{t-i} + e_{\Delta R,t}. \quad (11)$$

$$S_t = \sum_{i=1}^p c_i \Delta R_{t-i} + \sum_{i=1}^p d_i S_{t-i} + e_{S,t}. \quad (12)$$

In this section, we work under the assumption that the expectations theory is exactly true, which we relax later. Under this condition, term premia are

³² VECM regressions like (7) and (8) in the previous section are also restricted by the expectations theory. According to our simple model, the dynamics of short- and long-rate changes take the form

$$\begin{aligned} \Delta R_t &= e_{P_t} + e_{T_t} + \frac{1 - \beta\rho}{\beta} S_{t-1}, \\ \Delta R_{L,t} &= \Delta \tau_t + \theta \Delta x_t = e_{\tau,t} + \theta e_{x,t} + \frac{1 - \beta}{\beta} S_{t-1}. \end{aligned}$$

The second equation for the long-rate change is simply the efficient markets restriction.

³³ In our simple model, the VECM approach (discussed in the previous footnote) helps to correctly uncover some features of the data that are not known *a priori* by the econometrician. First, the temporary component x_t of the short rate is reflected in a temporary component of the long rate, but with a much dampened magnitude for plausible values of β and ρ . For example, if $1/\beta = 1.005$ and $\rho = 0.8$, then the composite coefficient θ takes on a value of $0.005/0.025 = 0.2$. Second, the spread is predicted to be a significant predictive variable for interest rates in the VECM, but especially for the temporary component of interest rates. These features of the model appear broadly in accord with the estimated VECM and its outputs, particularly in terms of the implication that there is a much smaller volatility of the temporary component of the long rate than the temporary component of the short rate. In addition, the generally poor predictive performance for changes in the long rate seems consistent with the importance of permanent shocks in that equation, relative to the small effect of the spread. Finally, the spread and the temporary component of the short-term interest rate are negatively associated in the example as in the outputs of the VECM. But other features of the model are at variance with the results obtained via estimating a VECM. In particular, the temporary component of the long rate has a strong positive association with the temporary component of the short rate in the model, while there is a negative correlation in the estimates discussed in the preceding section.

³⁴ The example we discussed above used one lag for analytical convenience, but in this empirical context we use multiple lags to capture the dynamic interactions between the variables more completely.

constant and the expression for the spread in (4) reduces to³⁵

$$S_t = \sum_{j=1}^{\infty} \beta^j E_t \Delta R_{t+j}, \quad (13)$$

as we saw in Section 3 above. This expression is important for two reasons. First, it says that according to the expectations theory the spread is simply the discounted sum of future expected short-rate changes. Second, in terms of econometrics, it reveals that as long as short rates are stationary in first differences, the spread must be stationary as well.

The derivation of testable restrictions that (13) imposes on (11) and (12) has four key ingredients. First, the *law of iterated expectations* implies that for any information set ω_t which is a subset of the market's information set Ω_t ,

$$E[E_t \Delta R_{t+j} | \omega_t] = E[E \Delta R_{t+j} | \Omega_t] | \omega_t = E[\Delta R_{t+j} | \omega_t].$$

Practically, this says that an econometrician's best estimate of market expectations of future short-rate changes, given a data set ω_t , is equal to the econometrician's forecast of these short-rate changes given his or her data. Thus, under the assumption that the expectations theory is exactly true and using the fact that the current spread is in the information set, we can rewrite (13) as

$$S_t = \sum_{j=1}^{\infty} \beta^j E[\Delta R_{t+j} | \omega_t]$$

so that the spread formula is unchanged when the information set is reduced.³⁶

Second, the *Wold decomposition theorem* guarantees that if ΔR_t is stationary, it can be well described by a vector autoregression (possibly of infinite order p) where the explanatory variables are composed of information Ω_{t-1} available to the market at date $t - 1$.

³⁵ Note that we have dropped the constant K from the equation for the sake of notational simplicity. In econometric terms, this simply means that, without a loss of generality, we have to test the expectations theory with demeaned data.

³⁶ As Campbell and Shiller (1987) stress, the explanation for this result is subtle: the expectations theory says that the spread is simply the discounted sum of future expected short-rate changes. Under the null that the theory is true, all the relevant information that market participants use to forecast future short-rate changes must by definition be embodied in the actual spread. As long as S_t is part of the econometrician's information set ω_t , it must thus be the case that $E[\sum_{j=1}^{\infty} \beta^j \Delta R_{t+j} | \Omega_t] = E[\sum_{j=1}^{\infty} \beta^j \Delta R_{t+j} | \omega_t]$. It is important to note that this result is conditional on the expectations theory holding exactly. If we relax the null to allow for time-varying term premia or even a simple error term, S_t no longer embodies all necessary information about expected future short-rate changes.

Third, since we want to derive restrictions on the bivariate system composed of (11) and (12), we define the data set ω_t as p lags of ΔR and S each.³⁷ The econometrician's best linear one-period forecast of short-rate changes thus becomes $E[\Delta R_{t+1}|\omega_t] = h_{\Delta R}E[\omega_{t+1}|\omega_t] = h_{\Delta R}M\omega_t$, where $h_{\Delta R}$ is a selection vector equaling $[1\ 0\ \dots\ 0]$ and where M is the companion matrix corresponding to (11) and (12), written in first order form as $\omega_t = M\omega_{t-1} + e_t$; i.e.:

$$\begin{bmatrix} \Delta R_t \\ \dots \\ \Delta R_{t-p+1} \\ S_t \\ \dots \\ S_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & \dots & a_p & b_1 & \dots & b_p \\ 1 & & & & & \\ & & & & & \\ c_1 & \dots & c_p & d_1 & \dots & d_p \\ \dots & & & 1 & & \\ & & & & & \dots \end{bmatrix} \begin{bmatrix} \Delta R_{t-1} \\ \dots \\ \Delta R_{t-p} \\ S_{t-1} \\ \dots \\ S_{t-p} \end{bmatrix} + \begin{bmatrix} e_{\Delta R,t} \\ \dots \\ 0 \\ e_{S,t} \\ \dots \\ 0 \end{bmatrix}, \quad (14)$$

Fourth, given $\omega_t = M\omega_{t-1} + e_t$, *multi-period linear predictions* of short-rate changes are easy to form:

$$E[\Delta R_{t+j}|\omega_t] = h_{\Delta R}M^j\omega_t.$$

Mapping these forecasts into $S_t = \sum_{j=1}^{\infty} \beta^j E[\Delta R_{t+j}|\omega_t]$ and expressing $S_t = h_S\omega_t$ where h_S is a selection vector with a one in the position corresponding to the spread and zeros elsewhere, we finally derive:

$$h_S\omega_t = \sum_{j=1}^{\infty} \beta^j h_{\Delta R}M^j\omega_t = h_{\Delta R}M[I - \beta M]^{-1}\omega_t,$$

or equivalently:

$$h_S = h_{\Delta R}\beta M[I - \beta M]^{-1}. \quad (15)$$

Expression (15) represents a set of $2p$ cross-equation restrictions that the expectations theory imposes on the bivariate VAR system and that are sometimes called the hallmark of rational expectations models. Specifically, (11) and (12) contain $4p$ parameters $\{a_i\}_{i=1}^p$, $\{b_i\}_{i=1}^p$, $\{c_i\}_{i=1}^p$ and $\{d_i\}_{i=1}^p$. However, under the null that the expectations theory holds true, only $2p$ of these parameters are free while the remaining half is constrained by the cross-equation restrictions in (15).³⁸

Working with the same vector autoregression in short-rate changes and the spread, Campbell and Shiller (1987) test such rational expectations restrictions on U.S. data between 1959 and 1983 by means of a Wald test and conclude

³⁷ This restriction to the past history of interest rates follows Sargent (1979) and Campbell and Shiller (1987). It would be of some interest to explore the implications of adding other macroeconomic variables.

³⁸ As Campbell and Shiller (1987) note, the cross-equation (15) can be simplified to a linear set of restrictions. Specifically, we can rewrite them as $h_S[I - \beta M] = h_{\Delta R}\beta M$, which implies that $a_i = -c_i$ for $i = 1, \dots, p$; $d_1 = 1/\beta - b_1$; and $b_i = -d_i$.

Table 6 VAR Tests of the Expectations Hypothesis

Full Sample Estimates (1951.4–2001.11)				
	ΔR_t		S_t	
	unconstrained VAR	VAR consistent with ET	unconstrained VAR	VAR consistent with ET
ΔR_{t-1}	0.5782 (0.1095)	0.5739 (0.1088)	-0.4927 (0.1171)	-0.5739 (0.1088)
ΔR_{t-2}	0.4580 (0.1124)	0.4604 (0.1116)	-0.5059 (0.1201)	-0.4604 (0.1116)
ΔR_{t-3}	0.2192 (0.1125)	0.2268 (0.1117)	-0.3701 (0.1202)	-0.2268 (0.1117)
ΔR_{t-4}	-0.0447 (0.0379)	-0.0464 (0.0377)	0.0767 (0.0405)	0.0464 (0.0377)
S_{t-1}	0.9254 (0.1021)	0.9218 (0.1014)	0.1507 (0.1091)	0.0838 (0.1014)
S_{t-2}	-0.2228 (0.1542)	-0.2159 (0.1532)	0.0875 (0.1649)	0.2159 (0.1532)
S_{t-3}	-0.4233 (0.1552)	-0.4184 (0.1541)	0.3263 (0.1659)	0.4184 (0.1541)
S_{t-4}	-0.1693 (0.1104)	-0.1761 (0.1096)	0.3023 (0.1180)	0.1761 (0.1096)

Notes: All variables represent deviations from their respective means. Numbers in parentheses represent standard errors. The likelihood ratio test of the unconstrained VAR against the VAR consistent with the expectations theory (ET) is 35.7131. Since the corresponding critical 0.1 percent χ^2 value for 8 degrees of freedom is only 26.1, the restrictions imposed by the ET are strongly rejected.

that the expectations theory is strongly rejected. Alternatively, Sargent (1979) advocates assessing the expectations theory by means of a likelihood ratio test with an asymptotic chi-square distribution, which is the approach that we follow here. The likelihood ratio is $2[L_{UVAR} - L_{ETVAR}]$, that is, the difference between the log likelihood values of the unrestricted VAR and the VAR subject to the restriction in (15), respectively. For a given significance level, the restrictions are then rejected if the likelihood ratio is larger than the critical chi-square value for $2p$ degrees of freedom.

Table 6 reports the unrestricted and the restricted VAR estimates for our 1951–2001 sample using our reference lag length of $p = 4$.³⁹ Remarkably, none of the restricted point estimates differ by more than two standard errors from their unrestricted counterparts.⁴⁰ However, the computed likelihood ratio of 35.71 is larger than the critical 0.1 percent chi-square value of 26.1. Our data set thus comfortably rejects the restrictions imposed by the expectations theory, confirming Campbell and Shiller’s result over a substantially longer time period and using a more appropriate testing procedure.⁴¹

Time-Varying Term Premia

The restrictions in (15) are derived from the strong assumption that the expectations theory is exactly true up to term premia that are constant through time, which precludes even measurement error in the spread. Alternatively, we can adapt the testing approach discussed above and derive testable restrictions that allow for certain forms of time-variation in the term premia. To this end, reconsider the general formula (4) that links the long rate to the present value of future expected short rates and the expected term premia. Without imposing any restrictions, the spread can thus be expressed as the sum of two unobserved components:

$$S_t = F_t + K_t, \quad (16)$$

where $F_t = \sum_{j=1}^{\infty} \beta^j E[\Delta R_{t+j} | \Omega_t]$ and $K_t = (1 - \beta) \sum_{j=0}^{\infty} \beta^j E[k_{t+j} | \Omega_t]$ denote the present value of the market’s expectations about future short-rate changes and term premia, respectively. Combining this expression with the VAR framework $\omega_t = M\omega_{t-1} + e_t$, we can rewrite (16) as

$$S_t = E[F_t | \omega_t] + K_t + \xi_t,$$

where $\xi_t = F_t - E[F_t | \omega_t] = \sum_{j=1}^{\infty} \beta^j \{E[\Delta R_{t+j} | \Omega_t] - E[\Delta R_{t+j} | \omega_t]\}$ is the error arising from the fact that the econometrician is using a smaller data set than the market to forecast future short-rate changes.⁴² Equivalently, we can

³⁹ The reported results hold true for alternative lag lengths as well.

⁴⁰ Because of the specific linear nature of the cross-equation restrictions noted above, the constraint estimates and the standard errors for different pairs of VAR coefficients are identical.

⁴¹ Bekaert and Hodrick (2001) show that in the context of cross-equation restrictions tests of present-value models such as the expectations theory, Wald tests suffer from substantially larger sample biases than likelihood ratio tests or Lagrangean multiplier tests.

⁴² As noted in a previous footnote, under the null that the expectations theory holds, S_t embodies all necessary information about future short-rate changes, and thus $E[\Delta R_{t+j} | \Omega_t] = E[\Delta R_{t+j} | \omega_t]$ as long as S_t is part of ω_t . However, since now we have relaxed the assumption of constant term premia (i.e., the expectations theory does not hold), we can no longer assume that S_t contains all necessary information about future short-rate changes. This means that replacing the market’s information set Ω_t with the econometrician’s information set $\omega_t \subset \Omega_t$ (potentially) introduces a forecasting error.

Table 7 VAR Tests Based on Lagged Information

Full Sample Estimates (1951.4–2001.11)	
Information Lag	Likelihood Ratio (between unconstrained and constrained VAR)
0	35.7131
1	32.8594
3	33.6881
6	33.6300
12	35.6203

form expectations conditional on data ω_{t-l} :

$$E[S_t|\omega_{t-l}] = E[F_t|\omega_{t-l}] + E[K_t|\omega_{t-l}], \quad (17)$$

where we recognize that $E[\xi_t|\omega_{t-l}] = 0$ since ξ_t is uncorrelated by construction with any information in ω_{t-l} .

Finally, we impose that the term premia K_t is unforecastable from information ω_{t-l} , that is, $E[K_t|\omega_{t-l}] = 0$. Under this assumption, which is weaker than the assumption $K_t = 0$ employed in the tests of the expectations theory discussed earlier, we obtain the following testable restrictions:

$$h_S M^l = h_{\Delta R} \beta M [I - \beta M]^{-1} M^l, \quad (18)$$

where we used the same arguments as above to rewrite $E[S_t|\omega_{t-l}] = h_S M^l \omega_{t-l}$ and $E[F_t|\omega_{t-l}] = h_{\Delta R} \beta M [I - \beta M]^{-1} M^l \omega_{t-l}$.⁴³ This strategy is suggested by the fact that Sargent (1979) actually tests the expectations theory by considering such a relaxed form of the cross-equation restrictions with $l = 1$ (i.e., a one-period lag in the information set).

The restrictions in (18) can be evaluated using a likelihood ratio test similar to that used above, which compares the fit of the constrained and unconstrained vector autoregressions. Because of the assumed stationarity of the joint process for spreads and short-rate changes, the eigenvalues of the companion matrix M are all smaller than one in absolute value. It must be the case, then, that the restrictions are satisfied as l becomes very large, since both sides of the equation contain only zeros in the limit. However, restrictions of the form of (18) are valid and interesting so long as the researcher is willing to assume that term premia are unforecastable at some intermediate horizon.

⁴³ It might appear that one could “divide out” the terms M^l from both sides of (18), restoring the restrictions (15). However, the matrix M can be shown to be singular if $E[K_t|\omega_{t-l}] = 0$ is true (Kurmann [2002a]).

Table 7 reports likelihood ratios of the unrestricted VAR against the VAR subject to the restrictions in (18) for the forecasting horizons $l = 1, 3, 6,$ and 12 .⁴⁴ Notably, the restrictions are rejected for all of these lags. Thus, while the cointegration tests of Section 3 indicate that variations in the term premia are stationary, the results of Table 7 show that departures from the expectations theory are not only due to high-frequency deviations but also occur at intermediate, business cycle frequencies.

5. EXPECTATIONS AND THE SPREAD

The preceding section illustrates that the cross-equation restrictions implied by the expectations theory are soundly rejected, even when we allow for some limited time-variation in the term premia. However, as Campbell and Shiller (1987) argue, statistical tests of the cross-equation restrictions may be “highly sensitive to deviations from the expectations theory—so sensitive, in fact, that they may obscure some of the merits.”⁴⁵ In other words, even if the theory is not strictly true, it may contain important elements of the truth. This section builds on the ingenious approach of Campbell and Shiller (1987) in computing an estimate of the expectations component of the spread—which they call a “theoretical spread”—in order to shed more light on this issue. This approach also permits us to (i) extract an estimate of the term premium and (ii) to derive a rule of thumb linking the observed spread to unobserved expectations concerning temporary variations in the short-term interest rate.

Decomposing the Spread in Theory

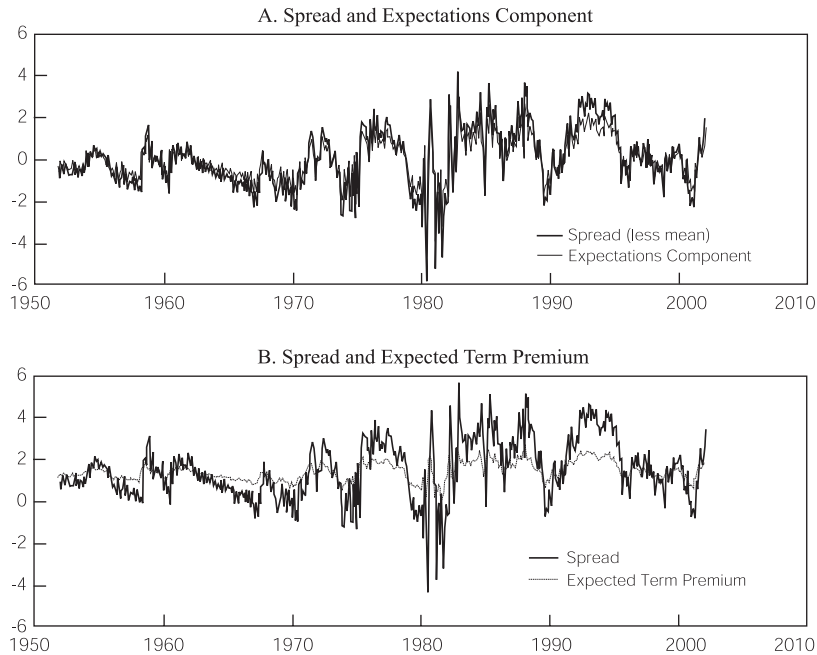
Our discussion above stresses that the observed spread is the sum of two unobserved components, $S_t = F_t + K_t$, which we call the expectations and term premium components. From (17) above, we know that the spread conditional on the econometrician’s information set ω_{t-l} can be written as:

$$E[S_t|\omega_{t-l}] = E[F_t|\omega_{t-l}] + E[K_t|\omega_{t-l}].$$

Under the expectations theory, we assumed that $E[K_t|\omega_{t-l}]$ is constant (or zero in deviations from the mean). In this section, we alternatively calculate an estimate of the expectations component given an information set and compare it to the prediction of the spread conditional on that same information set. From our results above, we know that the expectations component can be formed as $E[F_t|\omega_{t-l}] = \sum_{j=1}^{\infty} \beta^j E[\Delta R_{t+j}|\omega_{t-l}] = h_{\Delta R} \beta M [I - \beta M]^{-1} M^l \omega_{t-l}$, and

⁴⁴ The variables in the information set ω_{t-l} remain the same as for the cross-equation restriction tests above (i.e., ω consists of lags of ΔR and S). However, it would be interesting to assess the robustness of the reported results if we included additional variables that are likely to help forecast changes in the short rate.

⁴⁵ Campbell and Shiller (1987, 1080).

Figure 5 Spread, Expectations, and Term Premia

we also know that the predicted spread can be calculated as $E[S_t|\omega_{t-l}] = h_S M^l \omega_{t-l}$. In these formulas, the coefficients from an *unrestricted* VAR are used to provide the elements of the matrix M that are relevant to forecasting. The difference between the two expressions, $E[K_t|\omega_{t-l}] = E[S_t|\omega_{t-l}] - E[F_t|\omega_{t-l}]$, is an implied variation in the term premium.

Decomposing the Spread in Practice

In view of the results from the prior section, we calculate two decompositions of the spread, based on different information sets.

Current information: We begin by calculating an estimate of the expectations component and the residual term premium using current information ω_t . In this setting, which corresponds to the analysis of Campbell and Shiller (1987), $E[S_t|\omega_t]$ simply equals the actual spread and $E[F_t|\omega_t] = h_{\Delta R} \beta M [I - \beta M]^{-1} \omega_t$.

Panel A of Figure 5 shows that the expectations component (the spread under the expectations theory) is strongly positively correlated with the actual spread (correlation coefficient = 0.99) and displays substantial variability. Panel B of Figure 5 shows the spread and the term premium (the gap between

Table 8 Summary Statistics for Expectations Component/Term Premium Decomposition

Full Sample Estimates (1951.4–2001.11)		
A. Based on Current Information		
Spread	Expectations	Term Premium
1.9318	1.3339	0.5979
0.9923	0.9355	0.3984
0.9633	0.9225	0.1994
B. Based on 6-months Forecasts		
Spread	Expectations	Term Premium
0.6495	0.4264	0.2231
0.9998	0.2800	0.1464
0.9995	0.9987	0.0767

Notes: Statistics correspond to Figures 5 and 6. Each panel contains a 3 by 3 matrix. On the diagonal, variances are reported (e.g., the variance of 6-months forecasts of the spread is 0.6495). Above the diagonal, covariances are listed (e.g., the covariance between the spread and expectations in the current information case is 1.3339). Below the diagonal, the corresponding correlation is reported (e.g., the correlation between the spread and expectations in the current information case is 0.9923).

the spread and the expectations component). The residual term premium is much less variable.

It is useful to consider a decomposition of variance for the spread, similar to that which we used for permanent and temporary components in Section 3:

$$\begin{aligned} \text{var}(S_t) &= \text{var}(F_t|\omega_t) + \text{var}(K_t|\omega_t) + 2 * \text{cov}(F_t|\omega_t, K_t|\omega_t) \\ 1.93 &= 0.94 + 0.20 + 2 * (0.40) \end{aligned}$$

Panel A of Table 8 reports second moments of the spread, the expectations component and the term premia. The variance of the spread is 1.93 (as noted in the derivation of the first spread rule of thumb), while the variance of the expectations component is 0.94. Since their respective standard deviations are not too different (1.39 and 0.97, respectively) and since they are virtually perfectly correlated, it is not surprising that a glance at the first panel of Figure 5 leads one to think that the expectations component explains most of the spread. By contrast, the standard deviation of the estimated term premium is much smaller (0.45), so it is natural to downplay its contribution after glancing at the second panel. But as panel A of Table 8 indicates, there is a very high estimated correlation of changes in the term premium and changes in the expectations component (0.94), so there is a substantial contribution to variability in the spread that arises from the covariance term (0.80 of a total of 1.93).

Economically, the spread appears excessively volatile relative to the estimated expectations component because there is a tendency for periods of high expectations components to occur when the term premium is also high.⁴⁶ Looking back to the first test of rational expectations restrictions, Figure 5 provides insight into why the cross-equation restrictions are rejected, since it highlights the distinct behavior of the spread and the expectations component. The spread contains information about the temporary component of interest rates highlighted by the expectations theory, but there are important departures as well.

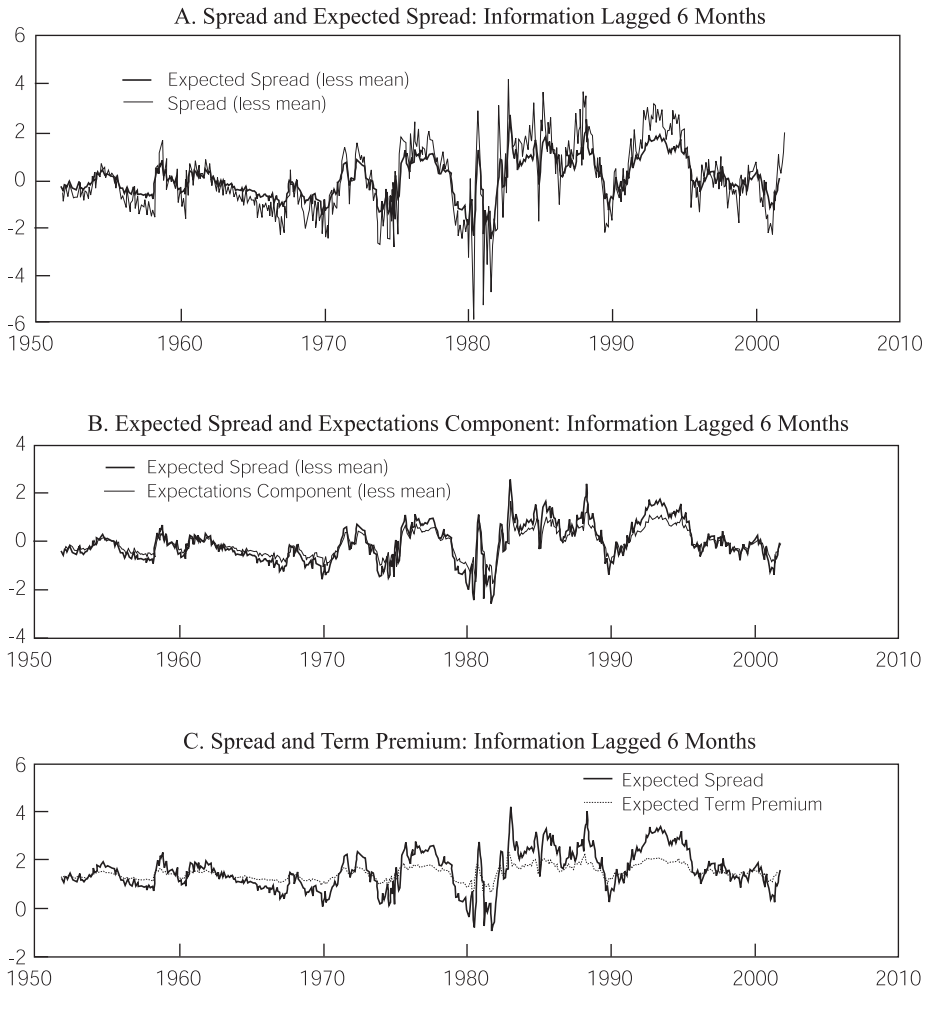
Results based on lagged information: Figure 6 and panel B of Table 8 use forecasts from the vector autoregression, using information six months previous. In panel A of Figure 6, the actual spread S_t and the forecast $ES_t|\omega_{t-6}$ are plotted. While these series move together, the forecasted spread is much less volatile than the actual spread (the variance of the forecasted spread is 0.65, which is about one-third of the actual spread's variance of 1.95). In panel B, the forecasted spread $ES_t|\omega_{t-6}$ and the forecasted expectations component $EF_t|\omega_{t-6}$ are plotted. While the forecasted expectations component is highly correlated with the forecasted spread, it is clearly less volatile as well. In panel C, the forecasted spread $ES_t|\omega_{t-6}$ and the forecasted term premium component $EK_t|\omega_{t-6} = ES_t|\omega_{t-6} - EF_t|\omega_{t-6}$ are plotted. This residual is positively associated with $ES_t|\omega_{t-6}$, with a near-perfect correlation. Its variance (0.076) is also somewhat more than one-third of the variance of the term premium measure $EK_t|\omega_t$ that is shown in Figure 5.

This figure illustrates, we conjecture, why the rational expectations restrictions are rejected when the information set is lagged, as reported previously in Table 7 and discussed in detail above. The deviations of the forecastable part of the spread $ES_t|\omega_{t-6}$ from the forecastable part of the expectations component $EF_t|\omega_{t-6}$ appear important. Indeed, there is some evidence that $EK_t|\omega_{t-6}$ are more serially correlated than either $ES_t|\omega_{t-6}$ or $EF_t|\omega_{t-6}$, as opposed to being unforecastable in the manner required for the rational expectations restrictions to be satisfied.

A Second Rule of Thumb for the Spread

If the spread rises by 1 percent, then how great a rise in the expectations component should an observer infer has occurred? This is a natural question, analogous to one earlier posed for the temporary component of the nominal interest rate, identified via the VECM. Since the variance of the spread is 1.93 and the covariance between the spread and the expectations component is 1.33, the rule

⁴⁶ While these are point estimates and do not take into account uncertainty implied by the fact that the unrestricted VAR is estimated rather than known, preliminary results in Kurmann (2002b) suggest that there may not be too much uncertainty in our context.

Figure 6 Conditioning S, F, and K on Lagged Information

of thumb coefficient is $b = 0.69 = 1.33/1.93$. Hence, we have the following.

Spread rule of thumb #2: *If the spread exceeds its mean by 1 percent, then our estimates suggest that the expectations component is high by 0.69 percent.*

Earlier, we derived a very similar implication—a coefficient of 0.71 but with an opposite rule sign—for the link between the temporary component of the short-term interest rate and the spread. It is not an accident that these two measures are very closely associated. The temporary component of the short-term rate is defined as $R_t - \bar{R}_t$, with $\bar{R}_t = R_{t-1} + \lim_{k \rightarrow \infty} \sum_{j=0}^k E_t \Delta R_{t+k}$. It

is accordingly given by $R_t - \bar{R}_t = -[\lim_{k \rightarrow \infty} \sum_{j=1}^k E_t \Delta R_{t+k}]$. The expectations component studied in this section is $E[F_t | \omega_{t-l}] = \sum_{j=1}^{\infty} \beta^j E_t[\Delta R_{t+j}]$. In each case, the expectations terms are made operational by use of very similar linear forecasting models; there are small differences because β is slightly smaller than one, but the essential theoretical and empirical properties are very similar except for the change in sign.

6. FOCUSING ON RECENT HISTORY

Many studies of recent macroeconomic history document changes in the pace and pattern of macroeconomic activity that have occurred over the past two decades.⁴⁷ Other studies suggest that a major reason for these changes is that the Federal Reserve System has altered its behavior in important ways. For example, Goodfriend (1993) argues that the U.S. monetary policy decision-making came of age—gaining important recognition and credibility—during this period, after having earlier traveled on a wide-ranging odyssey. Accordingly, in this section, we explore how some key features of our previous analysis change if we restrict attention to 1986.7–2001.11. The start date of this period was selected as descriptive of recent U.S. monetary policy with increased credibility, following the narrative history of Goodfriend (2002): it includes the last few years of the Volcker period and the bulk of the Greenspan period. We focus our attention on two sets of issues. First, how did the estimated variability in the permanent component of interest rates change during this period? Second, how did the estimated importance of the expectations effects on the long-short spread change during this period?

The Stochastic Trend in Interest Rates

One important conclusion from our earlier analysis is that there is a common stochastic trend in interest rates, which is closely associated with the long rate. To conduct the analysis for the recent period, we start by reestimating the VECM discussed in Section 3 and reported in Table 4. Then, we calculate the permanent component suggested by this specification, producing the results reported in Figure 7 and Table 9.

We focus on two main results. First, as Figure 7 shows, the stochastic trend continues to be an important contributor to the behavior of both the long-term and short-term interest rates. As in the full sample period, it is closely associated with the long rate. Further, it is much less closely associated with the short rate.

⁴⁷ For example, see Blanchard and Simon (2001) or Stock and Watson (2002).

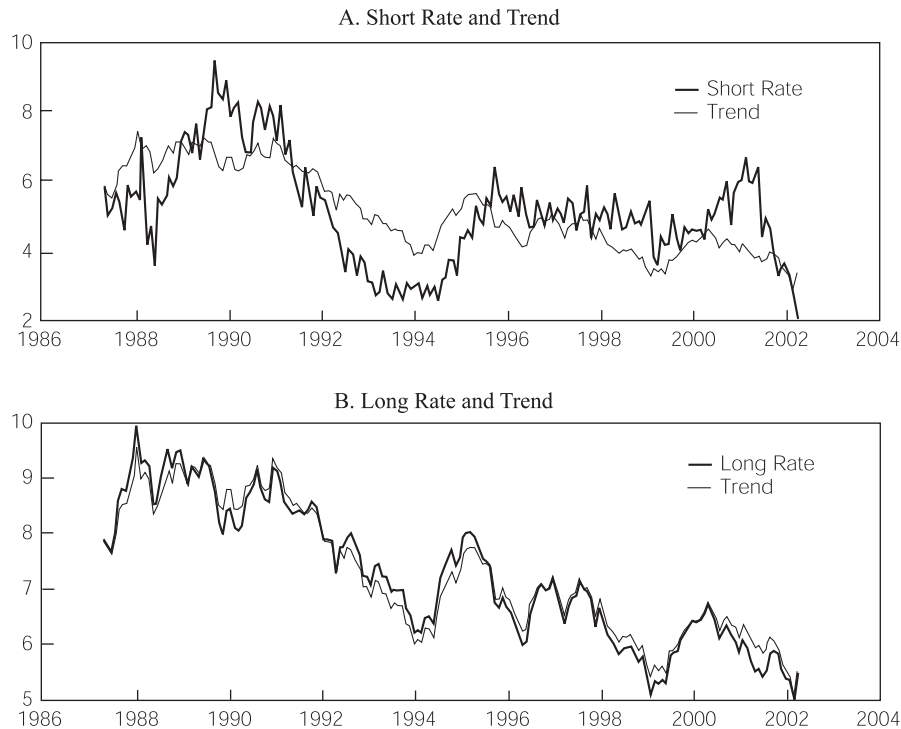
Table 9 Summary Statistics for Two Decompositions

Subsample Estimates (1951.4–2001.11)		
A. Short-Rate Changes		
Total	Permanent	Temporary
0.4409	−0.0003	0.4411
−0.0019	0.0476	−0.0478
0.9501	−0.3137	0.4890
B. Long-Rate Changes		
Total	Permanent	Temporary
0.0636	0.0536	0.0100
0.9747	0.0476	0.0061
0.6314	0.4422	0.0039
C. Spread		
Spread	Expectations	Term Premium
1.8917	1.4574	0.4343
0.9950	1.1341	0.3232
0.9475	0.9108	0.1111

Notes: Each panel contains a 3 by 3 matrix in a manner similar to Tables 5 and 8. On the diagonal, variances are reported. Above the diagonal, covariances are listed. Below the diagonal, the corresponding correlation is reported.

Panel B of Table 9 provides more detail. It shows that changes in the common stochastic trend (permanent component) have a variance of 0.048, which is less than one-half of the comparable variance reported in Table 5. Thus, there is evidence that the stochastic trend is less important for both short-term and long-term interest rates. We can measure this reduced influence on our rule of thumb. Based on the full sample, we calculate that a 1 percent rise in the long rate should bring about a 97 percent rise in the predicted permanent component. On the recent sample, this rule-of-thumb coefficient is smaller: a 1 percent rise should bring about only a 84 percent increase in the predicted permanent component.⁴⁸ Yet, while the effect is smaller, changes in long rate still strongly signal changes in the stochastic trend.

⁴⁸ In terms of elements of Table 9, the rule-of-thumb coefficient is calculated as $b = 0.0536/0.0636 = 0.84$.

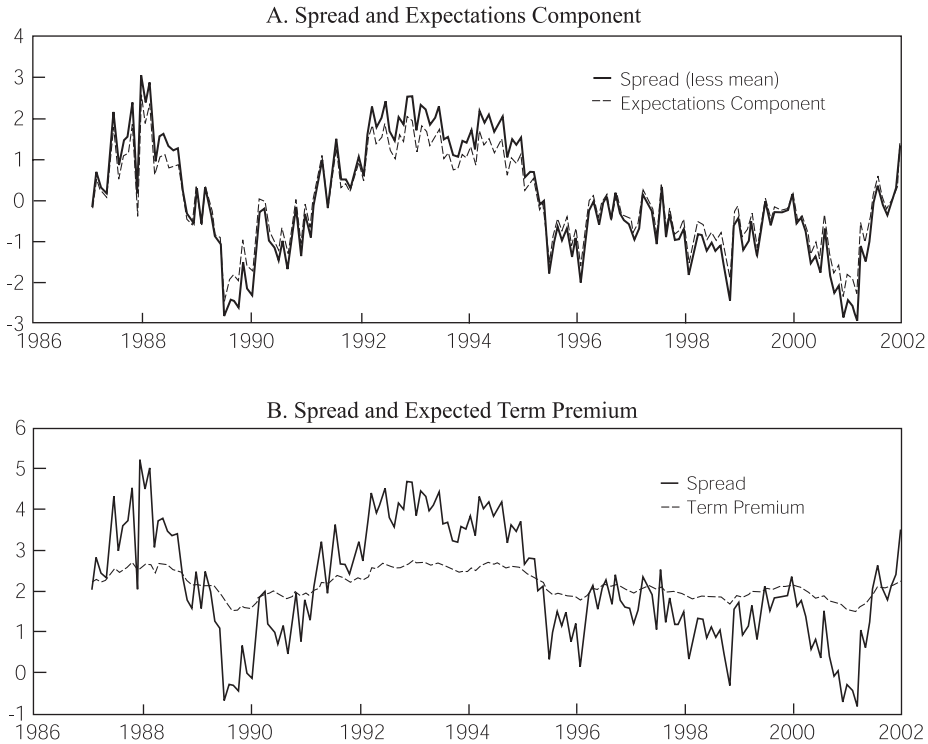
Figure 7 Interest Rates and Stochastic Trend, Once Again

Expectations and the Spread

Another important conclusion of our analysis above is that the spread is an indicator of forecastable temporary variation in short-term interest rates and, in particular, of market expectations of these variations. Figure 8 and panel C of Table 9 show that this relationship has been maintained and, indeed, has apparently gained strength during the recent period. In particular, if we look at rule of thumb #2 for the spread, which indicates the extent to which a high spread should be interpreted as reflecting a high expectations component, then the rule-of-thumb coefficient is $0.77 = 1.46/1.89$ for the recent period, whereas it was only 0.69 for the entire sample period.⁴⁹

In sum, the two reported differences for this more recent period are intriguing, and it is natural to think about possible sources of the change in

⁴⁹ We think that a natural next stage of research involves a more systematic inquiry into the evolving nature of the links between short-term rates and long-term rates. For example, Watson (1999) argues that increased persistence in short-term interest rates—which in our case would involve evolving VAR coefficients—helps explain the increased variability of long-term rates from the 1965–1978 period to the 1985–1998 period. This section, by contrast, argues that the changes in the persistent component in interest rates (the stochastic trend) were less important during 1986–

Figure 8 Spread, Expectations, and Term Premia, Once Again

stochastic properties of the term structure. For example, we might conjecture that the reduced importance of the permanent component is the result of a more credible, inflation-stabilizing monetary policy. Given the lack of structure in our present analysis, however, it is impossible to support such a claim with statistical evidence or to quantify its importance compared to other potential explanations. Rather, we consider that these findings highlight a topic that warrants further investigation.

7. SUMMARY AND CONCLUSIONS

We conclude that expectations about the level of interest rates are very important for the behavior of long-term interest rates on two dimensions. First,

2001 than over the 1951–2001 sample that includes the volatile 1979–1984 period not studied by Watson. A recent attempt to take into account time variations in the VAR parameters is Favero (2001), who computes the long rate under the expectations theory using a rolling regression VAR approach.

changes in the long-term interest rate substantially reflect changes in the permanent component (stochastic trend) in the level of the short-term rate. Second, the spread between long-term and short-term rates depends heavily on a temporary component (deviations from stochastic trend) of the level of short-term rates. Although the strong form of the expectations theory is rejected by a battery of statistical tests, it remains a workable approximation for many applied purposes. Changes in the long rate are largely a signal that the common trend in rates has shifted; a high spread is an important signal that future short rates will rise. More specifically, we provide rules of thumb for interpreting the expectations component of changes in long rates and the level of the long-short spread.

While the expectations theory is rejected, our rational expectations statistical approach is constructive in highlighting the ways in which the linear expectations theory of the term structure fails. The nature of predictable departures from the expectations theory, which we interpreted as time-varying term premia, suggests to us the importance of studying linkages between these factors and the business cycle, since our analysis indicates that these were not simply high frequency deviations.

Finally, the econometric methods that we use are nonstructural, in that they do not take a stand on the specific economic model that determines short-term rates. Nevertheless, the results of our investigation do make some suggestions about the shape that structural models must take, since they indicate the presence of a stochastic trend in the level of the interest rate. Recent research on monetary policy rules, as exemplified by Clarida, Gali, and Gertler (1999), almost invariably assumes that the short-term interest rate is governed by a stable behavioral rule of the central bank, linking it simply to the level of inflation and the level of the output gap, a specification which would preclude such shifts in trend interest rates when incorporated into most macroeconomic models. Our results suggest that a crucial next step in the analysis of monetary policy rules must be the exploration of specifications that can give rise to a stochastic trend in interest rates. In addition, most current macroeconomic models would generally ascribe such shifts in interest rate trends to shifts in inflation trends. Our results thus suggest the importance of an analysis of the interplay between trend inflation, the long-term rate, and monetary policy.

APPENDIX A: THE SHILLER APPROXIMATION

The purpose of this appendix is to derive and exposit Shiller's approximate equation for the yield on a long-term bond. For a coupon bond of arbitrary maturity, N , the yield-to-maturity is the interest rate that makes the price

equal to the present discounted value of its future cash flows $\{C_{t+j}\}$, which may include both coupons and face value:

$$P_t^L = \sum_{j=1}^N \frac{C_{t+j}}{(1 + R_t^L)^j}.$$

In the particular case of a bond with infinite term, which is commonly called a consol, the relationship is

$$P_t^L = \sum_{j=1}^{\infty} \frac{C}{(1 + R_t^L)^j} = \frac{C}{R_t^L}.$$

Between t and $t + 1$, the holding period yield on any coupon bond is given by

$$H_{t+1} = \frac{P_{t+1} + C - P_t}{P_t}.$$

Accordingly, the holding-period yield on a consol is given by

$$H_{t+1} = \frac{(C/R_{t+1}^L) + C - (C/R_t^L)}{(C/R_t^L)} = \frac{(1/R_{t+1}^L) + 1}{(1/R_t^L)} - 1.$$

The ratio R_t^L/R_{t+1}^L is approximately $1 + \theta(R_t^L - E_t R_{t+1}^L)$ via a first order Taylor series approximation about the point $R_t^L = R_{t+1}^L = R^L$, $\theta = 1/R^L$. It then follows that the holding-period yield is approximately

$$H_{t+1} = \theta(R_t^L - R_{t+1}^L) + R_t^L.$$

Notice that small changes in the yield $R_t^L - R_{t+1}^L$ have large implications for the holding-period yield H_{t+1} because θ is a large number. For example, if the annual interest rate is 6 percent and the observation period is one month, then $\theta = 1/(0.005) = 200$. Defining $\beta = 1/(1 + R^L)$, this expression can be written as $H_{t+1} = \frac{1}{1-\beta} R_t^L - \frac{\beta}{1-\beta} R_{t+1}^L$, which is convenient for the discussion below.

Suppose next that this approximate holding-period yield is equated (in expected value) to the short-term interest rate R_t and a term premium k_t . Then, it follows that

$$E_t H_{t+1} = E_t \left[\frac{1}{1-\beta} R_t^L - \frac{\beta}{1-\beta} R_{t+1}^L \right] = R_t + k_t$$

or

$$R_t^L = \beta E_t R_{t+1}^L + (1 - \beta)(R_t + k_t),$$

which is the form used in the main text. This derivation highlights the fact that the linear coefficient β may “drift” over time if the average level of the long rate is very different. It also highlights the fact that this term structure formula is an approximation suitable for very long-term bonds.

APPENDIX B: FORECASTING WITH THE VECM

We estimate a VECM of the form

$$\begin{aligned}\Delta R_t &= a(B)\Delta R_{t-i}^L + b(B)\Delta R_{t-i} + fS_{t-1} + e_{Rt} \\ \Delta R_t^L &= c(B)\Delta R_{t-i}^L + d(B)\Delta R_{t-i} + gS_{t-1} + e_{Lt},\end{aligned}$$

where B is the backshift (lag) operator. We note that this difference between these two equations is

$$\begin{aligned}\Delta S_t &= [c(B) - a(B)]\Delta R_{t-i}^L + [d(B) - b(B)]\Delta R_{t-i} \\ &\quad + (g - f)S_{t-1} + e_{Lt} - e_{Rt},\end{aligned}$$

so that we can write

$$\begin{aligned}S_t &= [c(B) - a(B)]\Delta R_{t-1}^L + [d(B) - b(B)]\Delta R_{t-1} \\ &\quad + (1 + g - f)S_{t-1} + (e_{Lt} - e_{Rt}),\end{aligned}$$

so that it is easy to write the system in state space form defining $x_{t-1} = [\Delta R_{t-1} \ \Delta R_{t-2} \ \dots \ \Delta R_{t-p} \ \Delta R_{t-1}^L \ \Delta R_{t-2}^L \ \Delta R_{t-p}^L \ S_{t-1}]$, which captures all of the predictor variables in these three equations. The main state equation is of the form $x_t = Mx_{t-1} + Ge_t$, with the elements being

$$\begin{aligned}x_t &= \begin{bmatrix} \Delta R_t \\ \dots \\ \Delta R_{t-p+1} \\ \Delta R_t^L \\ \dots \\ \Delta R_{t-p+1}^L \\ S_t \end{bmatrix} \\ M &= \begin{bmatrix} a_1 & \dots & a_p & b_1 & \dots & b_p & 0 \\ 1 & & & \dots & & & 0 \\ & & & \dots & & & \\ c_1 & \dots & c_p & d_1 & \dots & d_p & 0 \\ \dots & & & 1 & & & 0 \\ c_1 - a_1 & \dots & c_p - a_p & d_1 - b_1 & \dots & d_p - a_p & 1 + g - f \end{bmatrix} \\ Ge_t &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \dots & \dots \\ 0 & 1 \\ \dots & \\ -1 & 1 \end{bmatrix} \begin{bmatrix} e_{Rt} \\ e_{Lt} \end{bmatrix}.\end{aligned}$$

APPENDIX C: VARIOUS COINTEGRATED MODELS

In this appendix, we want to demonstrate that the vector autoregression system estimated by Campbell and Shiller (1987) implies a vector error correction model with the cointegrating vector $[1 \ -1]$. The discussion is a specific case of the existence of a Phillips triangular form for a cointegrated system (see Hamilton [1994, 576–78]).

We write the vector error correction model as

$$\begin{aligned}\Delta R_t^L &= a(B)\Delta R_{t-1}^L + b(B)\Delta R_{t-1} + f(R_{t-1}^L - R_{t-1}) + e_{Lt} \\ \Delta R_t &= c(B)\Delta R_{t-1}^L + d(B)\Delta R_{t-1} + g(R_{t-1}^L - R_{t-1}) + e_{Rt},\end{aligned}$$

where B is the back-shift (lag) operator.

We write the VAR system of the CS form as

$$\begin{aligned}S_t &= g(B)S_{t-1} + h(B)\Delta R_{t-1} + e_{St} \\ \Delta R_t &= i(B)S_{t-1} + j(B)\Delta R_{t-1} + e_{Rt}.\end{aligned}$$

Finding the first equation in the VECM: Add the second equation of the VAR to the first, resulting in

$$\begin{aligned}R_t^L - R_{t-1} &= [g(B) + i(B)]S_{t-1} \\ &\quad + [h(B) + i(B)]\Delta R_{t-1} + (e_{St} + e_{Rt}).\end{aligned}$$

Reorganize this as

$$\begin{aligned}R_t^L - R_{t-1}^L &= [1 + g(1) + i(1)]S_{t-1} \\ &\quad + [g(B) - g(1) + i(B) - i(1)]S_{t-1} \\ &\quad + [h(B) + i(B)]\Delta R_{t-1} + (e_{St} + e_{Rt}),\end{aligned}$$

where $g(1)$ is the sum of coefficients in the g polynomial (and similarly for i). Since the coefficients in $[g(B) - g(1)]$ sum to zero by construction, it is always possible to factor $[g(B) - g(1)] = \gamma(B)(1 - B)$ with $\gamma(B)$ having one less lag than $g(B)$. Further, we can similarly write $i(B) - i(1) = \phi(B)(1 - B)$. Hence, we can write the above equation as

$$\begin{aligned}R_t^L - R_{t-1}^L &= [1 + g(1) + i(1)]S_{t-1} \\ &\quad + [\gamma(B) + \phi(B)](\Delta R_{t-1}^L - \Delta R_{t-1}) \\ &\quad + [h(B) + i(B)]\Delta R_{t-1} + (e_{St} + e_{Rt}),\end{aligned}$$

which takes the general form of the VECM equation with suitable definitions of $a(B)$ and $b(B)$.

Finding the second equation in the VECM: Similarly, we can rearrange the second equation above as

$$\Delta R_t = [i(B) - i(1)]S_{t-1} + j(B)\Delta R_{t-1} + i(1)S_{t-1} + e_{Rt}.$$

Hence,

$$\Delta R_t = \phi(B)\Delta R_{t-1}^L + [j(B) - \gamma(B)]\Delta R_{t-1} + i(1)S_{t-1} + e_{Rt},$$

which is the same form as the second equation of the VECM system. Thus, the Campbell-Shiller VAR implies a VECM.

REFERENCES

- Bekaert, Geert, and Robert J. Hodrick. 2001. "Expectations Hypothesis Tests." *Journal of Finance* 56 (August): 1357–94.
- Beveridge, Stephen, and Charles R. Nelson. 1981. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'." *Journal of Monetary Economics* 7 (March): 151–74.
- Blanchard, Olivier, and J. Simon. 2001. "The Long and Large Decline in U.S. Output Volatility." *Brookings Papers on Economic Activity* 1 (March): 135–64.
- Campbell, John Y., and Robert J. Shiller. 1987. "Cointegration and Tests of Present Value Models." *Journal of Political Economy* 95 (October): 1062–88.
- _____. 1991. "Yield Spreads and Interest Rate Movements: A Bird's Eye View." *The Review of Economic Studies* 58 (May): 495–514.
- _____, and Kermit L. Schoenholtz. 1983. "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates." *Brookings Papers on Economic Activity* 1 (March): 173–217.
- Clarida, Richard, Jordi Gali, and Mark Gertler. 1999. "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature* 37 (December): 1661–707.
- Dickey, David N., and Wayne A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49 (June): 1057–72.
- Dotsey, Michael. 1998. "The Predictive Content of the Interest Rate Term Spread for Future Economic Growth." Federal Reserve Bank of Richmond *Economic Quarterly* 84 (Summer): 31–51.
- Engle, Robert F., and Clive W. J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55

- (March): 251-76. Reprinted in *Long-Run Economic Relations: Readings in Cointegration*, ed. Robert F. Engle and Clive W. J. Granger. New York: Oxford University Press, 1991.
- Fama, Eugene F. 1977. *Foundations of finance: portfolio decisions and securities prices*. Blackwell.
- Favero, Carlo A. 2001. "Taylor Rules and the Term Structure." Working paper, IGIER Università L. Bocconi (December).
- Goodfriend, Marvin. 1993. "Monetary Policy Comes of Age: A 20th Century Odyssey." Federal Reserve Bank of Richmond *Economic Quarterly* 79 (Winter): 1–22.
- _____. 2002. "The Phases of U.S. Monetary Policy: 1987 to 2001." Manuscript. Prepared for the Charles Goodhart Festschrift, Bank of England, November 2001, revised January 2002.
- Hetzl, Robert L., and Ralph F. Leach. 2001. "The Treasury-Fed Accord: A New Narrative Account." Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Winter): 57–64.
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Horvath, Michael T. K., and Mark W. Watson. 1995. "Testing for Cointegration When Some of the Cointegrating Vectors are Known." *Econometric Theory* 11 (December): 952–84.
- Ibbotson Associates. 2002. *Stocks, Bonds, Bills, and Inflation 2002 Yearbook*. Chicago: Ibbotson Associates.
- Johansen, Soren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12 (June/Sept.): 231–54. Reprinted in *Long-Run Economic Relations: Readings in Cointegration*, ed. Robert F. Engle and Clive W. J. Granger. New York: Oxford University Press, 1991.
- Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. "The Theory and Practice of Econometrics." New York: Wiley.
- King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81 (September): 819–40.
- Kurmann, André. 2002a. "Maximum Likelihood Estimation of Dynamic Stochastic Theories with An Application to New Keynesian Pricing." Chapter 2, *New Keynesian Price and Cost Dynamics: Theory and Evidence*, Ph.D. diss., University of Virginia.

- _____. 2002b. "Quantifying the Uncertainty about Theoretical Interest Rate Spreads." Working paper, University of Virginia.
- Mankiw, N. Gregory, and Jeffrey A. Miron. 1986. "The Changing Behavior of the Term Structure of Interest Rates." *Quarterly Journal of Economics* 101 (May): 211–28.
- Nelson, Charles R., and G. William Schwert. 1977. "Short-Term Interest Rates as Predictors of Inflation: On Testing the Hypothesis That the Real Rate of Interest Is Constant." *American Economic Review* 67 (June): 478–86.
- Owens, Raymond E., and Roy H. Webb. 2001. "Using the Federal Funds Futures Market to Predict Monetary Policy Actions." Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Spring): 69–77.
- Roll, Richard. 1969. *The Behavior of Interest Rates: An Application of the Efficient Market Model to U.S. Treasury Bills*. New York: Basic Books.
- Sargent, Thomas J. 1979. "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure." *Journal of Monetary Economics* 5 (January): 133–43.
- Shiller, Robert J. 1972. Rational Expectations and the Term Structure of Interest Rates. Ph.D. diss., M.I.T.
- Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* (December): 1097–107.
- _____. 2002. "Has the Business Cycle Changed and Why?" Manuscript. Prepared for NBER Macroeconomics Annual (April).
- Watson, Mark W. 1999. "Explaining the Increased Variability in Long-Term Interest Rates." Federal Reserve Bank of Richmond *Economic Quarterly* 85 (Fall): 71–96.