Japanese Monetary Policy: A Quantity Theory Perspective

Robert L. Hetzel

In the last half century, Japan has exercised an enormously beneficial influence on the world. After the end of World War II, it not only adopted democracy, but it also showed that non-Western countries could combine democracy with attainment of Western living standards. By embracing the use of Western technology and combining it with thrift and hard work, Japan modernized its economy and grew at double-digit rates for two decades in the 1950s and 1960s.

By the early 1970s, Japan took advantage of economies of scale in manufacturing and exported automobiles, steel, ships, and other manufactured goods to the world. By the early 1980s, it aggressively adopted new technology to create a range of electronic consumer goods for export, such as cameras and VCRs. At that point, Japan had also moved beyond importing innovative technology from the West to exporting its own innovations, such as just-in-time manufacturing techniques. More important, the range and quality of goods offered by Japanese firms forced Western companies to remain dynamic and competitive.

Over the post–World War II period, Japan has implemented a variety of monetary regimes. The world can learn valuable economic lessons from Japan by studying its monetary history. Japanese monetary policy divides naturally into two time periods separated by 1987. The first part includes the high inflation of the early 1970s and the establishment of price stability by the mid-1980s. The second part includes the boom-bust episode known as the bubble.[1]

Robert.Hetzel@rich.frb.org. The author benefited from early discussions with Milton Friedman and Allen Meltzer. He especially benefited from the friendly fire of colleagues—Michael Dotsey, Roy Webb, and John Weinberg. The views in this article are solely those of the author and should not be attributed to the Federal Reserve Bank of Richmond, the Federal Reserve System, or any other individuals.

[1] This article uses a quantity theory framework. "A Quantity Theory Framework for Understanding Monetary Policy," available upon request from the author, explains this framework.

The variety of monetary regimes implemented by Japan produced the results shown in Figures 1, 2, and 3. Figure 1 shows quarterly observations of four-quarter percentage changes in money (M2+CDs) and nominal output (GDP).[2] Figure 2 reproduces the nominal output growth series of Figure 1 and adds real output growth. Figure 3 shows inflation measured by the GDP price deflator. Inflation is the difference between nominal and real output growth. The rest of the article attempts to breathe some life into these data series.

1. BRETTON WOODS

Until the demise of Bretton Woods in early 1973, Japan pegged the foreign exchange value of the yen to the dollar. With a pegged exchange rate and with U.S. prices beyond Japanese control, the Japanese price level had to adjust to achieve balance of payments equilibrium. To maintain ongoing balance in its external accounts, Japanese baseline inflation had to match U.S. inflation.[3] Furthermore, as Japanese goods became more desirable to the rest of the world, their prices had to rise an additional amount. That is, a favorable change in the terms of trade required inflation in Japan beyond what was occurring at the time in the United States.

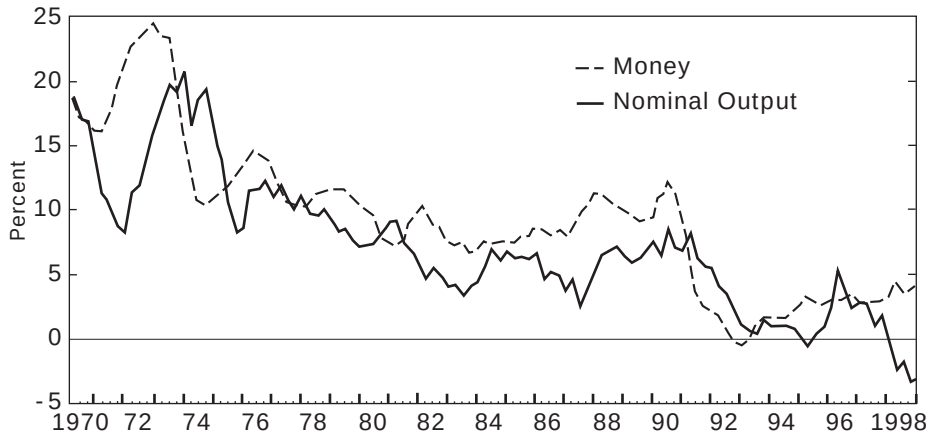
The external constraints imposed by a system of pegged exchange rates required the Bank of Japan (BoJ) to set its discount rate with the objective of targeting the current account balance rather than the state of the domestic economy. At times of current account surpluses, the BoJ lowered its discount rate and money growth and inflation rose. Similarly, at times of current account deficits, the BoJ raised its discount rate and money growth and inflation fell (see Ueda [1997]; Suzuki [1985]; and Yeager [1976]). Changes in economic activity followed changes in money growth with a lag of about three quarters (Yeager 1976, p. 526).

The pegged exchange rate regime served Japan well. In the 1960s, Japan's economy grew rapidly. From 1960 through 1969, Japanese real GDP grew at an annualized rate of 10.4 percent. A major reason for this growth was Japan's ability to mass-produce goods, such as compact automobiles, in big demand by consumers in the West. With its exchange rate pegged at 360 yen to the dollar, the resulting favorable change in the terms of trade required a rise in Japanese prices relative to U.S. prices. From 1961 through 1970, the difference between Japanese and U.S. (GDP deflator) inflation rates was 3 percentage points.

[2] All references to the Japanese money stock are for M2+CDs.

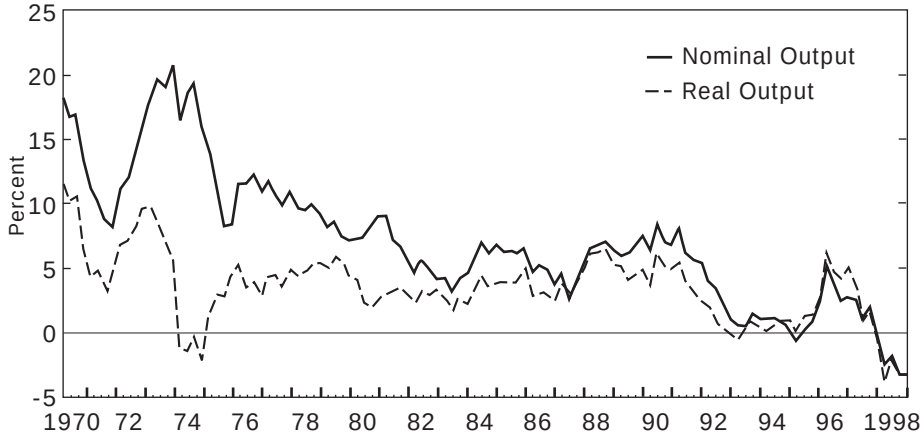
[3] Baseline inflation is for traded goods. High productivity growth in the traded goods sector meant that the prices of services and wages rose faster than the prices of traded goods.

Figure 1 Money and Nominal Output Growth for Japan



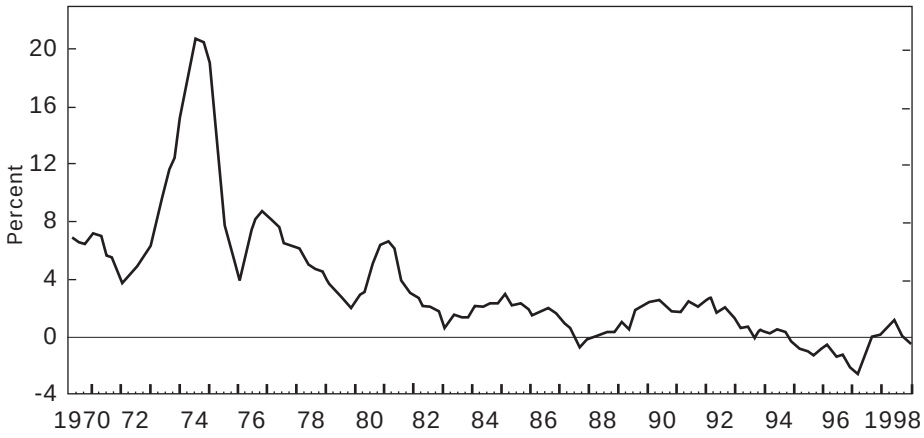
Notes: Quarterly observations of four-quarter percentage changes of money (M2+CDs) and nominal output (GDP). Heavy tick marks indicate fourth quarter of year.

Figure 2 Real and Nominal Output Growth for Japan



Notes: Quarterly observations of four-quarter percentage changes of real and nominal output (GDP). Heavy tick marks indicate fourth quarter of year.

Figure 3 Inflation in Japan



Notes: Quarterly observations of four-quarter percentage changes in prices (GDP price deflator). Heavy tick marks indicate fourth quarter of year.

Even with this difference in inflation rates, the yen was seriously undervalued by 1971. In the first half of the 1960s, the United States achieved near price stability. From 1960 through 1964, inflation rose in the United States at an annualized rate of only 1.4 percent. Then, from 1965 through 1970, U.S. inflation rose at an annualized rate of 4.1 percent. In Japan, on the other hand, inflation rose at an annualized rate of 5.8 percent between 1960 and 1965. Higher inflation in Japan than in the United States reflected the favorable change in Japan's terms of trade. However, from 1966 through 1969, the rate of Japanese inflation did not increase in line with U.S. inflation—in fact, at 5.7 percent, it remained at almost the same prior rate. So by early 1971, the yen was seriously undervalued and the dollar overvalued.

An overvalued dollar forced a change in U.S. policy in summer 1971. Earlier, in 1970, the Federal Reserve (the Fed) lowered short-term interest rates in response to U.S. recession. Capital then flowed out of the United States, further weakening the dollar. In 1971, the U.S. trade balance turned negative—for the first time in the twentieth century. On August 15, 1971, President Nixon announced measures to force a devaluation of the dollar. He closed the gold window to foreign central banks and imposed a 10 percent surcharge on imports into the United States. The surcharge gave the United States a club with which to force foreign countries like Japan to revalue their currencies relative to the dollar.

Figure 4 Yen/Dollar Exchange Rate

Notes: Quarterly observations of yen/dollar exchange rate. Heavy tick marks indicate fourth quarter of year.

Investors knew that foreign currencies could only rise in value relative to the dollar, so they flooded foreign central banks like the BoJ with dollars for conversion into their currencies. The BoJ purchased dollars in an attempt to maintain the existing parity of 360 yen to the dollar. The resulting increase in high-powered money allowed a surge in Japanese money creation starting in 1971Q3, followed by a surge in nominal GDP in 1972 (Figure 1). Money growth peaked at almost 25 percent in early 1973, and nominal GDP growth peaked at almost 20 percent in 1974.

In response to the turbulent events of the year, Japan, along with other member countries of the Bretton Woods system, agreed to a devaluation of the dollar in December 1971. In particular, Japan agreed to a revaluation of the yen relative to the dollar of 16.9 percent, along with a margin of fluctuation of 2 1/4 percent (Figure 4). However, the devaluation was insufficient, and by summer 1972, the yen had risen to its new ceiling. The BoJ then had to resume buying dollars and expanding its domestic money stock. In early February 1973, as inflation rose sharply in the United States, the dollar again weakened. Rather than continue large-scale intervention to support the dollar, the BoJ let the yen float. In early March 1973, the entire Bretton Woods system collapsed.

Japan's experience demonstrated to the countries of the Bretton Woods system (other than the United States) that if they wanted to control their own inflation rates they had to gain control over the rate at which their central

banks created money. They realized that to do so, they would have to allow their currencies to float against the dollar.

2. THE GOLDEN AGE OF JAPANESE MONETARY POLICY

After the system of pegged exchange rates came to an end, the BoJ moved decisively to gain control over money creation and inflation. The three-month Gensaki rate (a market rate on three-month repurchase agreements) went from 5.5 percent in 1973Q1 to 17.5 percent in 1974Q1 (Figure 5). Money growth, which had risen at an annual rate of 25 percent over the two and one-half years ending in 1973Q2, fell sharply to rates between 10 and 15 percent (Figure 1). Inflation (GDP deflator) peaked at about 27 percent (annualized rate) in 1974Q2 but then fell rapidly (Figure 3).

For more than a decade the BoJ concentrated on restoring and maintaining price stability. Suzuki (1985, p. 3) emphasizes the importance the BoJ attached to the behavior of money:[4]

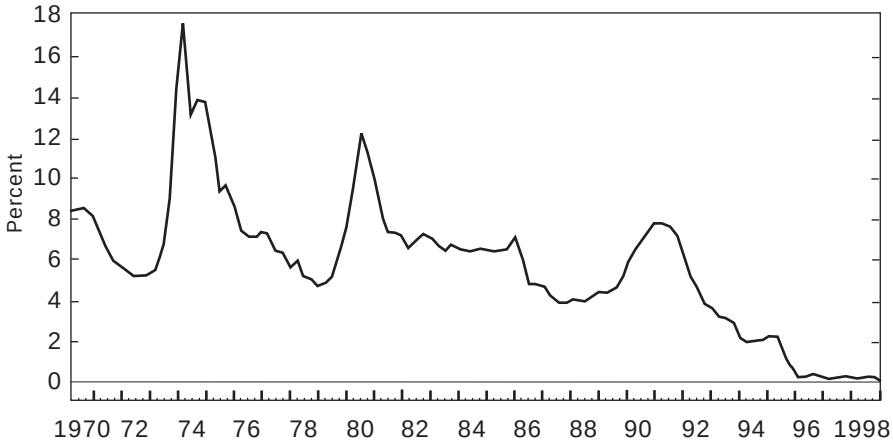
> Since 1975, when the Bank of Japan began to watch the money supply as an intermediate target, fluctuations in the growth rate of money stock have become small. They have remained in a range of $\pm$ 2.5% deviation from the trend line . . . . At the same time, the trend rate of monetary growth itself has been declining gradually, dropping from 15% [in the mid-1970s] to 7-8% recently [1985]. Consequently, the rate of growth of nominal GNP has also declined in the same way. It is interesting to note, however, that the decline in the growth rate of nominal GNP manifested itself as a decline in the inflation rate . . . without affecting the growth rate of real GNP.

Suzuki (1985, p. 6, and 1986, p. 192) also emphasizes that the BoJ viewed moderate money growth as essential to achieve credibility for its policy of lowering inflation. Hutchison (1988), on the basis of empirical estimation of a reaction function for the BoJ, concludes that the BoJ used money (M2+CDs) as an intermediate target.

The BoJ, however, never put in place explicit procedures for monetary control (Ito 1989). Suzuki (1985, p. 6) explains that an explicit target for money would have limited the independence of monetary policy because of political pressure to set the target to conform with the official economic forecast of the government. Further, Suzuki (1984, p. 39, and 1985, p. 5) states that the BoJ considered one to two years as the time period for which monetary control mattered. Moreover, based on an examination of the determinants of changes in the BoJ's discount rate, Ueda (1997, p. 263) writes, ". . . before 1974, monetary policy . . . responded only to the current account, while since then

[4] In the 1980s, Suzuki was Director of the Institute for Monetary and Economic Studies at the Bank of Japan.

Figure 5 Japanese Three-Month Gensaki Rate



Notes: Quarterly observations of three-month government Treasury bill with Repurchase Agreement (Gensaki) rate. Heavy tick marks indicate fourth quarter of year.

it has responded also to inflation and GDP developments." To summarize, in setting its discount rate, the BoJ responded to observed economic conditions rather than attempting to achieve short-term stability of money growth.

During the period that began in the last half of 1973 and continued through the end of 1986, the BoJ basically pursued nominal expenditure targeting using a simple feedback rule for changing short-term interest rates. That is, Japan relied on standard "lean-against-the-wind" procedures where it set its interest rate instrument in response to economic activity. In this way, the BoJ achieved the moderate money growth necessary to control inflation indirectly rather than directly through operationally binding short-term money targets. Nevertheless, money growth acted as a constraint because the BoJ tied the credibility of its goal of restoring price stability to long-run moderation in money growth.

Abandonment of pegged exchange rates in favor of floating exchange rates allowed the BoJ to gain control over the reserve creation of the banking system and ultimately over money creation. The BoJ-engineered reduction in money growth and inflation initiated a profound change in professional and popular views about the nature of inflation. Inflation came to be seen as a monetary rather than a nonmonetary phenomenon. In time, Japan, as well as other countries, assigned responsibility for the control of inflation to their central banks.

The 1970s produced not only high inflation but also stagflation—the simultaneous occurrence of inflation and unemployment. Stagflation provided the empirical impetus for the rise of the rational-expectations school, which

initially emphasized the role expectations play in explaining the interaction of real and monetary phenomena. As exposited by Milton Friedman (1968) and Robert Lucas (1981), the natural-rate/rational-expectations version of the Phillips curve implies that the existing empirical correlation between inflation and the unemployment rate derives from unanticipated, rather than anticipated, changes in inflation. From an inflation rate of 15 percent in 1976, Japan essentially returned to price stability by 1982. The fact that it did so without a recession supported this formulation of the Phillips curve.

3. THE OIL PRICE SHOCKS

In the 1970s, some economists blamed the dramatic surges in inflation that occurred during that period on special factors, especially the two oil price increases engineered by the Organization of Petroleum Exporting Countries (OPEC). The first increase began in October 1973 and the second in December 1978. It is difficult to assess this special-factors explanation of inflation because prior to both oil price shocks, money growth had risen significantly in major industrial countries. When one assigns responsibility for inflation, he must then disentangle the effects of a highly expansionary monetary policy from the effects of the oil price rise. In the case of the second oil price shock, however, Japan had broken ranks with other major countries in its pursuit of a disinflationary monetary policy. As a result, Japan offers as close to a controlled experiment as economists can obtain for assessing the oil price explanation of inflation.

In the case of the first oil price shock, a problem with the special-factors explanation is that inflation had already started to rise before the shock. The sharp rise in the price of oil occurred after 1973Q3 with the start of the October Yom Kippur war in the Middle East. From September 1973 to March 1974, the price of oil (Saudi Arabian light-34 crude) rose from $2.80 a barrel to $9.60 a barrel. In Japan, CPI inflation had run around 4.5 percent in the last half of 1972. However, by 1973Q3, it had already risen to 12.8 percent.

Prior to the second oil price shock, as opposed to the first, monetary policy had fairly persistently concentrated on lowering inflation. The second shock began in 1978Q4 when the price of oil jumped from $12.80 a barrel to $40 a barrel by 1979Q4. One can look at the inflation bulge following the oil price rise to measure the impact of that rise on inflation. Because Japan imports most of its fuel, the rise in the price of oil directly affected its wholesale price index of raw materials, which includes the prices of imported raw materials. This index increased 73.7 percent in the 12 months ending March 1980 (Suzuki 1981, pp. 408–09).

These price rises passed on to the broad-based GDP deflator but to a much lesser extent. The GDP inflation rate rose significantly only in two quarters, 1980Q2 and 1980Q3, when it rose at an annualized rate of 9.9 and 8.3 percent, respectively. Figure 3, which shows four-quarter percentage changes in the GDP

deflator, reveals the inflation bulge. By this measure, inflation rose temporarily from a low value of 2.1 percent (for the four quarters ending 1979Q3) to a peak of 6.6 percent (for the four quarters ending 1980Q3).

However, this rise of a maximum of 4.5 percentage points in the inflation rate for four-quarter changes in inflation overstates the impact of the oil price rise. Prior to the rise, the value of the yen had risen dramatically (Figure 4). The rise in the yen-dollar exchange rate from 290 in early 1977 to 190 by late 1978 amounted to a one-third appreciation. The appreciation produced a transitory reduction in inflation lasting into 1979 that exaggerated the size of the inflation bulge shown in Figure 3. At the time, the Organization for Economic Cooperation and Development (OECD) attributed some of the low inflation prior to this bulge to the one-time effect of lower import prices produced by a rising yen (Yeager 1981, p. 169).

The inflation rate for the eight quarters prior to the rise in GDP inflation (from 1977Q4 to 1979Q3) was 3.4 percent. In the subsequent four quarters (1979Q4 to 1980Q3), the inflation rate rose to 6.6 percent. In the next four quarters (1980Q4 to 1981Q3), it fell back to 3.3 percent. The rise in the middle four quarters of about 3.2 percentage points appears to be a reasonable estimate of the direct effect of the oil price rise.

At the time, some economists argued that, to avoid adverse effects on output, central banks should allow a rise in the money stock to accommodate increases in the price level caused by relative price increases. However, the BoJ continued to concentrate on maintaining the public's expectation of inflation at a moderate level. With the occurrence of the oil price shock in late 1978, the BoJ began to raise short-term interest rates from their level of about 4.25 percent. In late 1979, when broad measures of inflation began to rise, it raised rates sharply until they exceeded 12 percent by March 1980. Consequently, money growth fell from 11.7 percent in the four quarters ending 1979Q3 to 8.2 percent in the five quarters ending 1980Q4.

Despite the aggressive actions taken by the BoJ to contain inflation and despite the importance of imported oil to Japan, the effect of the oil price rise on output appears to have been moderate. Over the period of unusual price rise (1979Q4 to 1980Q3), real GDP growth was 2 percent. Over the subsequent ten years, real GDP growth averaged an annualized 4.2 percent. It thus appears that the oil price shock reduced real output growth about 2 percentage points for one year.

Although oil imports had risen in importance by the end of the decade, the second oil price shock affected inflation and real output much less adversely than the first.[5] Monetary policy made the difference. In the first instance, the

[5] In 1972 and 1973, oil imports amounted to 2 percent of GDP. In 1979, they amounted to 4.4 percent of GDP and in 1980, 6.6 percent (Hutchison 1991, p. 10).

BoJ concentrated on reducing the size of the balance of payments surplus and on minimizing the impact on the economy of the yen revaluation agreed to in December 1971 (Ueda 1993, p. 200). In the second instance, the BoJ concentrated on retaining credibility for its inflation objective. The change in the behavior of wages in the second period compared to the first shows that the BoJ succeeded. In 1973 and 1974, nominal wage growth reached 21.9 and 29.1 percent, respectively. In 1980, nominal wage growth rose only slightly to 6.6 percent, from 5.9 percent in 1978 and 6.5 percent in 1979 (Ueda 1993, p. 201).

Japanese experience with the second oil price shock demonstrated that relative price shocks can exert a transitory influence on inflation. However, the contrasting experience between the first and second oil price shocks demonstrated the importance of the monetary regime in place. Random perturbations in the price level due to one-time effects exert minimal impact on real economic activity and no influence on trend inflation if the public believes that the central bank will maintain long-run price stability.

4. A RETURN TO EXCHANGE RATE TARGETING

In the last half of the 1980s, the BoJ moved away from stabilizing nominal expenditure and money growth. Instead, it began to attach importance once again to stabilizing the exchange rate. Especially after the Louvre Accord of February 1987, the BoJ limited interest rate increases to restrain appreciation of the yen.

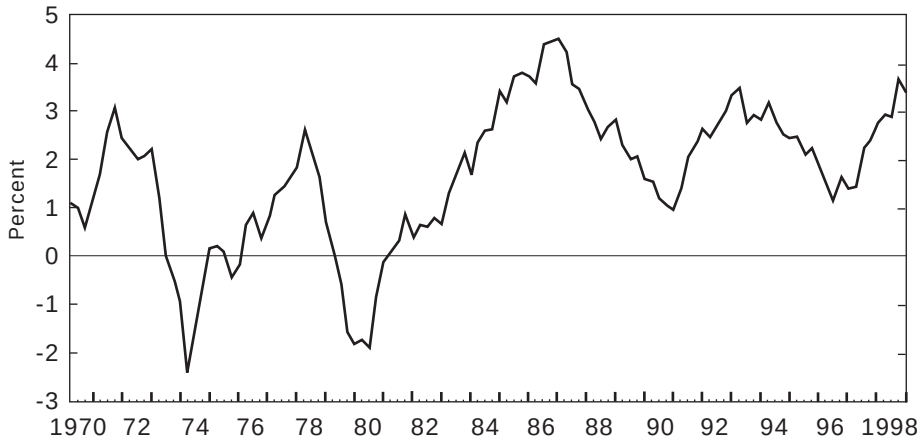
As explained in Section 1, in a fixed exchange rate regime, the central bank responds to a positive, rising balance of payments by lowering the discount rate. Starting in 1986, the BoJ moved partway back toward this monetary regime through a dirty float. That is, the BoJ lowered short-term interest rates in response to a positive current account balance and an appreciating yen. Lower interest rates made monetary policy expansionary.

An expansionary monetary policy dealt with both domestic and foreign pressures. Japanese exporters wanted the government to stimulate domestic demand to cushion the adverse effect of an appreciating yen on their exports. Furthermore, the U.S. government wanted Japan to stimulate domestic demand as a way of reducing its payments surplus through increased imports. By 1986, a rising Japanese trade surplus had kindled protectionist sentiment in the U.S. Congress, which threatened Japan with retaliatory trade measures.[6]

While the dollar rose in value against the currencies of other countries in the first half of the 1980s, the yen maintained its value. Starting in early 1985,

[6] These comments follow the discussion in Volcker and Gyohten (1992, pp. 248–58); James (1996, pp. 433–53); Ueda (1997, pp. 264–65, and 1993, pp. 207–09); and Sawamoto and Ichikawa (1994, pp. 94–96).

Figure 6 Japanese Current Account as a Percentage of GDP



Notes: Quarterly observations of current account balance divided by GDP. Heavy tick marks indicate fourth quarter of year.

the dollar began to depreciate against other currencies and the yen. The yen appreciated dramatically, rising from about 260 yen to the dollar in early 1985 to about 125 in early 1988. In response, the Japanese current account fell from early 1987 on, but only slowly (Figure 6).

In the September 1985 Plaza Agreement, the Group of Five countries had agreed to encourage the depreciation of the dollar that had begun earlier in the year. However, the continued strong Japanese current account balance eventually changed the focus from intervening in the foreign exchange market to raising Japanese imports by stimulating domestic demand in Japan. The Louvre Accord of February 1987 committed Japan to reducing its balance of payments surplus through an expansion of aggregate demand. From 7 percent in 1985Q4, the BoJ pushed the three-month Gensaki rate to 3.75 percent in 1987Q3.

Japan's money growth began to rise in 1987Q1. From an annualized rate of 8.1 percent from 1982Q1 through 1986Q4, it rose to 11.3 percent from 1987Q1 through 1990Q2. Because the Japanese public had come to expect price stability, the higher money growth initially impacted primarily real variables (GDP growth and real wealth) rather than prices. As a result, real GDP growth began to rise in 1987Q3. From an average annualized rate of 3.3 percent from 1980Q1 through 1987Q2, it rose to 5.7 percent from 1987Q3 through 1990Q2 (Figure 2). In 1989Q2 inflation began to rise. From an average annualized rate of 1 percent from 1985Q1 through 1989Q1, it rose to 2.8 percent from 1989Q2 through 1991Q4.

5. TARGETING ASSET PRICES

The rise in asset prices in the latter part of the 1980s came to be known as the bubble. As noted, CPI inflation rose only moderately in this period. The BoJ concluded that inflation targeting was inappropriate: instead of varying its policy instrument (the level of short-term interest rates) in response to the observed behavior of the prices of goods and services, it should respond to asset prices.[7] The BoJ believed that large rises in asset prices would inevitably be reversed with deleterious effects on the financial system and on real growth.[8] The belief held by many in Japan that the rise in asset prices, by exacerbating disparities of wealth, undermined the egalitarian character of Japanese society reinforced the BoJ's determination to limit those price rises (Ueda 1993, p. 205).

The asset price targeting of the post-bubble period can be understood as analogous to a commodity standard monetary regime. Instead of fixing the nominal price of an asset like gold, the central bank attempted to influence the nominal price of a much broader collection of assets—the country's reproducible and nonreproducible capital stock (equities and land). The situation was analogous to that of England's return to the gold standard in 1926 at the pre–World War I parity for sterling. The Bank of England had to force a deflation of the general price level to bring about a reduction in the price of gold. Assuming that a significant part of the rise in Japan's asset prices derived from real factors, the BoJ policy of lowering the yen price of assets required deflation in goods prices. As a result, Japanese monetary policy became strongly contractionary in mid-1990.

The expansionary monetary policy that began in 1987 had raised the real price of assets such as land. However, monetary expansion was not the sole source of the rise. The rise in asset prices also derived significantly from real factors. In the late 1980s, the price of land rose as Tokyo became an Asian center for growth of the service and information sectors. The deregulation of Japanese financial markets in the 1980s also increased Tokyo's attractiveness as an international financial center (Sawamoto and Ichikawa 1994, p. 91).

In May 1989, with an increase in the discount rate, the BoJ began pushing interest rates up sharply. The Gensaki rate went from 4.3 percent in 1989Q2 to 7.6 percent in 1990Q4 and 1991Q1. Reflecting the rate increase, money growth began to fall in 1990Q3, even becoming negative for three quarters

[7] For example, Sawamoto and Ichikawa (1994, p. 100) note, "It may be said that this argument is widely accepted, although opinions differ as to the desirability of making the stability of asset prices part of a target for monetary policy. . . ."

[8] For example, the governor of the BoJ, Yasushi Mieno (1994, p. 9), stated in a speech, ". . . it is apparent that with large fluctuations in asset prices since the second half of the 1980s, major fluctuations occurred in the economy as well, and sustainable growth was undermined. Therefore, asset price fluctuations are clearly a cause as well as a signal of fluctuations in economic activity. It is for this reason that the Bank of Japan includes asset price developments among the variables that it monitors."

after 1992Q1. Real GDP barely grew from 1992Q2 through 1995Q1.[9] From a peak of about 38,000 in early 1990, the Nikkei stock index fell to about 15,000 in 1992. Land prices began falling in 1991 and have continued falling to date.

It is essential to distinguish between shocks and the mechanisms that propagate those shocks. The initial shock was monetary—a rise in interest rates engineered through monetary contraction. The subsequent fall in asset prices reflected in part the direct effect of the rise in interest rates and in part the indirect effect of recession on expected future earnings. The reduction in wealth then propagated the real effects of the original monetary shock. One by-product of the fall in wealth was the creation of a mass of bad debts in the banking system. By reducing the efficacy of financial intermediation, those debts in turn amplified the original shock to the real economy.[10]

The fall in asset prices, which began in early 1990, came to be seen as precipitating the downturn—an inevitable collapse of a speculative bubble. The resulting mass of bad debts was the mechanism that propagated the shock. While superficially plausible, such an explanation is unsatisfactory. Japan had experienced a massive loss of wealth in World War II. Furthermore, the war

[9] For a discussion of Japanese monetary policy in the early nineties, see Goodfriend (1997).

[10] There is an analogy with U.S. monetary policy in the Depression. In the United States during the 1920s, the opportunities for growth appeared to have greatly expanded with the advent of mass production. The stock market rose accordingly. The Federal Reserve interpreted the rise as evidence of speculative extension of credit and initiated a rise in interest rates through a monetary contraction. The Fed then administered additional shocks. In fall 1931, it raised the discount rate sharply when Britain abandoned the gold standard. In spring 1932, it allowed a significant number of banks to close through bank runs.

The depressed state of the stock market during the Depression made the 1928 rise in the stock market appear to be a bubble. And the low level of interest rates made monetary policy appear impotent. The Fed restricted itself to preventing another presumed speculative rise in credit. Accordingly, in 1936 and 1937, it raised required reserves ratios sharply. The money stock then fell and economic recovery turned into renewed economic decline. Milton Friedman summarizes the conventional view of monetary policy in the Depression, a view which he and Anna Schwartz challenge in their 1963 book *A Monetary History of the United States,* in the following:

> Beginning in mid-1928, the Federal Reserve System, concerned about stock market speculation, adopted a monetary policy of nearly continuous restraint. . . . [In] the great contraction from 1929–1933 . . . the System pleaded impotence, arguing explicitly that the non-monetary forces making for contraction were so strong and violent that it was powerless to stem the tide, and implicitly that the depth of the decline in the money stock was due to the depth of the decline in business activity, rather than . . . the reverse (Friedman [1967] 1969, pp. 89 and 91).

The belief that the Depression was the result of a breakdown in financial intermediation that was produced by prior unrestrained speculation spurred the creation of numerous public agencies to revive lending. The Reconstruction Finance Corporation, started in the Hoover Administration, purchased the bad debts of banks and large corporations. Congress created the Federal Home Loan Bank System in 1932 to provide credit to housing. In 1933, Congress created the Federal Deposit Insurance Corporation. Also in that year, it created Production Credit Associations and Banks for Cooperatives to stimulate lending to farmers. However, the failure of these measures to end the Depression indicates that the bad debts were a by-product, not a cause, of the Depression.

had disrupted its financial system. However, this catastrophe did not produce economic stagnation. Instead the Japanese responded by increasing their supply of labor, and output rose dramatically. Why would they respond differently later?

The story that Japan's economic distress resulted from the collapse of a speculative bubble contained dramatic elements of greed and retribution, elements with mass appeal. That explanation relegated monetary policy to the role of avoiding future speculative bubbles. As a result, monetary policy did not relieve the relentless pressure for price deflation.

6. THE MAGNITUDE OF THE MONETARY SHOCK

One can measure the magnitude of the monetary shock by determining the extent to which actual money growth has fallen short of the growth consistent with price stability.[11] Formula (1), the equation of exchange, aids in the calculation:

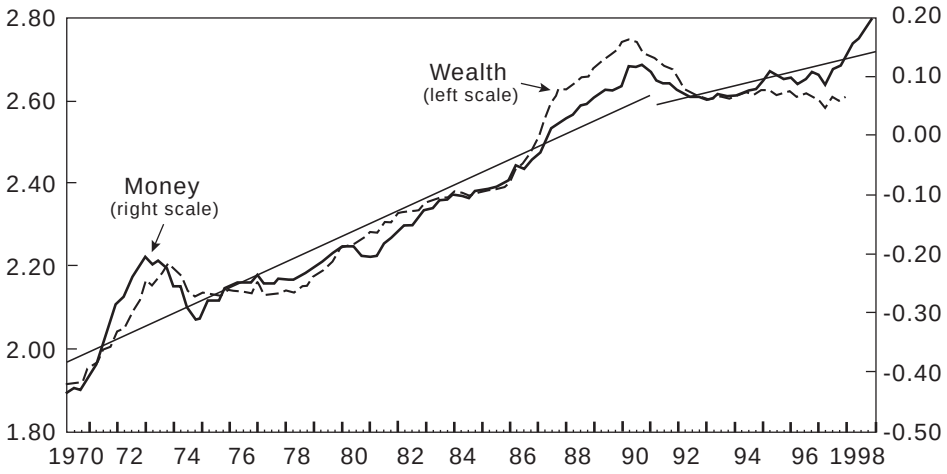
$$\dot{m} = \dot{p} + \dot{k} + \dot{y} \tag{1}$$

From left to right, the variables of (1) are growth in money, inflation, growth in the demand for real money expressed as a fraction of real expenditure, and growth in real expenditure. One can substitute into (1) trend values for growth of k and y that abstract from the monetary disturbances of the nineties. Figure 7 shows values of k from 1970Q1 through 1998Q4. (The slope of the trend line is the trend growth in k.) They rise secularly until the monetary contraction of the nineties. The growth of k from 1970 through 1990 is 2.2 percent. The figure used in equation (1) for growth in real expenditure (output) comes from the sum of labor-force growth and an estimate of longer-term productivity growth. From 1990Q1 through 1994Q4, the labor force grew at an annualized rate of 1 percent. From 1980Q1 through 1987Q2, the period before the monetary expansion of the late 1980s, labor productivity (real GDP per worker) grew at an annualized rate of 2.3 percent. The sum (1 plus 2.3) implies a figure for trend real output growth of 3.3 percent in the first half of the nineties. The sum of this figure and secular growth in k (2.2 percent) gives 5.5 percent as the rate of growth of money consistent with price stability.

The 5.5 percent figure for noninflationary money growth is based on labor-force growth of 1 percent. At present, the working-age population is no longer growing (OECD 1997, p. 113), so noninflationary trend money growth is probably near 4.5 percent. Nevertheless, the 5.5 percent figure applies to most of the nineties. From 1991Q1 through 1998Q4, money grew at an annualized rate of

[11] The calculations are for measured price stability. Biases in measurement of the price level in Japan make measured price stability consistent with deflation of 1 to 2 percent.

Figure 7 The Ratios of Nonmonetary Wealth and Money to Expenditure



Notes: The solid line is the logarithmic ratio of money (M2+CDs) to GDP. The trend lines are fitted to this series for the intervals 1970 to 1990 and 1991 to 1998. The dashed line is the logarithmic ratio of nonmonetary wealth to GDP. Wealth is from Sekine (1998) and includes financial assets such as land, housing, and inventories. Heavy tick marks indicate fourth quarter of year.

only 2.4 percent. This shortfall of about 3 percentage points in noninflationary money growth for eight years implies that the price level would have had to fall 24 percent to avoid any adverse impact on the real economy.

Even though this figure appears large, it ignores factors that might have increased the demand for money and thus required an even greater fall in the price level. First, the level of short-term interest rates is at historically low levels in Japan. The three-month Gensaki rate fell to only 0.3 percent starting in 1995Q4. At such low interest rates, the demand for money may increase significantly (Wolman 1997). Second, uncertainty may have increased the demand for money.[12] For example, uncertainty over the extent of the financial sector safety net in the fall of 1997 caused by the failure of Yamaichi Securities and Hokkaido Takushoku Bank may have increased the demand for money. The relatively rapid growth of narrow money (M1) in 1997 and 1998

[12] [Households] "have become extremely risk averse in their portfolio allocation behaviour. .... [H]ouseholds have shifted their portfolios, as mentioned, towards bank deposits [and] currency" (OECD 1998, p. 51).

relative to broad money (M2+CDs) may be evidence of an increased demand for liquidity.[13]

7. MONETARY NONNEUTRALITY

The magnitude of the monetary shock can also be seen through the fall in the trend rate of growth of money. Money growth was at an annualized rate of 9.5 percent over the ten-year period ending 1990Q2. It then fell to only 1.1 percent from 1991Q1 through 1994Q2. With a lag of almost two years, in 1992Q1 inflation did begin to fall, but only enough to produce moderate deflation. Economists disagree about why the price level adjusts sluggishly in response to nominal monetary disturbances. Ironically, the BoJ's hard-earned credibility for maintaining price stability very likely impeded a large-scale, sustained deflation.[14]

Growth of both real expenditure and real wealth adjusted in response to contractionary monetary policy.[15] Money growth fell 6.9 percentage points over the eight-year interval of 1990Q3 to 1998Q4 relative to the prior ten-year interval. Had inflation adjusted with complete flexibility, nominal GDP growth would have fallen by the same amount. Instead, it fell only 4.1 percentage points. That fall divided into a 1.1 percentage point reduction in GDP inflation and a 2.9 percentage point reduction in real GDP growth. Much of the real

[13] In 1997 and 1998, M1 grew at an annualized rate of about 8.5 percent, and M2 grew at about 3.6 percent.

[14] There are no direct observations of expectations of inflation in Japan. It seems plausible that the consistent policy of restoring price stability pursued by the BoJ from the mid-1970s to the mid-1980s earned it credibility. One piece of supporting evidence is the limited effect on inflation and wages of the price rises associated with the second oil price shock. Another piece of evidence that establishes the credibility of the BoJ's commitment to price stability comes from the coexistence of low government ten-year benchmark bond rates with a high and rising government debt-to-GDP ratio. At least through 1998, the public must have assigned no probability to central bank monetization of government debt.

[15] One reader pointed out that economists have been unable to construct models capable of explaining protracted slowdowns in real growth. The primary nonmonetary model is the real business cycle model, where recessions occur because of negative productivity shocks. However, economists have been unable to identify such shocks independently of the data series to be explained—real output. Monetary models emphasize either unanticipated monetary shocks or staggered price setting, perhaps because of contractual arrangements. In the former case, the monetary nonneutrality disappears when the public learns that the trend rate of growth of the money stock has fallen. In the latter case it disappears when the public resets or renegotiates fixed nominal prices. (Neither model deals with the length of time required for the public, long accustomed to a stable monetary regime, to adapt to a monetary regime requiring persistent deflation.) In order to understand the effects of their policies, monetary policymakers must then rely on a combination of models and historical episodes that possess the characteristics of controlled experiments. Friedman and Schwartz ([1963] 1969, pp. 216–22) list historical episodes where a prolonged economic slowdown has followed an independent monetary disturbance.

adjustment occurred in the form of a reduction in the growth rate of real nonmonetary wealth, which fell 9.3 percentage points over these two intervals.[16]

An interesting empirical fact is that the long-standing stable relationship between real money, real GDP, and real wealth continued even through the post-bubble period. In particular, the relationship between these variables remained consistent both in the event of negative and positive monetary shocks. This empirical consistency appears in the stability of the money demand functions estimated by the BoJ (1997) and Sekine (1998).[17] Both authors find long-run stability in the demand for money (M2+CDs) by the Japanese public. That stability survived the deregulation of the Japanese banking system in the 1980s (Sawamoto and Ichikawa 1994) and the bubble and post-bubble economy. That is, both authors find a stable relationship between real money, real GDP, and real wealth (defined to exclude money) that survived the boom and bust period. Figure 7 shows the positive relationship between the ratio of money to GDP (real money) and the ratio of nonmonetary wealth to GDP (nonmonetary wealth is from Sekine [1998]).

Why should a deflationary shock have prolonged real consequences? The answer given here is that the BoJ has created an environment where the public expects price stability. In that environment, it is hard for corporations to cut wages.[18] Without widespread wage cutting, corporations are reluctant to cut prices. Hourly wages continued to rise in the 1990s through 1997. The corresponding rise in real wages increased labor's share of national income from 68 percent in the 1980s to about 74 percent in 1998 (DIR 1998–99, p. 7). Correspondingly, corporate profits have fallen and so has the value of corporate equities. The fall in profits has led corporations to reduce capital expenditures in 1998, and the fall in wealth has led consumers to spend less. However, neither of these phenomena are causal; they propagate the monetary shock.

---

[16] Nominal GDP growth went from 6 percent to 1.9 percent. GDP deflator growth went from 1.8 percent to 0.7 percent. Real GDP growth went from 4.2 percent to 1.3 percent. Growth in real nonmonetary wealth went from 9.5 percent to 0.2 percent. The wealth figure is from Sekine (1998).

[17] Both authors use a cointegration analysis that involves regressing real money (M2+CDs) on an error-correction term, an opportunity cost variable, and scale variables. The opportunity cost variable is the difference between a rival interest rate and the own rate on money. The scale variables are real GDP and nonmonetary wealth. As a measure of wealth, the BoJ uses both an urban land price index and total financial assets of corporate business and personal sectors from flow of funds data. Sekine uses a broader measure of wealth from the System of National Accounts. The BoJ estimates its regression over the period 1968Q4 through 1996Q3, and Sekine estimates his over 1975Q1 through 1994Q4. The long lags in data availability on wealth determine the end dates. Yash Mehra and Jeffrey Walker, economists at the Federal Reserve Bank of Richmond, updated the work of the BoJ through 1998Q1. They found continued stability of the original money demand function.

[18] The Japanese system of lifetime employment has limited the ability of corporations to cut wages. The absence of a credible threat to fire employees limits corporate bargaining power.

8. RETURN TO A DIRTY FLOAT

Real GDP revived for a while in the mid-nineties, rising at 3 percent from 1995Q2 through 1997Q1. One contributing factor was a modest rise in money growth. From 1994Q2 through 1996Q3, money grew at an annualized rate of 3.2 percent. Moreover, inflation turned negative. The GDP deflator fell at an annualized rate of 0.7 percent from 1994Q2 through 1996Q4. An appreciating yen, which rose from about 160 yen to the dollar in 1990 to about 80 in 1995, facilitated the deflation. The combination of modest money growth and deflation increased real money balances. As a result, spending revived.

The revival in economic activity came to an end after 1997Q2. The precipitating factor in the onset of recession was the increase in the consumption tax rate from 3 to 5 percent on April 1, 1997. That increase produced a rise in inflation. CPI inflation, which had been running at a rate of 0.5 percent (year over year) jumped to almost 2.5 percent through March 1998. The BoJ did not accommodate the rise in inflation with higher money growth. Over the period subsequent to the tax hike, 1997Q2 through 1998Q1, M2 grew at an annualized rate of 3.3 percent, only slightly higher than the prior 3 percent growth.[19] The public reduced its real expenditures in response to the fall in its real money balances. An inventory cycle amplified the initial shock, and economic activity continued to decline throughout 1998.

As the economy weakened, the BoJ kept short-term interest rates unchanged. A primary reason was a fear that lower interest rates would produce a depreciation of the yen. The same fear kept the BoJ (1998, pp. 12–13) from aiming for faster money growth:

> . . . lower interest rates will reduce households' interest income. We then face the question whether consumer confidence, already low, would be further undermined. There is also the issue of whether lower interest rates would trigger an additional depreciation of the yen. . . . [T]he possibility cannot be denied that a depreciation of the yen could, by provoking falls in other Asian currencies against the U.S. dollar and causing a drop in stock prices in tandem, amplify people's anxieties. . . . [I]f a lowering of interest rates is problematic because it may have an adverse effect on consumer confidence and the yen's exchange rate, then we must realize that "monetary expansion" will give rise to exactly the same problems.

[19] The rise in inflation in itself would have produced a greater demand for nominal money and, at the prevailing call money rate of about 0.5 percent, a greater supply. However, the turmoil in financial markets in fall 1997, especially after the failure of Yamaichi Securities and Hokkaido Takushoku Bank, probably augmented the public's desire to save. (The widespread recognition at the time that the government would be liable for much of the bad debts of the banking system raised concerns that the government could have difficulty meeting its future pension liabilities.) Given no change in the level of short-term rates maintained by the BoJ, a reduction in the equilibrium real interest rate due to this increased desire to save made monetary policy more contractionary. The continued low rate of growth of money in this period probably reflected both forces.

Japan received strong external pressure not to allow a depreciation of the yen. In fall 1997, the International Monetary Fund negotiated rescue packages with Pacific Rim countries that required them to raise domestic interest rates to cushion the fall in their currencies. The fear was that a reduction in interest rates in Japan would put additional pressure on the currencies of these countries arising from a falling yen. For example, a falling yen would have put additional pressure on Korea, whose exports such as steel, cars, and consumer electronics compete with Japan's. Moreover, in early summer 1998, the Chinese Foreign Ministry and the governor of China's central bank made statements that observers interpreted as a threat to devalue the renminbi if the yen continued to fall. On June 17, 1998, concern over weakness in the yen led President Clinton and Prime Minister Ryutaro Hashimoto to agree to a joint intervention in the foreign exchange market to bolster the yen.

On September 9, 1998, the BoJ lowered its unsecured overnight call loan rate from around 0.5 percent to around 0.25 percent. Commenting on this rate reduction, a Knight-Ridder (1998) story stated that

> A senior official at the Economic Planning Agency . . . said the Japanese central bank had not been able to take easier credit policy as the yen's weakness (against the dollar) is feared to exacerbate the economic and financial crises in Asian countries. However, the yen's recovery from about 145 to the dollar in mid-August to below 131 . . . had given the BoJ a "leeway" to ease its credit policy.

9. MONETARY POLICY IN THE NINETIES

Popular commentary characterizes Japanese monetary policy as "easy" because of the low level of interest rates. The failure of this allegedly easy monetary policy to revive economic activity then leads to the conclusion that monetary policy in Japan is impotent. For example, a *Wall Street Journal* (1998, p. A19) article stated

> . . . sometimes, a nation's economy and financial system are so sick that conventional links between monetary policy and economic activity are damaged. . . . Many economists think that Japan has now fallen into a "liquidity trap," a bind no major economy has seen since the 1930s. In that state, interest rates hover near zero, but demand remains stagnant. There is no room to cut rates more.

It is true that interest rates are at historically low levels in Japan, but monetary policy is still contractionary if the BoJ maintains short-term rates above their equilibrium value. The equilibrium interest rate is the sum of the expected rate of inflation (deflation) plus the equilibrium real rate. Examination of each component indicates that the equilibrium value of the interest rate is near zero. First, expected inflation is probably zero or slightly negative. At present, the level of the CPI is basically steady, but the domestic wholesale

price index has fallen steadily since 1991. The result is to lower the equilibrium interest rate.

Second, the equilibrium real rate of interest is likely close to zero. In the investment boom of the late 1980s, private investment reached 30 percent of GDP. In the nineties, the capital stock continued to grow at a 3 percent or somewhat higher rate (OECD 1998, Figure 30). Continued investment probably sprang from the belief that Japan would resume normal growth after the 1991–1992 recession. The government further added to the capital stock by investing about 8 percent of GDP annually in infrastructure.

At the same time, real growth stagnated. Neither retail sales nor industrial production have grown from early 1992 to the present. It seems likely that the combination of sluggish domestic growth and continued growth in the capital stock produced a capital stock that is larger than optimal. Consequently, the marginal return to capital is probably near zero.

Despite low interest rates, saving remains high in Japan because the rapidly aging population needs to provide for its retirement. The unfunded liabilities (on a present-value basis) of Japan's public pension system are 180 percent of GDP (Matsuoka 1998). In addition, continued large government deficits are making the debt/GDP ratio in Japan among the highest in the world. General government (central and local government and social security funds) debt to GDP will rise to 130 percent of GDP in fiscal year 1999. The problem is that Japan's structural (cycle-adjusted) deficit is positive (DIR 1998–99, pp. 21–23).[20]

This imbalance creates a fear that the government will not be able to meet its pension obligations. At the same time, those saving for retirement have seen the value of their assets plunge.[21] House and stock prices are back to the levels of the early 1980s. Many Japanese now own homes whose value falls short of their mortgage obligation. As a consequence, Japanese saving is driving the equilibrium real interest rate toward zero.[22]

---

[20] Without a change in fiscal policy, the ratio of debt to GDP will rise indefinitely in Japan because the product of outstanding government debt and the difference between the long-term interest rate and nominal GDP growth exceeds the primary fiscal balance, where the latter is the fiscal balance excluding interest payments. The structural deficit in Japan, including national and local government and social security, is 2.4 percent of GDP (DIR 1998–99, p. 23).

[21] "Overall, the nation has had to confront cumulative capital losses [from 1990 through 1996] of around one quadrillion yen (about $7 trillion) which represents some two full years' worth of Japanese GDP and over 14 percent of the nation's total assets at the end of 1989. . . . [L]and prices have . . . fallen [from 1991 through 1998] by an average of around 4 percent per year" (OECD 1998, pp. 45–50).

[22] "The capital losses have had a number of obvious economic effects. . . . Households have been the biggest losers. . . . The resulting amount of negative equity for the owner of a typical 65-square-metre apartment in Central Tokyo purchased at the peak would reach 12 to 15 million yen ($100,000). . . . [T]hey [households] have probably boosted saving to try to restore some of this wealth. . ." (OECD 1998, p. 50).

The combination of zero, or negative, expected inflation with an equilibrium real rate near zero means that even the low market rates currently observed in Japan are consistent with contractionary monetary policy. A number of economists have argued that the BoJ should look at money growth rather than interest rates. Specifically, the BoJ should undertake open market purchases to increase the monetary base to the extent required to spur money growth (Friedman 1997; Goodfriend 1997; Laurent 1994, 1995; and Meltzer 1998).

If the BoJ were to expand the money stock, Japanese citizens would find themselves with larger amounts of money in their pockets.[23] Some extra money would be useful for emergencies, but beyond a point, individuals would spend additional money. Furthermore, the ability of the BoJ to create money is unlimited. It follows that the BoJ can stimulate spending. That spending would raise asset prices and spur both investment and consumption.

In opposition to this quantity theory of money argument, others have claimed that an expansionary monetary policy would be like "pushing on a string." The argument is as follows: If higher money growth is to affect the real economy, that money growth must still work through interest rates. And interest rates are so low that a further reduction could not make any difference. However, as economists like Pigou (1947) noted, this possibility ignores the wealth effects that higher real money balances would create in pushing interest rates down to zero. Bailey (1971, p. 113) expresses the argument as follows:

> Consider . . . a choice urban site whose expected annual rent is $1 per square foot. . . . [A] rate of interest of 5 percent per year implies a price of $20 per square foot. . . . Should the rate of interest fall to . . . 0.0000001 percent, [the price would be] $10 million per square foot, and so on. . . . It is hard to doubt that someone would be tempted by the time these [latter] prices were reached . . . to consume out of capital . . . enough to overwhelm other households' saving and eliminate aggregate net saving.

Moreover, at an interest rate close to zero, investors will again find the price of capital so cheap that they will want to accumulate capital. They will produce goods of great durability. And they will expand exports. Initially, the exchange rate would fall, but the economy would recover.

---

[23] The BoJ can expand the money stock by undertaking open market purchases, which enlarge bank reserves. In principle, the banks could limit an increase in the money stock by simply holding the additional reserves in their vaults. However, reserves yield no interest. At some point, open market purchases would make their reserves large enough that banks would surely purchase safe, short-term assets like government debt.

From February 1998 through February 1999, M2+CDs increased by 3.5 percent, while bank reserves increased by 9.4 percent. The BoJ could raise the rate of bank reserve growth to successively higher levels until it found a level of reserve growth that would produce the desired money growth. There is no limit to the ability of the BoJ to monetize assets and thus raise bank reserve growth. (Figures are from BoJ, *Reports and Statistics*, "Money Stock," and "Figures on Reserves," available on the BoJ web site.)

10. WHERE SHOULD JAPAN GO FROM HERE?

The stability of Japanese money demand and the difficulty of interpreting the economic impact of interest rates in an environment of deflation and depressed economic activity suggest that the BoJ should adopt an explicit target for money (M2+CDs) growth. The targeted growth rate should be high enough to relieve deflationary pressure on the price level. But what growth rate should the BoJ target? How does the BoJ maintain credibility for a policy of price stability?

The argument for temporarily high money growth is to relieve pressure for deflation. However, high money growth could become a source of instability if it were to cause the public to believe that inflation would become positive.[24] A policy of maintaining price stability would then be destabilizing (see Barro and Gordon [1983]). The public must believe that money growth above the level consistent with long-term price stability is temporary. A move to structural balance in the government deficit would help in that respect. As long as the government fiscal deficit maintains a structural imbalance, as opposed to a cyclical imbalance, the public might fear that the government would monetize its debt rather than repay it.[25]

The Japanese public must remain convinced that a reflationary monetary policy will not produce the alleged speculative excesses of the bubble period. The BoJ could target a rate of growth of money of, say, 9 percent (double the long-run growth rate assumed above to be consistent with price stability). At the same time, it could monitor the value of assets.[26] Over the period 1976Q1 to 1987Q1 (the period of stable money growth around a declining trend), real nonmonetary wealth grew at an annualized rate of 7.6 percent. The BoJ could lower its money target to 4.5 percent whenever the growth of wealth exceeded this value. It would also revert to the 4.5 percent figure when real GDP growth appeared to have stabilized at a sustainable value.

Japan should allow the money stock to rise sufficiently to remove the downward pressure on the price level. Over the longer run, it should return to

[24] This view is in opposition to Krugman's (1998) view that the BoJ should attempt to convince the public that it intends to inflate in the future.

[25] Starting with the locomotive strategy of the Carter Administration, Western governments have periodically pressured Japan to use deficit spending to stimulate its economy as a way of increasing its exports. (See the discussion of the Louvre Accord in Section 4.) The idea is that deficit spending would not only stimulate economic activity but would also raise interest rates. Higher interest rates would strengthen the yen and make Japanese exports less competitive. An expansionary monetary policy would also have stimulated economic activity, but by lowering interest rates. However, lower interest rates would weaken the yen and make Japanese exports more competitive, thereby exacerbating trade frictions. In actual fact, deficit spending has done little for Japan. In the 1990s, despite a rise in the level of general government debt from around 70 percent of GDP in 1990 to well over 100 percent by the end of 1998, real growth has stagnated.

[26] The BoJ would need to construct a proxy for wealth based on the stock market and recent information on land prices. The proposal reflects a suggestion in Meltzer (1998).

its "Golden Age" of monetary policy, 1975 to 1986, when the BoJ stabilized money growth and nominal expenditure growth to achieve price stability.

REFERENCES

Bailey, Martin J. *National Income and the Price Level.* New York: McGraw-Hill, 1971.

Bank of Japan. "Money Stock," and "Figures on Reserves," *Reports and Statistics.* Available: http://www.boj.or.jp.

_____. Speech by Governor Masuru Hayami to the Yomiuri International Economic Society in Tokyo, July 29, 1998, reprinted in Bank for International Settlements *BIS Review,* 75/1998. Available: http://www.bis.org/publ.

_____. Bank of Japan *Quarterly Bulletin.* "On the Relationship between Monetary Aggregates and Economic Activities in Japan: A Study Focusing on Long-Term Equilibrium Relationships," November 1997, pp. 104–24.

Barro, Robert J., and David B. Gordon. "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy,* vol. 91 (August 1983), pp. 589–610.

Daiwa Institute of Research (DIR). *Japan's Economic Outlook* (Winter 1998–99).

Friedman, Milton. "R_x for Japan: Back to the Future," *The Wall Street Journal,* December 17, 1997, p. A22.

_____. "The Monetary Theory and Policy of Henry Simons" (1967), and "The Role of Monetary Policy" (1968), in Milton Friedman, *The Optimum Quantity of Money.* Chicago: Aldine, 1969.

Friedman, Milton, and Anna J. Schwartz. "Money and Business Cycles" (1963), in Milton Friedman, *The Optimum Quantity of Money.* Chicago: Aldine, 1969.

_____. *A Monetary History of the United States, 1867–1960.* Princeton: Princeton University Press, 1963.

Glick, Reuven, and Michael M. Hutchison. "Monetary Policy, Intervention and Exchange Rates in Japan," in Reuven Glick and Michael M. Hutchison, eds., *Exchange Rate Policy and Interdependence.* New York: Cambridge University Press, 1994.

Goodfriend, Marvin. "Comments," in Iwao Kuroda, ed., *Towards More Effective Monetary Policy.* Tokyo: Bank of Japan, 1997.

Humphrey, Thomas M., and Robert E. Keleher. *The Monetary Approach to the Balance of Payments, Exchange Rates, and World Inflation.* New York: Praeger Publishers, 1982.

Hutchison, Michael M. "Aggregate Demand, Uncertainty and Oil Prices: The 1990 Oil Shock in Comparative Perspective," Bank for International Settlements Economic Papers No. 31. August 1991.

_____. "Monetary Control with an Exchange Rate Objective: The Bank of Japan, 1973–86," *Journal of International Money and Finance*, vol. 7 (1988), pp. 261–71.

Ito, Takatoshi. "Is the Bank of Japan a Closet Monetarist? Monetary Targeting in Japan, 1978–1988," Working Paper 2879. Cambridge, Mass.: National Bureau of Economic Research, March 1989.

James, Harold. *International Monetary Cooperation Since Bretton Woods.* Washington: International Monetary Fund, 1996.

Knight-Ridder. Story #9830, September 10, 1998.

Krugman, Paul. "Setting Sun; Japan: What Went Wrong," posted June 11, 1998. Available: http://web.mit.edu/krugman/www/japan.html.

Laurent, Robert. "Shades of the Thirties," *The Wall Street Journal*, August 1, 1995.

_____. "Monetary Policies in the Early 1990s—Reflections of the Early 1930s," Working Paper 94–26. Chicago: Federal Reserve Bank of Chicago, December 1994.

Lucas, Robert E. "Expectations and the Neutrality of Money," in Robert E. Lucas, Jr., ed., *Studies in Business-Cycle Theory.* Cambridge, Mass.: The MIT Press, 1981.

Matsuoka, Mikihiro. "Japan's Government Debt." Daiwa Institute of Research, August 12, 1998.

Meltzer, Allan. "Time to Print Money," *Financial Times*, July 17, 1998.

Mieno, Yasushi. "The Conduct of Monetary Policy by the Bank of Japan," Bank of Japan *Quarterly Bulletin* (August 1994), pp. 6–12.

Nikkei Weekly. "Pension Reform Panel Skirts Key Funding Issues," October 19, 1998, p. 3.

Organization for Economic Cooperation and Development. *OECD Economic Surveys, 1996–1997 and 1998, Japan.* Paris: OECD, 1997 and 1998.

Pigou, A. C. "Economic Progress in a Stable Environment," *Economica* (August 1947), pp. 180–88.

Sawamoto, Kuniho, and Nobuyuki Ichikawa. "Implementation of Monetary Policy in Japan," in Tomas J. T. Balino and Carlo Cottarelli, eds., *Frameworks for Monetary Stability.* Washington: International Monetary Fund, 1994.

Sekine, Toshitaka. "Financial Liberalization, the Wealth Effect, and the Demand for Broad Money in Japan," Bank of Japan *Monetary and Economic Studies,* May 1998, pp. 35–55.

Suzuki, Yoshio. *Money, Finance and Macroeconomic Performance in Japan.* New Haven: Yale University Press, 1986.

_____. "Japan's Monetary Policy over the Past 10 Years," Bank of Japan *Monetary and Economic Studies,* vol. 3 (September 1985), pp. 1–9.

_____. "Financial Innovation and Monetary Policy in Japan," Bank of Japan *Monetary and Economic Studies,* vol. 2 (June 1984), pp. 1–45.

_____. "Why Is the Performance of the Japanese Economy so Much Better?" *Journal of Japanese Studies* (1981), pp. 403–13.

Ueda, Kazuo. "Japanese Monetary Policy, Rules or Discretion? A Reconsideration," in Iwao Kuroda, ed., *Towards More Effective Monetary Policy.* Tokyo: Bank of Japan, 1997.

_____. "Japanese Monetary Policy from 1970 to 1990: Rules or Discretion?" in Kumiharu Shigehara, ed., *Price Stabilization in the 1990s.* Tokyo: Bank of Japan, 1993.

Volcker, Paul A., and Toyoo Gyohten. *Changing Fortunes: The World's Money and the Threat to American Leadership.* New York: Random House, 1992.

The Wall Street Journal. " 'Wealth Effect' Tends to Linger on for Years," September 10, 1998.

The Washington Post. "Japan Pressed to Aid Slumping Economy," April 4, 1998.

Wolman, Alexander L. "Zero Inflation and the Friedman Rule: A Welfare Comparison," Federal Reserve Bank of Richmond *Economic Quarterly,* vol. 83 (Fall 1997), pp. 1–21.

Yeager, Leland B. *Experiences with Stopping Inflation.* Washington: American Enterprise Institute, 1981.

_____. *International Monetary Relations: Theory, History, and Policy,* 2d ed. New York: Harper & Row, 1976.

Pooling or Purchase: A Merger Mystery

John R. Walter

On September 14, 1998, WorldCom merged with MCI to form MCI WorldCom, a global telecommunications giant. On September 30, NationsBank of Charlotte, North Carolina, and BankAmerica of San Francisco merged to form BankAmerica, one of the largest banks in the United States. While each case involved the combination of two firms, each used a different accounting method. MCI WorldCom's merger announcement noted that the combination would be accounted for as a "purchase"; on the other hand, BankAmerica's merger used a method called "pooling of interests" accounting.

In May 1991, American Telephone and Telegraph (AT&T) acquired computer manufacturer NCR Corporation (formerly National Cash Register) for $110 per share, in what was to that date the largest-ever computer industry merger. Press reports indicated that during negotiations AT&T upped its offer by $5 per share, an increase of about $325 million, to secure NCR's cooperation in accounting for the acquisition as a pooling of interests.[1]

Here is the mystery. AT&T paid the additional $325 million to use pooling accounting rather than the alternative—purchase accounting—a choice that affected accounting numbers but neither added assets, reduced liabilities, nor changed tax treatment. Why then was AT&T willing to expend an additional $325 million? Both anecdotal and empirical evidence indicate that AT&T's preference for pooling is not unusual. Corporate managers frequently go to some expense to employ pooling, though there are no obvious benefits.

The article benefited greatly from discussions with and comments from J. David Carroll, Thomas Humphrey, Jeffrey Lacker, Wenli Li, Raymond Owens, Warren Trepeta, and John Weinberg. The opinions expressed herein are the author's and do not represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] For a detailed discussion of AT&T's acquisition of NCR, see Lys and Vincent (1995). For press reports, see Smith (1991) and Cowan (1991).

These cases raise questions for those not acquainted with the features of merger and acquisition procedure. What are the differences between purchase and pooling of interests accounting? Should the choice of accounting method be of concern to analysts, investors, or others interested in business activity? Why are two different forms of accounting—purchase and pooling—used for otherwise similar acquisitions? What drives the choice between the two methods, and why are acquirers willing to take expensive steps that have only cosmetic consequences? This article addresses these questions.

Despite firms' express preference for pooling, the body responsible for setting U.S. accounting standards, the Financial Accounting Standards Board (FASB), recently proposed eliminating pooling, even though the accounting treatment has been used for years. While the change would bring U.S. merger and acquisition accounting standards more in line with standards used in other countries, acquisitive corporations are likely to oppose it. The change might offer some benefits, but the benefits could be offset by efficiency losses.

1. POOLING AND PURCHASE: THE NUTS AND BOLTS

Accountants attempt to report in balance sheets an accurate valuation of a firm's assets, liabilities, and equity. But how should accountants value a firm arising from the combination of two separate businesses? One approach is to simply sum the dollar amounts of assets, liabilities, and equity of the two firms as they stood before the combination. This is pooling of interests accounting. Or, since business combinations are typically one firm's purchase of another firm, another valid method would value the purchased firm at its purchase price, and add the purchase price to the assets of the acquiring firm, as one would if the acquisition were of a piece of equipment. In broad terms, the latter approach is purchase accounting. The financial statements of a combined firm will vary with the choice between pooling or purchase accounting. While accounting methods for business combinations have changed over time, under today's accounting rules both pooling and purchase are acceptable means of valuing combinations in the United States.

The terms merger, acquisition, consolidation, reorganization, and combination are often used interchangeably (none is particularly associated with either pooling or purchase accounting). While no single term predominates, throughout this article the term *business combination* will be employed to indicate the uniting of two firms, regardless of the features of the unification.

Pooling of Interests Accounting

As already implied, pooling of interests accounting is conceptually quite simple. When a business combination is completed, the balance sheet of the combined firm reflects assets, liabilities, and owners' equity at the sum of these accounts

as recorded by the separate companies immediately before the combination was completed. Income statements will show income and expenses for the statement period in which the combination occurs as if the companies had been combined from the beginning of the period (FASB 1992, pp. 213–14).

Purchase Accounting

Purchase accounting is somewhat more complicated. Under purchase accounting the acquiring and acquired firms are treated differently, so the first step is to identify which is which. FASB holds that in a typical combination the acquiring company pays out cash or other assets or issues the stock used in the acquisition and is the larger of the firms (FASB 1992, pp. 213–14).

Once acquirer and acquired are identified, accounting for the acquisition can proceed. The acquirer is to record on its books the acquisition at the price paid to the acquired firm's owners, using a two-step process. First, assets and liabilities from the acquired firm (target) are recorded on the acquirer's books at individual market values. Second, any positive difference between acquisition price and market value of net assets (assets minus liabilities) is recorded as an asset called goodwill. Once recorded, goodwill is depreciated by equal annual charges against the combined firm's earnings for a period of years over which, in the accountant's estimate, the combined firm benefits from the goodwill built by the acquired firm. The amortization period is limited to at most 40 years (FASB 1992, pp. 227–28). If the market values of the acquired assets and liabilities are accurately measured, goodwill is the value of the acquired firm as a going concern. Alternatively, goodwill can represent promising products developed by the target, or the price the acquirer is willing to pay for economic gains, such as economies of scale, expected from the merger (Brealey and Myers 1996, p. 930).

The following example may help illustrate purchase accounting. Assume Honest Auto Maintenance, Inc. (HAM), an auto repair shop management company, agrees to pay $100 million cash to acquire Wally's Import Repair, a regional chain. Following the acquisition, the assets and liabilities purchased in the acquisition are recorded on HAM's books at their current market values as determined by appraisers hired by HAM. The appraisers value the assets at $160 million and the liabilities at $90 million. So HAM has purchased net assets with a market value of $70 million ($160M–$90M). To record the difference between the market value of the net assets and the $100 million purchase price, $30 million of goodwill is recorded on HAM's balance sheet. For the next 40 years (the estimated life of the goodwill according to HAM's accountants) HAM will record on its income statement an after-tax expense of $750,000 ($30M/40 years), decreasing its reported net income each year by this amount.

The Logic Underlying Acquired Goodwill

The same example can be used to illustrate the logic of the purchase accounting treatment of goodwill. Assuming HAM viewed as accurate its appraisers' assessment of the market value of Wally's assets, $70 million of the $100 million it paid for Wally's company was for Wally's tangible assets. The remaining $30 million was to acquire Wally's good name in the community, an intangible asset, but an asset nonetheless. The asset will yield a future return. Following the acquisition, the value of Wally's good name must be recorded as an asset (called goodwill) on HAM's books; if it is not, HAM's worth is understated.

The $30 million expense borne to purchase the good name is *not* realized (charged against earnings) when the purchase is made but over time as the asset produces matching revenues. If the firm does not match expenses with the revenues these expenses produce, outsiders viewing HAM financial statements could be misled. For example, if instead all of the expense were recorded on the date the asset was purchased, profits would appear inappropriately low during the year of the purchase and too high in later years when the revenues generated by Wally's good name are received. So the annual $750,000 charges against earnings must be recorded on HAM's income statements throughout the estimated life of the asset.

The value of the asset must also be depreciated (lowered) over time. The logic here is that like most assets, Wally's good name has a limited life. Over time Wally's customers will move away or die off or, alternatively, learn that Wally is no longer running the operation and shift their business to competitors. Therefore the value of the good name declines over time, and goodwill is depreciated by $750,000 each year.[2]

When to Purchase and When to Pool

How do accountants determine whether to pool or purchase? A set of rules specifies the characteristics of combinations that can receive pooling or, alternatively, purchase treatment. The rules are intended to distinguish between two types of combinations: one represents firms joining forces, and the other represents one firm buying the assets and liabilities of another. Theoretically, in the first instance, the combined firm receives pooling treatment, and in the second, it receives purchase. Regardless of the intent of the rules, in reality firms can often choose their accounting treatment by structuring the combination carefully, though at possible extra cost.

[2] This example assumes straight-line depreciation whereby the asset is depreciated by equal yearly amounts. Other depreciation schedules are allowed under acquisition accounting rules.

The Rules

The rules are part of generally accepted accounting principles (GAAP). Combined firms' financial statements, like those of all firms, must adhere to GAAP, and therefore be in accord with the rules, to be deemed transparent, or not misleading, by the Securities and Exchange Commission (SEC) and for independent auditors to grant an unqualified audit report (Woelfel 1994, pp. 518, 1037).

Today GAAP is established by FASB, a private sector organization funded by contributions from professional accounting associations. The ultimate authority for determining GAAP rests with the SEC, however. The Securities Exchange Act of 1934 gave the SEC authority for establishing accounting and reporting standards for publicly held companies. The SEC delegates the setting of accounting standards to FASB, subject to SEC review.

The rules (called *conditions* by accountants), of which there are 12, were established in 1970 by the Accounting Principles Board (predecessor to FASB). If one or more of the conditions is violated, the combination must be accounted for as a purchase. If all conditions are met, pooling treatment is obligatory. (See the Box on page 44 for a list of the conditions.)

In general terms, the 12 pooling conditions prohibit certain financial transactions for specified periods before and after the acquisition and place restrictions on the terms of the acquisition. For example, one condition requires that the owners of the target be compensated predominantly with acquirer stock (specifically, 90 percent of consideration must be in stock—see condition 4 in the Box). Fundamentally, this condition and several others are intended to prohibit from pooling those combinations in which most target owners do not remain combined firm owners. This intention is based on the idea that to receive pooling treatment the combination must simply be the joining of two firms, with the former owners of the two firms continuing as owners of the combined firm. Other conditions prohibit acquirer and target repurchases of stock for a period before and after the acquisition. In a stock repurchase owners are bought out for cash and therefore relinquish their ownership interest in the combined firm.

Motivation for the Rules

By long-standing accounting convention, assets are recorded in financial statements at their original purchase price and liabilities at the amount of the original debt. The convention is known as historical cost accounting. But when does the original purchase of assets occur, and when are the liabilities assumed? Normally the answer is simple, but not necessarily when firms combine.

If the combination is simply the purchase of all of the assets and liabilities of the target firm, then historical cost accounting implies that the combined firm originally purchased the assets and originally assumed the liabilities on

the date the combination occurred. Accordingly, purchase accounting offers the appropriate treatment for such combinations. Operating under the notion that acquirers pay market value for assets and liabilities of targets, purchase accounting demands that each individual acquired asset be recorded at its market value at the time of the combination. As previously noted, the difference between the purchase price and the market value of net assets is recorded in a new asset account, goodwill, to account for the going concern value of the target at the time of the combination.

If, instead, two firms have combined with neither of them buying the other, no assets or liabilities change hands at the time of the combination. In this case, historical cost accounting demands that assets and liabilities be valued at the prices the target firm paid for them when it originally purchased or assumed them, perhaps years before the date of the combination. In other words, here historical cost accounting demands pooling treatment, in which the values as reported on the target's financial statements are carried over to the combined firm's statements.

There is no economic content to the distinction between combinations characterized by a joining of forces and those characterized by one firm buying the assets and liabilities of another. The bottom line is that the assets and liabilities of the two firms are merged together; therefore, the resulting firm is equivalent in either case. Furthermore, in a dynamic stock market, ownership changes hands constantly anyway as shares of stock are traded, so it makes no difference whether the same set of owners remains immediately before and after the combination.

Even if Pooling Rules are Violated, Firms May Still Manage to Pool

A firm that has violated one of the 12 conditions may still be able to pursue remedial steps that would allow it to pool. For example, prior to AT&T's purchase of NCR in 1991, NCR had repurchased several million shares of its own stock, thus violating a condition that seemed to make pooling impossible. So that AT&T could continue to employ the pooling accounting method, NCR agreed to reverse the stock repurchase by placing an equivalent amount of Treasury stock before the acquisition was completed (Gilson and Black 1995, p. 537; Lys and Vincent 1995, p. 367). Afterwards, the private placement was deemed sufficient to mend the violation of the condition.

2. THE MYSTERY: WHY DO ACQUIRERS PREFER POOLING?

The financial press and specialists in mergers and acquisitions maintain that acquirers prefer pooling to purchase accounting. Empirical analysis supports this view as well. It provides evidence that acquirers are willing to pay higher

bid prices in acquisitions that are pooled than in those that use purchase accounting. Likewise, even though there can be additional costs of qualifying for pooling, discussed below, the current significant use of pooling argues for the presence of a fairly strong preference. So what benefit underlies this strong preference?

At first blush one might imagine that pooling has some tax advantage. But, as shown later, advantageous tax treatment plays only a peripheral role, so taxes alone cannot explain the preference for pooling. Instead, both observers and acquirers themselves often argue that pooling is advantageous because reported earnings will generally be larger with pooling than with purchase.

For those who believe that markets are efficient whereby stocks are priced accurately based on all available information, the reported earnings explanation is puzzling, however. While accounting numbers are enhanced using the pooling method, firm performance is unaffected by the accounting choice. Moreover, information is available that allows investors to eliminate the effect on reported earnings. What follows will examine the puzzle but can offer no solution.

Acquirers Prefer Pooling

Attorneys and accountants specializing in mergers and acquisitions, as well as the financial press, report a strong preference among acquirers for pooling treatment.[3] To use pooling, all 12 conditions must be met (or violations must be mended). Some conditions severely restrict the structure of combinations and restrain the future actions of acquirers. Under the circumstances, the fact that pooling is chosen at all, much less that it predominates for large combinations, seems strong evidence that the reports are accurate.

AT&T's acquisition of NCR illustrates some of the costs of meeting pooling conditions. AT&T bore the expense of reversing NCR's earlier stock repurchases, an expense estimated to be $50 million (Lys and Vincent 1995, p. 367). Additionally, during negotiations with NCR, AT&T offered to increase its bid by $5 a share to $110 a share, or by $325 million, if NCR would make it possible to use pooling of interests accounting (Lys and Vincent 1995, p. 368). Similarly, acquisitive firms often give up or at least put off stock repurchases to avoid running afoul of pooling conditions.

As noted in a 1997 *Wall Street Journal* article, "[c]learly, companies prefer pooling over . . . purchase accounting. . . . Since 1992, there have been 357 poolings vs. 36 purchase acquisitions in deals valued at over $100 million. . . . So far this year, there have been 41 poolings vs. four purchase acquisitions" (MacDonald 1997).

[3] Attorney and accountant reports of a preference come from the author's interviews of merger and acquisition specialists.

Tax Avoidance Offers Only a Partial Explanation of the Pooling Preference

Does pooling offer a tax advantage not available under purchase accounting? The answer is that while no tax advantage results directly from pooling, one of the conditions that must be met to qualify a combination for certain tax advantages is also a key condition for pooling treatment. Consequently, the desire to attain tax advantages might account for some of the apparent preference for pooling treatment. Still, factors beyond taxes must explain most of the preference.

To What Extent Do Taxes Explain Pooling's Predominance?

In broad terms, under the Internal Revenue Code a combination is treated either as a *tax-free reorganization,* in which no taxes are assessed in response to the combination, or as *taxable,* in which case certain taxes are typically imposed. In a tax-free reorganization, the target's shareholders face no capital gains taxes on their stock as a result of the combination; instead, these taxes are deferred. While target shareholders receive the direct benefit, acquirers can expect to benefit as well, since target shareholders are likely to agree to a lower bid price if they are assured the acquisition will be deemed tax free. Alternatively, if a combination is deemed taxable, the target's shareholders must pay capital gains taxes on the exchange or sale of their stock.

One of the major conditions for a combination to be deemed tax free is that a majority of the consideration paid the target's shareholders be stock of the acquiring firm. In certain types of combinations, all consideration must be stock for the combination to be tax free. The upshot is that a firm wishing its combination to receive tax-free status automatically meets one of the fundamental conditions for pooling treatment—that the consideration be largely in stock. One might imagine, therefore, that the numerical predominance of firms choosing pooling treatment to some degree results from a preference for tax-free status.

The Pooling Preference Goes Beyond Taxes

If the conditions for tax-free treatment equaled those for pooling treatment, then the desire to avoid taxes might completely account for firms predominantly choosing pooling treatment. But the two sets of conditions are not equivalent. Instead, acquirers often choose to structure a combination to meet conditions for tax-free reorganization while stopping short of encumbering themselves with the conditions for pooling, some of which restrict valuable future actions.

A prominent example of a tax-free combination that used purchase rather than pooling accounting was when NationsBank Corporation, a large U.S. banking organization based in Charlotte, North Carolina, acquired Boatmen's

Bancshares of St. Louis in 1997. In this case the acquirer did not wish to pool because it intended to undertake stock repurchases in conjunction with the acquisition (NationsBank 1996, 1997).

Yet other firms are willing to take the extra step of meeting all of the pooling conditions, apparently for some benefit they perceive beyond tax-free status. For example, to qualify for pooling, acquisitive firms often willingly forgo planned stock repurchases, a step necessary for pooling treatment but not required for tax-free status.

Beyond the logical argument, there is empirical evidence that supports the theory that a pooling preference exists apart from any tax incentive. Robinson and Shane (1990) test for an association between accounting treatment and the premium paid for an acquisition, holding tax status constant. A premium is the difference between the bid price paid by the acquirer and the stock market value of a firm before its acquisition. Their hypothesis is that if acquirers prefer pooling to purchase accounting, one would expect acquirers to pay higher premia for such acquisitions. Using a sample of 95 tax-free acquisitions made between 1972 and 1982, including 59 accounted for by pooling of interests and 36 by purchase, Robinson and Shane find that, other things equal, acquirers pay more for pooling acquisitions than for purchase acquisitions. They provide estimates of the premia difference using various procedures. The average premia in pooled acquisitions exceeded the average in purchase acquisitions by between 29 and 66 percent. All the estimated differences were statistically significant. Consequently, the study suggests two conclusions, (1) that there is a measurable preference for pooling, borne out in the higher price paid for pooled acquisitions, and (2) since the study includes only tax-free acquisitions, acquirers prefer pooling for reasons beyond tax considerations.

Improved Reported Earnings as an Explanation for the Pooling Preference

If taxes cannot explain the pooling preference, what other motives remain? One frequently discussed explanation is that pooling allows acquirers to avoid purchase's negative effects on accounting earnings reported in financial statements in the years after the acquisition is completed. When purchase price exceeds book value of the target's equity, pooling accounting will produce higher reported earnings than purchase accounting for two reasons. First, goodwill depreciation lowers earnings but is not present in a pooled combination. Second, future reported earnings are reduced as higher depreciation expenses are recorded for the marked-up (to market value) assets in a purchase, but not pooled, combination. The view is that acquisitive firms are concerned that lower future reported earnings depress post-transaction stock prices.

The financial press often promotes this view (though articles typically focus only on the effect of negative earnings from goodwill depreciation and ignore

higher depreciation from marked-up assets). For example, a March 1997 *Institutional Investor* article points to an instance in which a deal failed, in part, over the effect on reported earnings of purchase accounting's goodwill amortization. One reason for the failure of Paramount Communication's 1989 bid for Time, Inc. was the "staggering load [of goodwill amortization that] was expected to depress the earnings of the acquiring company—and its stock price—for years to come" (McGoldrick 1997, p. 145). Likewise, a 1998 *Wall Street Journal* article on SEC efforts to reduce the use of pooling notes that the "drawback to purchase accounting is that it [imposes] an earnings penalty" (MacDonald 1998b).

Concern for reported earnings was apparently a major factor underlying AT&T's willingness to go to considerable additional expense to pool. In their review of AT&T's acquisition of NCR, Lys and Vincent (1995) conclude that a concern for a negative effect of goodwill amortization on future AT&T stock price was responsible for AT&T's willingness to expend at least an additional $375 million (i.e., 5 percent of the total acquisition price) to pool. Specifically, Lys and Vincent report that when they interviewed AT&T spokesmen, the spokesmen indicated that AT&T management believed that "financial analysts would . . . penalize AT&T's stock price for lower earnings [resulting from purchase's goodwill amortization]" (Lys and Vincent 1995, p. 370).

Outside of these reports, a number of empirical studies have investigated earnings maximization as a possible motive in acquirers' accounting choices. The studies find that acquirers act predictably if accounting choice (i.e., the choice between pooling and purchase) was motivated to maximize reported earnings.

When purchase price is higher than the target's book value, acquired asset depreciation and goodwill amortization charges will result. If managers are interested in maximizing *reported* earnings, they will prefer pooling to purchase when purchase price is higher than book value. Moreover, the larger the amount by which purchase price exceeds target book value, the more likely the manager should choose pooling. In research reviewed in Robinson and Shane (1990), a number of analysts found statistical evidence that managers choose to pool when the purchase/book value differential is positive; furthermore, the more positive the differential, the more likely pooling is chosen.

By All Logic, Acquirers Should Not Prefer Pooling to Purchase

The choice between pooling and purchase has no apparent economic consequences. Goodwill amortization charges are not cash flows and only serve to lower reported earnings. These charges have no effect on the current or future income produced by the activities of the firm. As a result, one would expect investors to ignore such changes to reported income and instead focus on changes in financial reports that signal changes in future income or in the health of the firm. While there appears to be abundant evidence of managers' preference

for pooling, as previous sections have described, acquisitive managers should be indifferent between pooling and purchase if investors look beyond reported earnings.

Nevertheless, the question arises: Do investors have sufficient information with which to remove the distortion produced by the pooling-purchase choice? The answer is yes. Such information is typically available in proxy statements associated with an acquisition so that interested investors should be able to remove the effects of goodwill on future earnings.

Empirical investigation supports the argument that the differences in reported earnings under pooling and purchase do not sway investors. Hong, Kaplan, and Mandelker (1978) examined empirically the stock price reaction to acquirers' pooling-purchase choice. They did so by using a statistical technique known as event study, which isolates the stock market's reaction to the event in question. In this case the event was business combinations accounted for either by pooling or by purchase methods.

Hong, et al. investigated acquisitions made in the period 1954 through 1964, when rules that limited the choice between the two accounting methods were more lax than those adopted in 1970. All acquisitions in the authors' sample were tax free.[4] They found no evidence that stock prices of acquirers using purchase accounting suffered relative to that of acquirers using pooling accounting. In fact, just the opposite was true—the stock price of acquirers using purchase accounting rose around the time of the acquisition while the stock price of poolers demonstrated no significant change.

Davis (1990) repeated the Hong, et al. study using a somewhat more advanced methodology and sampling business combinations that occurred after rules were tightened in 1970. Davis's sample covered the period between 1971 and 1982, and as with Hong, et al., it included only tax-free combinations. His results were equivalent to those in the earlier study. In other words, there was little evidence of a stock price reaction to pooled mergers but a significantly positive stock price reaction to mergers accounted for as purchases. Nonetheless, the positive reaction for purchases occurred largely in the six months prior to the mergers' announcements, presumably before the market knew the accounting method to be used. In the period after the merger announcements, the stock price reaction was positive for purchase and negative for pooling, but neither reaction was statistically significant. In sum, Davis's results show no stock price benefit from pooling. Therefore, Hong, et al.'s results seem to hold up well across time and across accounting rule regimes.

[4] Typically, combinations in which the acquired firm's shareholders sell their shares for cash are taxable to the acquired firm's shareholders, while those in which the acquired firm's shareholders are paid in stock are not taxable. Hong, et al. include in their sample only combinations in which the compensation was stock.

The Hong, et al. and Davis findings are in line with those of other event studies examining various types of firm choices that alter accounting numbers but have no economic effect (Brealey and Myers 1996, pp. 342–44; Copeland and Weston 1983, pp. 319–27). Such findings indicate that investors make use of information beyond reported earnings to make their stock investment decisions, ignoring accounting numbers that have no real economic significance.

3. FASB MAY ELIMINATE POOLING

Ever since pooling was first officially recognized almost 50 years ago, firms wishing to qualify have had to meet certain conditions. Yet over the years, businesses have attempted to test and expand the limits set on pooling treatment. Accounting rulemakers have consistently viewed this expansion as an abuse of the proper use of pooling and have responded by tightening the conditions. Nevertheless, inventive acquirers have successfully circumvented the stricter conditions. The opportunity for such expansion could end shortly, however. On April 21, 1999, FASB announced a proposal to eliminate pooling before 2001 (FASB 1999). FASB has noted several reasons, but two other factors discussed below seem worthy of consideration.

History of the Two Methods

While there were no official pronouncements before World War II, purchase accounting was apparently the more uniformly accepted method among accountants. For example, according to a 1943 version of the *Accountants' Handbook*, pooling treatment was "questionable," and purchase accounting was the "proper" means.[5] Pooling treatment began to gain favor among accountants in the late 1940s as more acquirers paid target firm owners with shares of stock in the combined company rather than with cash (Wyatt 1963, p. 24). As the means of payment shifted to stock, the outcome in these combinations was that owners from both the target and acquirer remained owners of the combined firm. As such, no sale of the assets had occurred, and there was no reason to revalue assets of the acquired firm.

The first authoritative pronouncement on accounting for business combinations embraced this interpretation. The Committee on Accounting Procedure, a predecessor of FASB and the body responsible for establishing GAAP at the time, issued *Accounting Research Bulletin No. 40* in September 1950, which stated that pooling treatment was allowed if several broadly specified conditions could be met; otherwise, purchase accounting was to be used. The first and pri-

[5] W. A. Paton, editor, *Accountants' Handbook*, third edition, 1943, as cited in Wyatt (1963), p. 21.

mary condition required that "substantially all" of the owners of predecessor corporations be paid with shares of the combined company rather than with cash or other assets. The other conditions were that neither of the combining firms be "minor" in size relative to the other, managers from both firms be retained in the combined firm ("management continuity"), and the activities of the combining firms be "either similar or complementary."[6] (Wyatt 1963, pp. 24–25, 123–26.)

Still, businesses had no strong preference for pooling at first because until 1953 purchase accounting did not necessarily result in charges against future reported earnings. Firms were allowed to charge the excess of purchase price over the acquired firm's book value to a capital account (called surplus). As a result, acquirers avoided the depreciation and amortization charges that lowered reported earnings. This capital account option was an alternative to establishing a goodwill account and recording future charges against earnings as under today's purchase accounting (Wyatt 1963, pp. 34, 38–39, 59). But in 1953, *Accounting Research Bulletin No. 43* eliminated the option of deducting the excess from surplus. With this change, a strong preference for pooling over purchase accounting began to surface apparently because of acquirers' aversion to charges against reported earnings (Wyatt 1963, pp. 38–39, 59–60).

By the late 1950s, accountants were expanding the vaguely specified pooling conditions to include more and more combinations (Wyatt 1963, pp. 61–62). For example, the condition requiring that the target's owners be compensated in stock was effectively relaxed as deals that involved substantial amounts of cash were soon treated as poolings. Likewise, the condition that neither combining firm be minor in size was continually tested so that by the 1960s, deals were handled as poolings even when the seller made up only 1 percent of the combined firm (Scharf, Shea, and Beck 1991, pp. 177–78). As a result, by the second half of the 1960s, pooling became predominant (Hong, Kaplan, and Mandelker 1978, p. 34). Clearly, businesses could and did choose pooling regardless of the features of the transaction.

In 1970 the Accounting Principles Board (APB; descendent of the Committee on Accounting Procedure and predecessor of FASB) issued new rules for business combinations intended to limit the use of pooling (FASB 1997, p. 2). Before the passage of these stricter rules, which are still in effect today, the SEC had itself threatened action to limit poolings. Initially the APB proposal completely prohibited pooling, but because the business community reacted

[6] The wording of *Accounting Research Bulletin No. 40* indicates a willingness to allow some violations of the size, management continuity, and activity similarity conditions without necessarily requiring purchase treatment. In reference to these three conditions, it says: "No one of these factors would necessarily be determinative, but their presence or absence would be cumulative in effect." The *Bulletin* offers no signs of flexibility in the requirement that payment to acquired firm shareholders be in the form of stock of the surviving firm. (*Bulletin No. 40* is reproduced in Wyatt [1963], pp. 123–26.)

negatively, it settled on new rules restricting pooling somewhat (Gilson and Black 1995, p. 517).

Today the tighter conditions make qualification more difficult. Nevertheless, as demonstrated by AT&T in 1991, combinations that might at first appear to be in violation *can* meet the pooling conditions if they follow some creative and at times costly procedures. The recent predominance of pooling for large combinations attests to the ability of firms to arrange their transactions to meet pooling conditions.

FASB's Proposed Change and Its Motivations

In August 1996 FASB once again began to consider changing the rules governing U.S. accounting treatment of business combinations. After reviewing a number of options, on April 21, 1999, FASB announced a proposal to eliminate pooling of interests accounting. FASB plans to release for comment its formal proposal early in the third quarter 1999 (FASB 1999). According to press reports, the international accounting community had encouraged FASB to adopt standards closer to international standards. The SEC likewise had urged FASB changes, and FASB itself had shown interest in aligning its standards more closely with international standards prior to the recent announcement (FASB 1997, p. 3; McGoldrick 1997, p. 147; MacDonald 1998a, pp. A2, A7; FASB 1999). FASB's proposed elimination of pooling would align the standards since most other nations allow pooling only infrequently or not at all.[7] However, U.S. companies are likely to oppose the proposed elimination.

FASB expressed several motives for eliminating pooling both prior to and as part of its April 21 announcement. Given that pooling is infrequently used in other countries, one reason FASB is planning to eliminate it is to simplify international accounting comparisons (FASB 1997, p. 21). FASB also wishes to prevent "abuses" of pooling; presumably, by "abuse" FASB means that firms are constantly stretching the definition of pooling to fit more and more combinations (FASB 1997, p. 2). Recently the SEC penalized just such stretching in some highly publicized cases. Several acquirers were required to revise financial statements to record as purchases combinations that had previously been shown as poolings (McCafferty 1998, p. 23; MacDonald 1998b, pp. C1, C2). Last, FASB wants to make comparison across U.S. firms simpler. Comparison is costly now, involving as it does two very different accounting treatments for business combinations (FASB 1999).

[7] See McGoldrick (1997) and FASB (1997) for a listing of business combination accounting standards in other major countries. For example, Australia allows no poolings at all, while Canada allows them only when an acquirer cannot be identified. In the case of Canada, the merger partner with more than 50 percent of the combined firm is assumed to be the acquirer.

An Additional Rationale for Eliminating Pooling

One can imagine an additional reason for preferring purchase to pooling. Purchase accounting requires acquiring firms to reveal more up-to-date information about the value of the target firm's assets and liabilities than that revealed by pooling, thereby potentially lowering information costs to investors. Specifically, target firm assets and liabilities must be recorded on the combined firm's books at their market values as of the time of the combination. Alternatively, under pooling treatment the target's assets and liabilities are recorded at historical cost, or value at the time they were originally obtained by the target, with no adjustment for price changes occurring since.

The current market values can be useful in the following manner. Imagine first Conceal, Incorporated, a firm arising out of a pooled combination of Conceal, Inc. (the surviving firm) and Target, Inc. An investor considering a purchase of stock in Conceal will wish to determine whether Conceal received a good price on its purchase of Target in order to judge the quality of Conceal's management. Because the Conceal-Target combination was pooled, Conceal's financial statements reveal only the historical asset and liability values (the prices originally paid for the assets and the debts originally contracted for) of Target. Yet the investor is concerned that the true market values may be quite different due to influences or disturbances (shocks) that may have affected prices of some of Target's assets.

The investor believes that analysts who have studied Target's stock make recommendations influencing its price that fail to account for the effect of the shocks on Target's asset values. So the investor is left to make his stock purchase decision based on historical asset values for Target and on his own admittedly rough estimate of how the shock might have affected Target's true asset values.

On the other hand, imagine Divulge, Inc., a new firm resulting from a combination of Divulge, Inc. (the surviving firm) and S. Target and Sons. Suppose this combination received purchase accounting treatment. Here the situation is exactly the same as the Conceal case except the investor has access, from Divulge's financial statements, to the current market values of S. Target's assets and liabilities. In this case, the investor need not estimate the effect of the shocks on various S. Target assets because the figures are provided on Divulge's financial statements.

Are the market values of S. Target's assets and liabilities, as reported on Divulge's financial statement, likely to be accurate, or will Divulge be able to manipulate their values to paint a more positive picture? For assets and liabilities with no secondary market, the estimates of market values will necessarily involve some subjective judgments. Still, there is a fairly well-established body of rules and procedures used to guide the pricing of such assets. These rules, developed by the appraisal industry, have been tested in court cases over the

years. If the appraiser hired by Divulge to evaluate S. Target does not abide by the rules, he could be sued. Therefore, while manipulation is certainly possible, the threat of lawsuits brought by disgruntled stockholders tends to encourage reasonable valuations.

Consequently, purchase accounting may provide investors with superior information compared to pooling of interests accounting. At the very least, purchase provides all the information pooling does along with an appraiser's estimate of market values of the target's assets and liabilities.

A Move to Eliminate Pooling Should be Undertaken Cautiously

As shown previously, not only has FASB identified several reasons to eliminate pooling, but also this article has noted that purchase accounting may provide superior information as compared to pooling. Nevertheless, FASB might wish to consider one factor before proceeding with its proposal to eliminate pooling.

Pooling may produce some real cost savings that are not immediately evident, for managers prefer pooling in spite of no apparent stock price benefit and therefore no benefit to owners. Are managers simply acting foolishly? This seems unlikely, since one expects foolish behavior to be punished and eventually extinguished by a reasonably efficient stock market. Yet poolers are not penalized by the stock market, and the pooling preference has continued for years, perhaps because pooling allows some firms cost savings, offsetting the additional costs of meeting conditions to pool.

It is not difficult to imagine, at least in broad terms, how benefits from pooling might arise. For example, perhaps pooling may give the combined firm greater flexibility in its future accounting. Managers may use that flexibility to increase their salaries and bonuses. While such managerial actions may appear to enrich managers at stockholders' expense, they may be necessary to achieve a preferred compensation and incentive arrangement between management and ownership. The benefits of achieving the preferred arrangement might exceed the costs of achieving pooling.

However, if benefits exceed the costs of achieving pooling, would we not then expect empirical tests (e.g., Hong et al.) of stock price reactions to have shown pooling to be rewarded more highly than purchase? Not necessarily. Some firms may not need to allow managers the added flexibility to extract higher compensation. Since these firms may prefer other management compensation arrangements, they are free to use purchase accounting, thereby avoiding the additional costs of meeting the conditions to pool. Any firm, whether it is one that chooses pooling or one that chooses purchase, will be rewarded for choosing its optimal accounting strategy.

The important point here is that unless managers are irrational, and as long as the market is fairly efficient, there is reason to believe that firms' preference for pooling is driven by some real cost savings. Attempting to identify that benefit perhaps should be a prerequisite to eliminating pooling.

4. CONCLUSION

In the post–World War II period, business combinations have received one of two accounting treatments, pooling of interests or purchase accounting. From the start, businesses preferred pooling and are apparently willing to pay for it, by most accounts because it allows them to report higher earnings. Yet that accounting choice neither increases assets, reduces liabilities, nor modifies tax treatment, so in theory it ought to be ignored by investors. The preference for pooling is especially puzzling since empirical research implies that investors are not swayed by whether merged firms employ pooling or purchase accounting.

While the puzzle has yet to be solved analytically, there is a good chance that it will diminish in importance for business decisions. The Financial Accounting Standards Board has indicated that it may well eliminate pooling, bringing U.S. business combination accounting standards in line with those in other industrialized nations. This article suggests that while the elimination of pooling in favor of purchase accounting could produce benefits by requiring more complete disclosures in financial statements, it also might eliminate arrangements that owners and managers find to be cost-saving. Evidence that acquirers are willing to bear additional costs to pool, combined with the lack of a stock price penalty for poolers, at least hints at the presence of some as-yet-undetermined cost savings.

The Twelve Conditions For Pooling

1. Each of the combining companies is autonomous and has not been a subsidiary or division of another corporation within two years before the plan of combination is initiated.

2. Each of the combining companies is independent of the other combining companies, meaning that none of the combining companies have significant equity investments (greater than 10 percent of outstanding voting common stock) in one another.

3. The combination is effected in a single transaction or is completed in accordance with a specific plan within one year after the plan is initiated.

4. Payment is effected by one corporation offering and issuing only common stock in exchange for substantially all (meaning 90 percent or more) of the voting common stock interest of another company. The common stock issued must have rights that are identical to those of the majority of the issuing company's outstanding voting common stock.

5. None of the combining companies changes the equity interest of its voting common stock for two years before the plan to combine is initiated or between the dates the combination is initiated and consummated. Changes to equity interests may include distributions to stockholders and additional issuances, exchanges, and retirements of securities.

6. None of the combining companies reacquires shares of its voting stock except for purposes other than business combinations. Examples of allowable share repurchases might include shares for stock option and compensation plans and other recurring distributions provided a systematic pattern is established at least two years before the plan of combination is initiated.

7. The ratio of the interest of an individual common stockholder to those of other common stockholders in the combination is unchanged before and after the combination. In other words, each individual common stockholder who exchanges his stock receives a voting common stock interest exactly in proportion to his relative voting common stock interest before the combination is effected.

8. Voting rights in the combined company are exercisable by the stockholders. This condition is not met, however, if shares of common stock issued to effect the combination are transferred to a voting trust, in which case the individual stockholders would lose the ability to vote.

9. The combination is resolved at the date the plan is consummated. In other words, the combined corporation does not agree to issue additional shares of stock or other consideration on any contingency at a later date to former stockholders of the combining companies.

10. The combined corporation does not agree to retire or reacquire any of the common stock issued to effect the combination.

11. The combined corporation does not enter into other financial arrangements—such as a guaranty of loans secured by stock issued in the combination, which in effect negates the exchange of equity securities—for the benefit of former stockholders of a combining company.

12. The combined corporation does not intend to dispose of a significant part of the assets of the combining companies—other than disposals in the normal course of business or to eliminate duplicate facilities or excess capacity—within two years after the combination.

Source: Financial Accounting Standards Board (1992), pp. 209–12.

REFERENCES

Brealey, Richard A., and Stewart C. Myers. *Principles of Corporate Finance.* New York: McGraw-Hill, 1996.

Copeland, Thomas E., and J. Fred Weston. *Financial Theory and Corporate Policy.* Reading, Mass.: Addison-Wesley Publishing Company, 1983.

Cowan, Alison Leigh. "NCR Accounting Tactic Could Hurt Any Merger," *New York Times,* March 22, 1991.

Davis, Michael L. "Differential Market Reaction to Pooling and Purchase Methods," *Accounting Review,* vol. 65 (July 1990), pp. 696–710.

Financial Accounting Standards Board. "FASB to Eliminate Pooling of Interests Accounting," News Release, April 21, 1999. Available: www.rutgers.edu/accounting/raw/fasb/news/nr42199.html.

_____. "Project Update: Business Combinations," posted September 30, 1998, at http://www.rutgers.edu/accounting/raw/fasb/project/buscomb.htm [October 21, 1998].

_____. "Financial Accounting Series Special Report: Issues Associated with the FASB Project on Business Combinations," No. 174–A. Norwalk, Conn.: FASB, June 1997.

_____. *Original Pronouncements: Accounting Standards, 1992/93.* Homewood, Ill.: Irwin, 1992.

Gilson, Ronald J., and Bernard S. Black. *The Law and Finance of Corporate Acquisitions.* Westbury, N.Y.: The Foundation Press, 1995.

Hong, H., R. S. Kaplan, and G. Mandelker. "Pooling Vs. Purchase: The Effects of Accounting for Mergers on Stock Prices," *Accounting Review,* vol. 53 (January 1978), pp. 31–47.

Lys, Thomas, and Linda Vincent. "An Analysis of Value Destruction in AT&T's Acquisition of NCR," *Journal of Financial Economics,* vol. 39 (October–November 1995), pp. 353–78.

MacDonald, Elizabeth. "FASB Weighs Tighter Acquisition Rules," *The Wall Street Journal,* October 23, 1998a.

_____. "SEC Moves To Tighten Merger Ploy," *The Wall Street Journal,* July 14, 1998b.

_____. "Merger-Accounting Method Under Fire," *The Wall Street Journal,* April 15, 1997.

McCafferty, Joseph. "Pooling Under Fire (Again)," *CFO,* September 1998, p. 23.

McGoldrick, Beth. "Goodwill Games," *Institutional Investor,* March 1997, pp. 145–49.

NationsBank Information Center. "Boatmen's Shareholders Choose NationsBank Stock," Press Release, January 8, 1997. Available: http://www.nationsbank.com/newsroom/press/1997/nbstk.htm [January 18, 1999].

_____. "NationsBank and Boatmen's Bancshares Agree to Merge," Press Release, August 30, 1996. Available: http://www.nationsbank.com/newsroom/press/1996/boatmens.htm [January 18, 1999].

Robinson, John R., and Philip B. Shane. "Acquisition Accounting Method and Bid Premia for Target Firms," *Accounting Review,* vol. 65 (January 1990), pp. 25–48.

Scharf, Charles A., Edward E. Shea, and George C. Beck. *Acquisitions, Mergers, Sales, Buyouts, and Takeovers: A Handbook with Forms,* 4th ed. Englewood Cliffs, N.J.: Prentice Hall, 1991.

Smith, Randall. "NCR Board, Facing Hard Proxy Fight, Agrees to Talks on AT&T Merger Offer," *The Wall Street Journal,* March 25, 1991.

Woelfel, Charles J. *Encyclopedia of Banking and Finance,* 10th ed. Chicago: Probus Publishing Company, 1994.

Wyatt, Arthur R. *A Critical Study of Accounting for Business Combinations.* New York: American Institute of Certified Public Accountants, 1963.

A Primer on Moral-Hazard Models

Edward S. Prescott

Moral Hazard — *The effect of insurance on insureds' behavior.*

Moral hazard, a long-time concern in the insurance industry, is increasingly being recognized as a concern in the regulation of banking and other financial industries. A classic example of its possible perverse effects is the selling of a fire insurance contract to a group of uninsured individuals. If the premiums are based on the actuarial data of this group's loss experiences, then the contract will be unprofitable. The reason for this loss is that with the introduction of fire insurance, insured people take fewer precautions than before against fires, raising losses above historical levels. It is this adverse effect of insurance on people's behavior that is moral hazard, and it is because of these adverse effects insurance contracts frequently contain clauses that attempt to minimize this behavior such as deductibles and copayments.

Ever since moral-hazard models were formalized mathematically in the early 1970s, applications of the models have burgeoned.[1] The models have been applied to just about any field where contractual relationships play an important role. In development economics, they have been used to study agricultural sharecropping contracts. In corporate finance, they have been used to study capital structure and executive compensation. In labor economics, they have been used to study employee compensation.

Moral-hazard analysis plays an important role in the theories of bank regulation and, more generally, financial regulation. For instance, government guarantees of bank deposits, be they explicit or implicit, reduce the incentive for depositors to monitor their banks. This lack of monitoring can create

The author would like to thank Andreas Hornstein, Brent Hueth, Jeff Lacker, Wenli Li, and John Weinberg for helpful comments. The views expressed in this article do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] Some of the earliest work on this topic was done by James Mirrlees, one of the 1996 Nobel Laureates in Economics. See Dixit and Besley (1997) for a review of his contributions.

incentives for financial institutions to take on excessive amounts of risk. One striking example of the perverse risk-taking effects caused by poor regulatory design is the U.S. savings and loan crisis in the 1980s.[2] Moral hazard may also be a problem when financial institutions are large enough to be "too big to fail." Some commentators fear that too-big-to-fail institutions have an incentive to take on excessive risk because these institutions (and even their "uninsured" creditors) will be bailed out in the event of a failure.[3]

Moral hazard is also assigned a prominent role in some analyses of causes of the recent Asian financial crisis. In these analyses, lax domestic financial regulation led to excessive risk-taking by Asian financial institutions. Furthermore, as argued by Calomeris (1998), the expectation of International Monetary Fund bailouts for developing country banking sectors gave foreign investors an incentive to finance risky activities.

Moral-hazard models are studied by analyzing constrained maximization programs, an important class of optimization problems. Though some of these programs are easy to study, the moral-hazard class is a particularly difficult one to analyze. Consequently, an extensive literature has developed that provides conditions that simplify the program. Unfortunately, these conditions are unappealingly restrictive.

To avoid these simplifying assumptions, we present another approach to analyzing moral-hazard models whereby we compute solutions to numerical examples. There are two advantages to this approach. First, it can be used to study problems that are not amenable to analytical methods. Indeed, the methods for computing numerical examples in this article succeed in cases in which the standard analytical simplification does not apply. The second advantage to computing solutions is that it gives one the ability to answer quantitative questions. For instance, the effect of deposit insurance on bank risk-taking is an inherently quantitative question. So is the question of whether a smaller amount of deposit insurance would result in a more efficient mix of risk and insurance. If one does not compute examples, the answers to these questions, as well as others, cannot be quantified.

In this article, we use linear programming as the computational technique for solving moral-hazard programs. Linear programming has been widely applied in management science, operations research, engineering, industry, and economics. Much is known about this class of programs, and practical algorithms for computing solutions to them have existed since the 1940s. Today,

[2] See Kareken (1983) for an early warning about this crisis. See also Benston and Kaufman (1998), who not only discuss this episode but also the effectiveness of The Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA), the reform that resulted from the episode.

[3] For a recent statement of this view, plus a proposed solution, see Feldman and Rolnick (1998).

there are numerous commercial and publicly available software codes that implement them.

Developing the moral-hazard program as a linear program requires making two departures from its standard formulation. One change is minor but the other change, allowing randomization or lotteries in the contract, is not. The implications of making this latter change will be discussed later. It is important, however, to understand that the change is not merely a technical assumption. Rather, it is one that can be justified on economic grounds.

As we review the moral-hazard model in the following paragraphs, we discuss the role of private information and describe the moral-hazard program as it is usually seen in the literature. This section is self-contained and can be read as such if the reader is interested only in the basic intuition of moral-hazard problems.

Using the standard formulation as a starting point, the next section introduces lotteries. We discuss the economic reasons for using them and present the linear program. (Two appendices spell out the linear program in more detail: Appendix A contains a short review of linear programs, and Appendix B derives the linear programming representation of the moral-hazard program.) Next, we compute the solution to a bank regulation example followed by some concluding comments. Finally, Appendix C contains a list of the papers that formulate private-information problems other than the moral-hazard problem as linear programs.

1. THE MORAL-HAZARD PROBLEM

The moral-hazard problem usually is formulated in terms of a contract between a principal and an agent who "works" for him. The principal and the agent can be people or institutions. With regards to agricultural sharecropping, the principal is the landowner and the agent is the tenant. With regards to banking, the principal is the bank regulator and the agent is the bank. With regards to executive compensation issues, the principal represents the collective interests of the shareholders while the agent is the chief executive officer.

In the moral-hazard problem, the agent works on a project for the principal. The amount of work the agent performs affects the probability distribution of the project's return. The problem is that the principal cannot monitor the agent's work, so the agent's effort is *private information*; that is, it is observed only by the agent himself.[4] In some models, the agent's amount of effort is not observed. In other models, precisely how the task is performed is not observed. Our model captures either specification, so we will refer generally to the agent taking an action, be it a level of effort or a specific task. In agricultural sharecropping arrangements, the action is the amount of effort the tenant applies to working

[4] Actions that are private information are also sometimes called hidden.

the sharecropped land, and the output is the crop yield. In bank regulation, the action is the risk-return profile of the bank's investments, and the output is the bank's return. In the executive compensation example, the action is the amount of resources aimed towards increasing the corporation's profits rather than the CEO's own personal satisfaction (for example, expanding the size of the corporation or avoiding painful decisions like layoffs), and output is the corporation's profit. In each example, the agent takes actions, unobserved by the principal, that affect output.[5]

Moral-hazard models are normally developed so that there is a conflict between the agent and the principal over the action the agent should take. For example, the agent might prefer a low-effort action because he dislikes hard work, while the principal might prefer a high-effort action because it increases the expected output of a project. This conflict in and of itself does not cause moral hazard but it does so when combined with private information on the agent's action.

To understand the role of private information, it is helpful to first consider the opposite situation, that is, when both the principal and the agent observe the action taken (commonly called the full-information case). In this case the principal and the agent could simply make a contract that fixes the level of action the agent should take. The exact action they would agree upon depends on many factors, such as the outside opportunities of the agent's labor, the outside opportunities of the principal's investment project, and the amount of compensation the agent receives. Nevertheless, once the two parties agree upon an action, they can enforce it because both parties observe it.[6]

Returning to the private-information assumption, one would see that the principal and the agent could still write a contract specifying the agent's action. But in this case, how could the contract be enforced? After all, if the principal cannot observe the agent's action, how can he make sure the agent took it? Although the principal could ask, there is no guarantee that the agent would respond truthfully; the agent can always reply that he fulfilled his end of the contract, whether he did or not.

While the principal cannot make the agent work a specified action directly, he may be able to induce the agent to take the desired action using the

[5] Not allowing the principal to observe the action at all is, admittedly, an extreme assumption. Landowners can inspect the fields, regulators can monitor asset quality. Still, these measures are not perfect. It is expensive to audit and difficult to interpret signals. The basic moral-hazard problem laid out here should be considered a starting point for detailed analysis of any particular contractual arrangement. Adding features like auditing to the moral-hazard problem can be, and has been, done. We will return to this issue later.

[6] The private-information literature usually assumes that if both parties observe a variable, then they can write enforceable contracts on it. Clearly, contracting is not so simple in practice. The largest literature that tries to address this problem is the one on incomplete contracts. See Hart (1995).

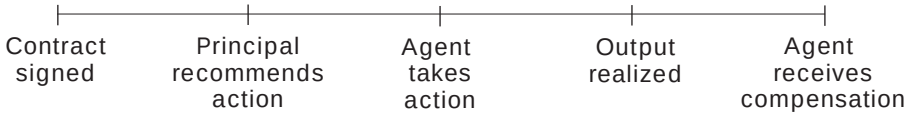
information he does have, namely, the project's output. In moral-hazard problems, one usually assumes that the output is publicly observed, that is, seen by both the principal and the agent. It follows that the only device the principal can use to encourage the agent is to make compensation depend on output. The principal, of course, wants to do this in such a way that the agent willingly takes the contracted action. The idea is that actions can be contracted upon, but *only* if they are consistent with the agent's incentives as determined by the compensation schedule. As we will see, compensation schedules can be effective at inducing actions but not without a cost.

It might be helpful to return to the agricultural example. Think of a landowner who lets a tenant farm a piece of his land. If the landowner spends his time on other activities, such as farming other pieces of land he owns, then he cannot monitor the tenant's efforts on the plot. He does not observe how carefully the tenant weeds the plot. Nor does he observe how many hours a day the tenant works the plot. All he can do is infer from the plot's yield the care and effort the tenant put into working the land. But this inference is far from perfect. If the crop fails, it could be because the tenant did not properly apply himself. But, it could also be the result of disease, insect infestation, insufficient rain, or a host of other factors beyond the tenant's control. Disentangling the effects of factors under the agent's control (his effort) and factors not under his control (the weather) is the essence of the moral-hazard problem because it leads to a trade-off between insurance and incentives. The contract should insure the tenant against events beyond his control, but it should also provide him an incentive to do what he is supposed to. Often, these are conflicting goals.

Formal Development of the Model

There are five sequential stages to the model. First, the principal and the agent agree to a contract. Second, as part of this contract, the principal recommends an action and then the agent decides whether or not to take it. These two steps are separated so that it is clear that the agent is choosing his action at that point. Then, the output is realized and, finally, the principal compensates the agent. Figure 1 summarizes these steps in a timeline.

Figure 1 Timeline of Moral-Hazard Problem



Environment

There are three variables that matter in this problem. First, there is the action that the agent takes. We identify an action by a and restrict it to lie in a set A, where A is an interval. (For example, A could be the interval between 0 and 1.) The second variable is the output, which we call q. Possible outputs lie in the set Q. For simplicity, we assume that there are only a finite number of elements in this set. The final variable is consumption c, which is restricted to lie in the set C. Like A, it is an interval.

The output q is determined by the agent's choice of action and a random shock that occurs after the agent has taken his action. The principal observes the output but not the shock or the agent's action. The idea is that he cannot infer from the output alone how hard the agent worked. It is most convenient to drop explicit references to the shock and instead describe the relationship between the action and the output by the conditional probability $p(q|a)$. This function is the probability distribution of output given the action. Because it is a probability distribution, $\sum_q p(q|a) = 1$ for each a. Finally, for simplicity we assume at this time that each output is possible for each action, that is, $p(q|a) > 0$, $\forall a \in A$, $q \in Q$.

Preferences

The agent cares about his consumption and his effort. We write his utility function as $u(c, a)$. For the moment, we do not make any assumptions on the form this function takes. The principal only cares about the project's surplus, that is, $q - c$. Depending on the model, the surplus may be negative, that is, the principal pays the agent out of his own funds. The principal's utility function is $w(q - c)$.

Deterministic Contracts

In the standard formulation of the moral-hazard problem, the model is solved for the optimal *deterministic contract*. A contract consists of the action the agent is supposed to take and the compensation schedule, that is, consumption as a function of output. The term "deterministic" refers to an assumed property of the contract, namely, that no randomization is allowed in the contract's terms. What we mean precisely by randomization is described in the next section.

Definition 1 *A deterministic contract is a recommended action* a *and an output dependent compensation schedule* c(q).

The Approach

Our goal is to find one of the best feasible contracts that satisfies some criterion. Economists do this by solving a constrained-maximization program.

These programs consist of an objective function and a set of constraints. The objective function ranks alternative contracts according to some criterion. The constraints describe the set of contracts that are feasible. In this problem, the constrained-maximization program represents the problem facing a principal who is trying to determine the best feasible contracts to give the agent.

Objective Function

In moral-hazard problems, it is usually assumed that the principal owns the project and designs the contract. The objective function then is the principal's utility. It is written

$$\sum_q p(q|a)w(q - c(q)).$$

Participation Constraint

The first constraint is the participation constraint; it is also sometimes called an individual rationality constraint. The variable U represents the value of outside opportunities to the agent, opportunities that are not usually explicitly modeled. The constraint represents the idea that the principal-agent relationship does not exist in a vacuum. Since the agent has other activities that he can do, he will only sign the contract if it is at least as good as the best of these outside opportunities. Though we do not explore the issue in this article, the level of U can have a strong effect on the optimal contract. The higher U is the less surplus the principal will be able to get from the project. The participation constraint is

$$\sum_q p(q|a)u(c(q), a) \geq U. \tag{1}$$

Incentive Constraints

Incentive constraints are the formal method of accounting for private information. To see how moral hazard restricts the set of feasible contracts, consider the following problem. Assume that the principal is risk-neutral, that is, $w(q-c) = q-c$, and that the agent's preferences are separable $u(c, a) = U(c) - V(a)$. We assume that $U(c)$ is concave ($U'(c) > 0$, $U''(c) < 0$) indicating that the agent dislikes riskiness in consumption. We also assume that $V(a)$ is nonnegative and increasing in the action, indicating that the agent prefers a lower action to a higher one. (The action can be thought of as the level of effort here.)

Now consider the following compensation schedule, $c(q) = \bar{c}$, that is, consumption is independent of the realized output. We assume that this $c(q)$, together with the assigned a, satisfy (1). Under this contract, the principal bears all of the risk. His consumption (the surplus) moves one-for-one with

fluctuations in the output. In contrast, the agent's consumption is unaffected by fluctuations in output because he receives a fixed payment. In fact, it is easy enough to show that this contract is the optimal one if the only constraint is (1). The first-order conditions on consumption are

$$\forall q, \quad \frac{1}{U'(c(q))} = \lambda, \tag{2}$$

where λ is the Lagrangian multiplier on the participation constraint. Only a contract with constant consumption satisfies these constraints.

Now consider what happens if this contract is chosen (and the contract does not assign the lowest action to the agent), but the agent's action is private information. Because the principal does not observe the agent's action, he cannot make him work the contracted action. The agent will take the action that is in his best interest, which is defined by *his* maximization problem, that is, $\max_a \sum_q p(q|a)[U(\bar{c}) - V(a)]$. But because consumption does not depend on output, the agent's action does not affect his consumption! Consequently, he takes the action that gives him the least disutility rather than the action recommended by the contract. Thus, the example contract, despite its desirable insurance properties, is not feasible.

Economists capture the effect of private information by using *incentive* or *incentive-compatibility* constraints. These constraints are simply a way of recognizing that any feasible contract will have to be compatible with the agent's incentives. For a contract $(a, c(q))$ to be incentive compatible, it must satisfy

$$a \text{ solves } \max_{\tilde{a}} \sum_q p(q|\tilde{a})u(c(q), \tilde{a}). \tag{3}$$

This constraint just says that the agent will take the action that is in his own best interest as determined by the compensation schedule $c(q)$.

Another way to write this constraint, one which will be more convenient later, is to express it by a direct pairwise comparison between taking the recommended action and taking all other actions. This representation is

$$\sum_q p(q|a)u(c(q), a) \geq \sum_q p(q|\hat{a})u(c(q), \hat{a}), \ \ \forall \hat{a} \in A. \tag{4}$$

As before, this constraint says action a must be optimal from the agent's perspective. Notice that there are many constraints.

We can now proceed to the constrained-maximization program. To repeat, the problem is to find one of the best feasible deterministic contracts. For a contract to be feasible, it must satisfy both the participation and the incentive-compatibility constraints. An optimal deterministic contract is a solution to the program below.

Program with Deterministic Contracts

$$\max_{a,c(q)} \sum_q p(q|a)w(q - c(q))$$

s.t. (1) and (4).

Properties of Solutions

This program is surprisingly difficult to analyze. The problem is that if A is a continuum, then there are a *large* number (a continuum) of incentive constraints. To put these constraints into a manageable form, researchers normally try to substitute the first-order condition (FOC) from the agent's problem (3) for (4). When the agent's preferences are separable, that is, $u(c, a) = U(c) - V(a)$, the first-order condition is $\sum_q p_a(q|a)U(c(q)) = V_a(a)$. The programming problem with this constraint is much easier to analyze than the programming problem with (4).

Unfortunately, this "first-order approach" is not valid in general. First-order conditions are only sufficient for describing an optimum of a concave function. There is no guarantee, however, that at the optimal $c(q)$ the function $\sum_q p(q|a)u(c(q), a)$ is concave in the action.

What researchers have attempted to do is to find conditions that make this substitution valid. These conditions, however, are restrictive. In particular, the most commonly used assumptions are that the agent is risk-averse, his preferences are separable, the principal is risk-neutral, and the technology $p(q|a)$ satisfy the monotone likelihood ratio property (MLRP) and the convexity of the distribution function condition (CDFC).[7]

The MLRP condition is $\frac{p_a(q|a)}{p(q|a)}$ increasing in q. This condition guarantees that output is increasing stochastically in effort, that is, higher output is more likely for higher effort than lower efforts. If $P(q|a)$ is the cumulative distribution function, then CDFC is

$$P(q|\alpha a + (1-\alpha)a') \leq \alpha P(q|a) + (1-\alpha)P(q|a'), \quad \forall a, a' \in A, \quad \forall \alpha \in (0,1),$$

and for all q. This condition provides a form of diminishing returns in effort. Both of these conditions are rather restrictive, and many natural technological specifications do not satisfy them both. For example, the distribution $q = a + \theta$, where θ is normally distributed, satisfies MLRP but does not satisfy CDFC.

When the first-order approach is valid, the optimal contract can be fairly well characterized. For example, if the previous assumptions on preferences and

[7] For more details than are presented here, see Hart and Holmström (1987). Rogerson (1985) contains an important early proof of this result. Jewitt (1988) extends these results for the case of a risk-neutral principal, and Alvi (1997) extends these results for nonseparable preferences.

technology holds, then the first-order condition to the constrained-maximization problem (not the agent's subproblem) is

$$\frac{1}{U'(c(q))} = \lambda + \mu \frac{p_a(q|a)}{p(q|a)}.$$

The variables λ and μ are the Lagrangian multipliers on the participation constraint (1) and the incentive constraint (FOC version), respectively. Compare this with (2), the FOC for the full-information program. The only difference is the term $\mu \frac{p_a(q|a)}{p(q|a)}$, which is the effect on the solution from adding the incentive constraints. Now, because of private information, consumption does depend on output. In particular, one can easily verify that when MLRP is satisfied, the term $\frac{p_a(q|a)}{p(q|a)}$ is increasing in q. Because the Lagrangian multipliers are nonnegative in this problem, this property implies that consumption is increasing in output. Unfortunately, most technologies do not satisfy MLRP. Consequently, optimal consumption-sharing rules need not be monotonic.

An Example

The following executive compensation problem illustrates what a "typical" compensation schedule may look like.[8] In this example, the owners of a bank (the principal) must devise a compensation schedule for their chief executive officer (the agent). The CEO is risk-averse and would rather apply his effort to activities that do not necessarily increase bank profits. To simplify the model, we assume that an increase in his action gives the bank higher expected profits but gives the CEO lower utility. Furthermore, we assume that the technology satisfies the MLRP.[9]

The solid line in Figure 2 shows the optimal compensation schedule as a function of the bank's profits. Compensation increases with bank profits as one would expect for a technology that satisfies MLRP. To illustrate the role of private information, we also report the optimal full-information compensation schedule, calculated from the solution to the full-information program. (That program is the private-information program without the incentive constraints.) Under full information, the risk-averse executive bears no risk.[10] He is fully insured against all fluctuations and the risk-neutral bank absorbs all fluctuations in output.

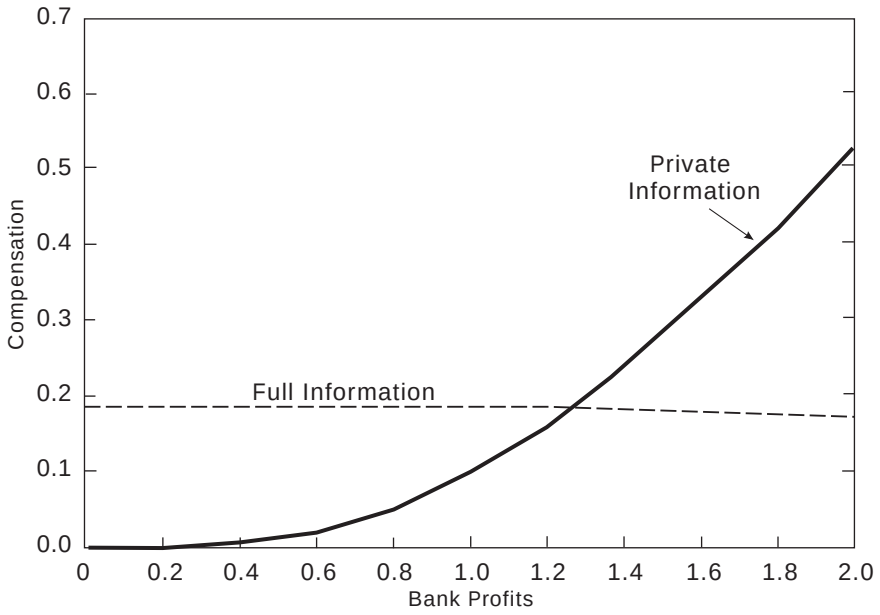
Our comparison of the compensation schedules reveals one cost of private information. For incentive reasons, some insurance must be sacrificed. There

[8] There is a literature that examines executive compensation using moral-hazard models. See Jensen and Murphy (1990) and Haubrich (1994).

[9] The technology does not satisfy CDFC. We computed this example using the techniques described later in the article.

[10] Any variation in the schedule is due to numerical approximation.

Figure 2 Optimal Compensation Schedules



is another cost, however, not illustrated in Figure 2: the program implements a lower action with private information than with full information. Apparently, in this example, it is too expensive to implement the optimal full-information action when there is private information.

2. CONTRACTS WITH LOTTERIES

In our review of the moral-hazard problem we restricted the contract space to deterministic contracts. In many cases, however, randomization in the terms of the contract may improve welfare.[11] This section covers these issues and explains how lotteries are a necessary component in developing the linear program.

There are two types of randomization that one may place in the contract. The first type involves making the recommended action a random. Instead of choosing a single action a, the principal chooses a probability distribution

[11] With randomization one can show by *Revelation Principle*-type arguments that no communication game between the principal and the agent can improve upon the direct mechanism considered in this problem. See Myerson (1982).

over all of the possible recommended actions. We will call this probability function $\pi(a)$. The second type of randomization is contained in the compensation schedule. Instead of choosing $c(q)$, the principal chooses $\pi(c|q, a)$, that is, a probability distribution of consumption given each realized output and each recommended action with positive probability. It is important to note that contracts with lotteries do not preclude deterministic contracts. Deterministic contracts are still feasible, but they are simply degenerate lotteries over the relevant sets.

Definition 2 *A contract with lotteries is a probability distribution over recommended actions,* π(a)*, and a probability distribution over consumption as a function of the output and the recommended action,* π(c|q,a).

Economic Role of Lotteries

There has been little research on the role of lotteries in moral-hazard problems. What we do know is that under certain strong conditions, randomization is undesirable, and under certain weaker conditions, it is desirable. In general, analytical results concerning randomization appear to be quite difficult to derive. We will not attempt to summarize the results but will point out a few apparent ones. The reader interested in more detail can examine Arnott and Stiglitz (1988) and the citations contained within.

If the agent is risk-averse and his utility function is separable, and if the principal is risk-neutral, then we know that consumption lotteries are *not* optimal. To see this result, imagine a contract with a nondegenerate consumption lottery. Replace the consumption lottery with a deterministic compensation schedule that leaves the agent's utility unchanged for any realization of q and a. Because of concavity of the utility function, the schedule uses less consumption in expectation without violating the participation and incentive constraints. The reduced expected payment can be returned to the principal, increasing his utility. With nonseparable preferences, however, such similar conditions are much harder to provide. See Arnott and Stiglitz (1988) and the citations contained within.

Cases with optimal nondegenerate consumption lotteries require either a nonconvexity or nonseparability in preferences. If the agent is a risk-seeker, consumption lotteries (for reasons having nothing to do with incentives) will be valuable. Consumption lotteries may also be desirable if the agent's action affects his risk-aversion. Cole (1989), in a different model, provides one such example.

Action lotteries have been less systematically studied than consumption lotteries. Arnott and Stiglitz (1988) demonstrate that action lotteries will occur if in the space of contracts with deterministic actions, the principal's expected utility is nonconcave in the agent's expected utility. Unfortunately, conditions

under which this situation occurs are far from obvious. Lehnert (1998) and Prescott (1998) contain examples with action lotteries (or something similar).

To summarize, analytical results on the optimality of nondegenerate lotteries are difficult to derive. It seems then that computation of examples should play an important role in examining this issue.

Concerns with Lotteries

One common criticism of lotteries is that they are typically not found in the explicit terms of contracts. While this may be true, the complicated terms of optimal deterministic contracts are also typically not found in the explicit terms of contracts. This criticism then is not one aimed at the use of lotteries per se but at contract theory in general.

One response to the criticism is that explicit terms in contracts are not necessarily an accurate guide to its true terms. Frequently, there are implicit terms or contingencies in contracts. For example, Townsend and Mueller (1998) contains examples of implicit contingencies in sharecropping contracts.

Another response to the criticism is that lotteries may be indistinguishable from other state-contingent transfers and may represent unmodeled transactions. Cole and Prescott (1997) demonstrate that ex ante randomization, the sort defined by the action lotteries, need not be implemented through individual contracts but, instead, can be implemented through ex ante gambling that generates wealth inequality. In a developing economy context Lehnert (1998) argues that ex ante randomization may be implemented by financial intermediaries like Rotating Savings and Credit Associations (ROSCAs). ROSCAs frequently use lotteries to determine the order in which members receive funds. In both papers, there is a continuum of agents so the ex ante lotteries represent equilibrium *fractions* of the population. Interpreting the lotteries as equilibrium fractions of the population is a particularly intuitive interpretation that we will return to later in our example. It is also possible that the consumption lotteries may represent state-contingent transfers tied to exogenous events, transfers that one might interpret as part of a normal financial contract.

In some arrangements, lotteries are an explicit part of the economic mechanism. For example, the Internal Revenue Service randomly audits tax returns. Some indivisible goods or duties are assigned by lottery. For example, citizens are assigned to jury duty by lot. No doubt in the case of employment decisions sometimes the difference between a successful job candidate and an unsuccessful one is simply luck.

In my view, if we are to take the discipline of optimizing seriously, we should not restrict the contract space unless there is an economic argument for excluding these types of contracts.[12] This issue is not merely a technical

[12] See Boyd and Smith (1994), who argue that the gains from randomized auditing in resolving firm bankruptcy are miniscule.

one. In Lehnert (1998), policy prescriptions differ between an economy with deterministic contracts and one with lotteries. If for a particular problem one decides that the evidence does not support lotteries, then one needs to choose exogenous parameters so that the optimal contract is deterministic. The theory then provides an additional layer of requirements on the class of models we work with.

Making the Moral-Hazard Problem a Linear Program

We now show how to develop the moral-hazard program with lotteries as a linear program. This article does not deal with the topic of linear programming in any detail. There are plenty of well-written references on the topic. It does, however, contain a brief review of linear programs in Appendix A, summarized in the following paragraph.

A linear program is a constrained-maximization program that satisfies the following conditions:

1. The objective function and the constraints are linear;
2. There is a finite number of variables; and
3. There is a finite number of constraints.

By adding lotteries, condition 1 is satisfied. Lotteries have the further advantage that they are nonnegative by definition so the choice variables in the program are nonnegative, a requirement of the standard form of linear programming. (See Appendix A.) To satisfy conditions 2 and 3 another assumption is necessary, namely, that the sets C and A, in addition to the set Q, each contain a finite number of elements. For example, if A has two elements $\{a_1, a_2\}$, then, with respect to the recommended action, the principal chooses two variables, $\pi(a_1)$ and $\pi(a_2)$, the probability of a_1 and the probability of a_2. In general, the elements in these sets—often called the grids—can be made arbitrarily large, providing an approximation to the continuum case.

The derivation of the linear program is contained in Appendix B, but to summarize here, the deterministic contract $(a, c(q))$ is replaced with the contract $(\pi(a), \pi(c|q, a))$. Next, the objective function and constraints are algebraically manipulated so that the choice variable is the probability distribution $\pi(c, q, a)$. The term $\pi(c, q, a)$ is the unconditional probability distribution of each possible consumption, output, and action triplet.

Usually, the new choice variable is the probability distribution of each point in the grid $P = C \times Q \times A$, though sometimes P is a subset of the grid. For expositional ease, we assume the former in this section. Since C, Q, and A all have finite numbers of elements, P has a finite number of points. For example, if $C = \{c_1, c_2, \ldots, c_l\}$, $Q = \{q_1, q_2, \ldots, q_m\}$, and $A = \{a_1, a_2, \ldots, a_n\}$, then P has lmn elements. Each element is indexed by a (c_i, q_j, a_k) vector. One way to list all the elements is

$$
\begin{gathered}
(c_1, q_1, a_1), \\
(c_2, q_1, a_1), \\
\ldots \\
(c_n, q_1, a_1), \\
(c_1, q_2, a_1), \\
(c_2, q_2, a_1), \\
\ldots \\
(c_l, q_m, a_n).
\end{gathered}
$$

The choice variable $\pi(c, q, a)$ is an lmn-dimensional vector. The value of $\pi(c_i, q_j, a_k)$ is the probability of the (c_i, q_j, a_k) triplet occurring.

Moral-Hazard Program with Lotteries

$$\max_{\pi} \sum_{c,q,a} \pi(c, q, a) w(q - c)$$

$$\text{s.t.} \quad \sum_{c,q,a} \pi(c, q, a) u(c, a) \geq U, \tag{5}$$

$$\sum_{c,q} \pi(c, q, a) u(c, a) \geq \sum_{c,q} \pi(c, q, a) \frac{p(q|\hat{a})}{p(q|a)} u(c, \hat{a}), \quad \forall a, (\hat{a} \neq a) \in A \times A, \tag{6}$$

$$\forall \bar{q}, \bar{a}, \quad \sum_{c} \pi(c, \bar{q}, \bar{a}) = p(\bar{q}|\bar{a}) \sum_{c,q} \pi(c, q, \bar{a}), \tag{7}$$

$$\sum_{c,q,a} \pi(c, q, a) = 1, \text{ and } \forall c, q, a, \quad \pi(c, q, a) \geq 0. \tag{8}$$

This program is the same as the deterministic program discussed earlier but includes lotteries. Indeed, except for the addition of two constraints, the programs are identical in structure. Again, the objective function is the principal's utility. Constraint (5) is the participation constraint. The incentive constraints are (6). On the right-hand side of (6) is the likelihood ratio $\frac{p(q|\hat{a})}{p(q|a)}$. This ratio is very important in private-information models because it influences the ability of the principal to reward and punish the agent. For example, if the ratio is low for some q, then high compensation would reward an agent for taking the recommended action more than it would reward him for taking the deviating action $\hat{a}$. Similarly, if the ratio is high for some q, then low compensation would punish an agent who takes the deviating action $\hat{a}$ more than it punishes him for taking the recommended action a.

The last two sets of constraints are specific to problems with lotteries. They are not analogous to any constraints in the deterministic formulation.

Constraints (7) are the technology, or Mother Nature, constraints. As Appendix B describes in more detail, these constraints are added to the program so that despite choosing $\pi(c, q, a)$, the principal really only chooses the contract $(\pi(a), \pi(c|q, a))$; that is, the constraints ensure that a feasible $\pi(c, q, a)$ is consistent with $p(q|a)$. The final constraint (8) ensures that $\pi(c, q, a)$ is a lottery, i.e., it sums to one and each element is nonnegative.

It is straightforward to verify that this problem is a linear program. Lotteries deliver linearity of the objective function and the constraints. The grids then ensure that $\pi(c, q, a)$ is a finite-dimensional vector and that there are a finite number of constraints.

Computation

In general, the two limitations to computing linear programs are computer speed and memory. Roughly, the larger the dimensions of the problem, the more memory is required and the longer the problem takes to solve. Unfortunately, as the sizes of the grids increase, these problems quickly grow in size. For example, if l, m, and n are the number of elements in C, Q, and A, respectively, then the number of variables in the problem is lmn. The number of constraints grows similarly. There is one participation constraint, $n(n-1)$ incentive constraints, nm technology constraints, and one probability measure constraint. For example, if there are 10 actions, 20 outputs, and 50 consumptions, then the linear program has 10,000 variables and 292 constraints.

The size of the linear program that can be computed depends on the user's software and hardware. Although recent advances in computer technology have greatly increased the size of programs that may be computed, size limits may still arise. If size limits bind, the programmer may be forced to restrict the number of elements in the grids. Whether this is a serious issue for the results depends on the problem. For some problems, grids are the natural specification, such as when there are indivisibilities. For many problems, the grid does not need to be very fine.[13]

With commercial code or high-quality public domain code written in a compiled language, the biggest limitation on computation is probably memory. When this problem arises, the Dantzig-Wolfe algorithm can be used to com-

[13] The discussion of the grid brings up the issue of *grid lotteries*. Grid lotteries are lotteries over adjacent points in the grid, be they consumption or action lotteries. They frequently appear when computing these problems, so it is important to understand that they have no economic content. They only reflect the approximation of a continuum inherent in the grids. For example, the executive compensation schedule presented in Figure 2 was computed as a linear program and the optimal solution contained some grid lotteries. This situation is most obvious in the full-information contract, where the compensation schedule is not quite horizontal. This deviation would disappear if the grid was made successively finer. Indeed, one strategy for computing these problems is to solve them first with a relatively coarse grid and then add grid points to areas where there is positive weight in the solution.

pute a moral-hazard program. This algorithm solves linear programs that have a special constraint structure, a structure that the moral-hazard program satisfies. Prescott (1998) solves several moral-hazard programs using this algorithm. He finds that for given memory limitations substantially larger problems can be computed using this method.

3. A BANK REGULATION EXAMPLE

In this example, we study a moral-hazard problem where the principal is a bank regulator and the agent is a bank. The bank is funded by insured deposits. Because depositors are insured, they are unconcerned about the bank's return. For simplicity, we drop all reference to them from the problem. Since depositors have no incentive to monitor, it is the job of the bank regulator to devise a regulatory regime.

The bank can engage in either an opaque investment strategy or a transparent one. The opaque strategy consists of investing in assets that are difficult for outsiders to evaluate and are potentially risky. Examples of opaque strategies include business loans or complicated derivative contracts. The transparent strategy consists of investing in safe assets that are easy for outsiders to evaluate. Examples could include holdings of Treasury bills or other money market assets. We assume that the bank may engage in only one investment strategy. It cannot split its assets between an opaque and a transparent strategy. Also, for simplicity we assume that the bank has a fixed deposit base so it cannot choose the amount of funds to invest.

The information assumptions on the bank's actions are slightly different from the standard formulation, although except for a minor modification to the incentive constraints, the program will not change. The strategy the bank engages in is public information, that is, both the bank and the regulator know whether the bank invests in opaque or transparent assets. But if the bank engages in the opaque strategy, there is some private information. Specifically, we allow the bank to choose the riskiness of its opaque strategy without letting the regulator observe this choice. In contrast, if the transparent strategy is chosen, we assume that there is no private information. Finally, the bank's return (net of payments to depositors) is public information.

This problem is designed to study two issues about bank regulation. First, the model can exhibit coexistence of multiple regulatory regimes, one for each possible investment strategy adopted by the bank. Second, it can be used to study the role of fines in mitigating excessive risk-taking. Fines and other penalties are an essential part of the precommitment approach, a recent proposal for regulating banks' trading accounts.[14]

[14] See Kupiec and O'Brien (1995), Prescott (1997), and Marshall and Venkataraman (1998). This approach advocates letting banks choose their capital level but fining them if losses exceed

If the bank engages in the opaque investment strategy, we assume that the bank chooses only the variance of its returns. Unlike most problems, the bank's choice of action has no affect on the mean of returns. Also unlike most problems, the bank's utility is not directly affected by the action chosen. To create a conflict of interest between the regulator and the bank, we first assume that the bank has limited liability. Because of its limited liability the bank prefers a high variance return as its losses are limited to zero.[15] Our second assumption that will guarantee a conflict of interest is to assume that there is a social loss to bankruptcy, an event defined as a negative return by the bank. The loss creates a dislike for variance on the part of the regulator who we assume is benevolent and wishes to minimize social losses. Our setup creates a trade-off between risk-seeking behavior by the bank (driven by the limited liability) and the social cost from poor realizations of the bank's investment. The difference between the trade-off here and the one described earlier in the executive compensation problem is that in the compensation problem there was a trade-off between insuring the executive and giving him incentives to take the desired action.

The regulator has two devices available to influence the bank's actions. First, it can regulate the investment strategy of the bank. The regulator by fiat can decide which lines of investment that the bank may engage in. Second, the regulator may impose return-dependent fines, though the regulator's ability to do this is limited.

We limit regulators' ability to levy fines in two ways. First, with limited liability, a fine cannot exceed the bank's return. Second, the amount of the fine that the regulator may impose in any state must lie between zero and an upper bound. Reasons for these bounds are not modeled, but they could exist for political reasons. For example, it might be difficult politically to make explicit payments (negative fines) to banks. Furthermore, we assume that fines are not merely a transfer; instead, we assume that a fraction of collected fines represent a social loss to society. Consequently, not only does the regulator want to prevent bankruptcy, but it also prefers not to impose fines.

Setting up the Linear Program

In this problem it is more convenient to put the level of fines in the grid rather than the bank's profit (consumption). Bank profits are easy enough to calculate, as they are the difference between the return and the fine, that is, $c = q - f$.

Grid

The bank can engage in either of two investment strategies. If the bank engages

capital.

[15] Limited liability makes the bank a risk-seeker over certain ranges of returns.

in the transparent strategy, then we assume that there is only one action the bank can take. We write this set as $A_{tr} = \{0.0\}$. If the bank engages in the opaque investment strategy, its set of feasible actions is $A_{op} = \{0.4, 0.7, 1.0, 1.3\}$. The value of the action corresponds to the standard deviation of the project's return. To be consistent with previous notation we can write $A = (A_{tr}, A_{op}) = \{0.0, 0.4, 0.7, 1.0, 1.3\}$. Because the investment strategy is public information, the incentive constraints only apply to actions in the set A_{op}.[16]

It is easiest if we make the return grid depend on the investment strategy. We will make the extreme assumption about the transparent strategy that there is no variance in its return. In particular, we assume that $Q_{tr} = \{0.5\}$ and that $Q_{op} = \{-0.3, -0.25, -0.2, \ldots, 1.7\}$.

Because both the principal's and the agent's utilities in fines are linear (they are both risk-neutral), it is necessary to have only two points in the fine grid, an upper and a lower bound. Lotteries over the bounds can obtain any intermediate level of fine without being an approximation. Indeed, the weighted sum of the bounds will be interpreted as the actual level of the fine.

By assumption the lower bound on fines is zero. But because of limited liability, the upper bound on fines must be less than the bank's return. The fine must also be less than the exogenously set upper bound on fines, which is 0.12 in this example. We capture these constraints by writing the fine grid as return dependent. It is $F(q) = \{0\}$ if $q < 0$ and $F(q) = \{0, \min\{q, 0.12\}\}$ if $q > 0$. The return and action grid for this problem is thus $(Q_{tr} \times A_{tr})$ followed by $(Q_{op} \times A_{op})$, that is, $\{(0.5, 0.0), (-0.3, 0.4), (-0.25, 0.4), \ldots, (1.7, 1.3)\}$. The entire grid is then created by appending the appropriate $F(q)$ for each return, action pair.

Preferences

The bank is risk-neutral with limited liability so its preferences are $u(f, q, a) = \max\{q - f, 0\}$. Notice that the bank's utility is not affected, at least not directly, by the action it chooses. The regulator only cares about the social costs of bankruptcy and fines. Bankruptcy is defined as negative return, that is, if $q < 0$. The regulator's preferences are

$$w(f, q) = \begin{cases} -1 & \text{if } q < 0, \\ -0.5f & \text{otherwise.} \end{cases}$$

These preferences assume that it is costly to resolve bankruptcies and to collect

[16] Here is the minor change from the standard formulation that we referred to earlier. Keeping the notation of the standard formulation, the incentive constraints are now

$$\sum_{c,q} \pi(c, q, a)u(c, a) \geq \sum_{c,q} \pi(c, q, a)\frac{p(q|\hat{a})}{p(q|a)}u(c, \hat{a}), \quad \forall a, (\hat{a} \neq a) \in A_{op} \times A_{op}.$$

Notice that there are no incentive constraints if the bank adopts the transparent strategy.

fines. By maximizing this utility function, the regulator is minimizing the social costs of bankruptcy and fines.[17]

Technology

The probability distribution of the bank's return when it adopts the transparent strategy is predetermined by our assumption that there is no variance in the return. The probability of $q = 0.5$ when $a = 0.0$ is 1.0. If the bank adopts the opaque strategy, then its action affects the variance of its returns. Let $f(q|a)$ denote a normal distribution with mean -0.7 and standard deviation a evaluated at q. The conditional probability of each output for the opaque strategy is

$$p(q|a) = \frac{f(q|a)}{\sum_{Q_{op}} f(q|a)}.$$

The denominator normalizes the function to sum to one, which we need to do because of the grid. The technology is probably better understood by considering Figure 3. Notice that the higher the standard deviation (the higher actions), the higher the probability of bankruptcy (the area under each curve for $q < 0$).

In this problem, the regulator dislikes bankruptcy and fines because of their social cost. Without limited liability, the bank would not care which action it takes since each has the same mean. With limited liability, however, the bank receives no disutility from negative returns, so it prefers high-variance actions. This preference is the source of the conflict between the bank and the regulator. The only tools available to the regulator in this problem are its limited ability to impose fines on the bank and its ability to order the bank to engage in either of the investment strategies.

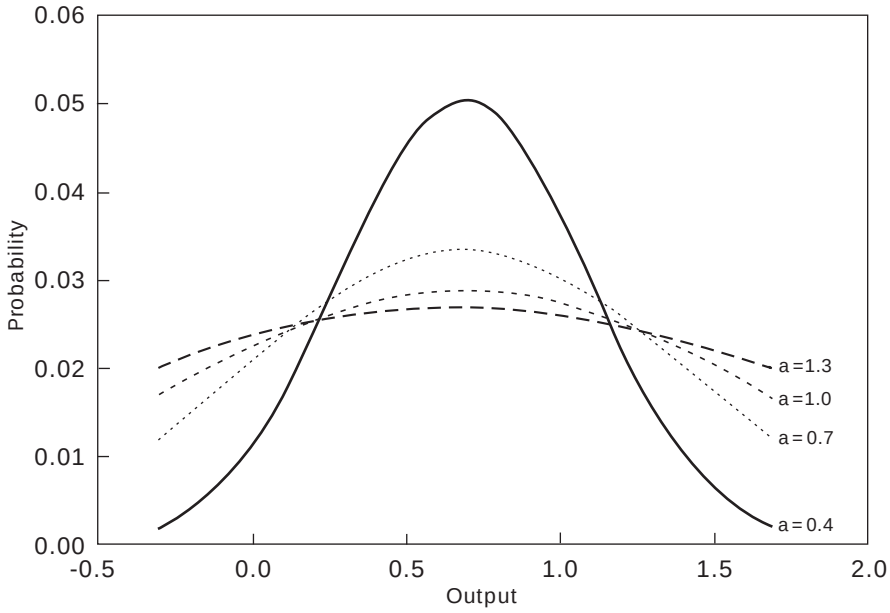
Solution

The linear program was solved for the case where the bank receives $U = 0.60$ utils. For this parameterization, the optimal action for the bank is a lottery over the transparent and opaque strategies. The bank is assigned action $a = 0.0$ (the transparent investment strategy) with probability 0.47, and it is assigned action $a = 0.4$ under the opaque investment strategy with probability 0.53.

When the bank is assigned the transparent strategy, no fines are imposed because there are no incentive constraints for this strategy. When the bank is assigned the opaque strategy, however, the regulator imposes fines to prevent

[17] Arguably, one would want to add a constraint that the regulator recovers enough resources in fines to compensate depositors in the event the bank loses money, that is, when returns are negative. In the interest of keeping the example as close as possible to the structure of the basic moral-hazard program, we left out this constraint. However, in the solution to the example computed later, fines are sufficiently high that in expectation they cover expected losses (even assuming that half of the fines are lost as part of the fine collection process).

Figure 3 Probability of Each Output Given Each Action for the Opaque Strategy



the bank from taking the higher risk actions. Figure 4 shows the optimal fine schedule if the bank is assigned the opaque strategy.

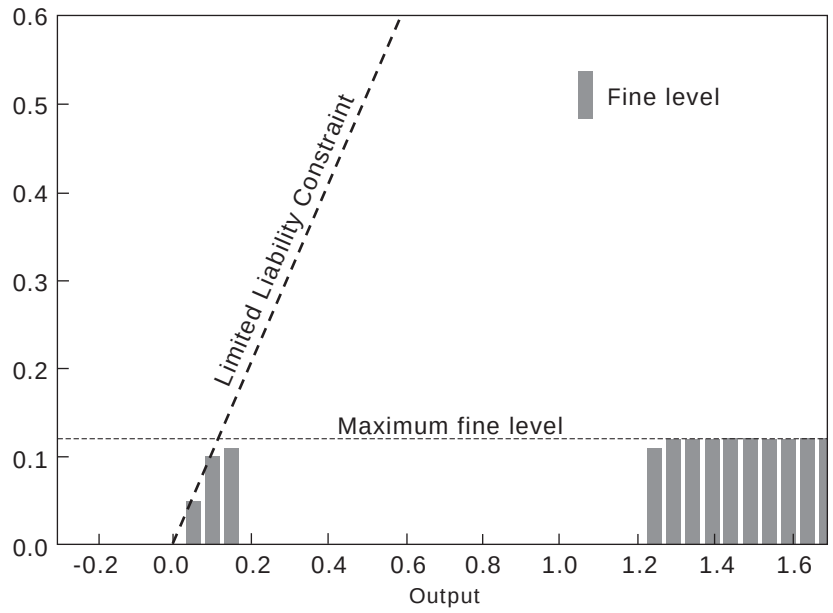
Because the regulator wants to minimize the amount of fines imposed, it wants to limit their use to when they are most effective. They are most effective if imposed on returns that are likely if the bank takes the high-variance action and unlikely if the bank takes the low-variance action (like it is supposed to). In this problem, low and high returns satisfy this condition. With limited liability, however, fines can hardly be imposed on the bank when it realizes a low return. In contrast, although the limited liability constraint is not binding for high returns, the fine ceiling binds instead.[18]

Interpretation

We interpret the investment-strategy lottery as representing the coexistence of two different regulatory regimes. To see this, imagine that instead of there

[18] One more point to note about this example. The binding incentive constraint in this example is not the one corresponding to the downward adjacent action (the discrete analog to the FOC on the agent's subproblem); instead it is the one corresponding to the highest variance action. Consequently, the first-order approach would not work in this example.

Figure 4 Optimal Fine Schedule



being one bank, there are a large number, or a continuum, of them. Each of these banks is identical to the one in the example and is treated identically, at least ex ante. Mathematically, the linear program is unchanged but now the lotteries represent equilibrium fractions of the population of banks. Under this interpretation, 47 percent of the banking sector would be engaged in transparent investment strategies and would not be subject to any fines. These banks resemble narrow banks. The remaining 53 percent of the banking sector would be engaged in the opaque investment but would be subject to a very strong regulatory regime.

Parallels to these results exist in U.S. financial regulation. Thrifts, credit unions, banks, and mutual funds all face different limitations on their investment strategies. For example, thrifts must hold a certain fraction of assets in housing and consumer lending. Credit unions are limited in their commercial lending. Money market mutual funds limit investment to money market instruments. Furthermore, all of these institutions have their own regulators and set of rules.

The investment strategy lottery is an example of an action lottery. As we discussed earlier, these lotteries can occur if there is a nonconvexity in utility space. In this example, banks get higher utility (and the regulator less utility) if the bank takes the opaque strategy than if it takes the risky strategy. The

lottery ensures that the bank gets its reservation utility in expectation. For different levels of reservation utility it is possible that all banks may engage in deterministic strategies.

In this example, banks receive higher utility from taking the opaque investment, so why would any bank take the transparent strategy? For the lottery to be implemented there needs to be some kind of control over the number of banks that may engage in the opaque investment strategy. In financial regulation, chartering may serve such a role. Many financial institutions must receive a charter to operate. Regulators' control of charters can limit the fraction of banks that can engage in certain strategies. A less command-and-control method would be to price deposit insurance so that banks pay the social costs of their investment strategy. Transparent banks would pay no deposit premiums, while opaque banks would pay enough to cover the expected social costs they cause, net their expected fine payments.

Another important feature of this example is the important role that fines play in mitigating risk-taking behavior. Like the action lottery, the optimal fine schedule (or at least a portion of it) parallels existing regulatory practice. The fines imposed for low returns are suggestive of the early closure rules under FDICIA. Under FDICIA, regulators can close banks when their capital drops to a low enough level. There is no direct parallel, however, for the high return fines. If we take the extreme view that existing regulatory rules cannot be improved upon, then we might say that the technology (the $p(q|a)$) in the example is incorrectly specified. And no doubt the technology is incorrectly specified; it was chosen to illustrate lotteries and the role of fines, not to match data. Another source of misspecification is that the model is missing crucial elements that would preclude the imposition of such fines. For example, well-run and innovative organizations should generate higher returns than other banks and such activities should not be discouraged.

Still, the high-return fines are suggestive of an interesting modification to this model. Dye (1986) adds to the moral-hazard model by also allowing the principal to verify ex post the agent's action but only at a cost. The problem for the principal is to determine the returns (if any) under which he should spend resources to verify the agent's action. For the parameters in our example, we conjecture that it would be optimal for the regulator to verify whenever there are high returns (the same returns under which it is optimal to impose fines in the present example). Indeed, the optimal regulatory regime would probably entail a combination of costly ex post verification with fines.[19]

[19] Selective costly ex post verification is used in fair lending examinations by the Federal Reserve. Presently, one portion of the Federal Reserve's fair lending enforcement procedures starts with a statistical analysis of Home Mortgage Disclosure Act (HMDA) data. This data, which banks and other financial institutions are required by law to collect and report, contains information on each loan applicant. The information includes variables such as the loan applicant's race, the loan

4. CONCLUDING COMMENTS

The strength of computation lies in its ability to answer quantitative questions and its ability to investigate models that are difficult to analyze. The moral-hazard problem is one such model, of particular importance to many problems in economics, including financial regulation. Still, the basic model is limiting in many ways. Compensation is the only device available to the principal but in practice other mechanisms are important. For example, bank regulators receive reports, they monitor, and they observe signals. Many papers have investigated variants on the moral-hazard model, though not many have used the methods developed in this article. Appendix C lists the few such papers that use linear programming to solve examples. Since there are considerable difficulties in analyzing the basic moral-hazard model, one would expect that these difficulties would still exist in variants on the model and, therefore, numerical methods should be increasingly valuable as an analytical tool. With continued rapid advances in computer hardware and software, computation should become an increasingly effective way to study moral-hazard and other private-information problems.

APPENDIX A LINEAR PROGRAMMING

A linear program written in standard form is

$$\max_{\mathbf{x} \geq 0} \mathbf{cx}$$
$$\text{s.t. } \mathbf{Ax} = \mathbf{b},$$

where $\mathbf{c}$ is a $(1 \times n)$ vector, the choice variable $\mathbf{x}$ is an $(n \times 1)$ vector, $\mathbf{b}$ is an $(m \times 1)$ vector, and $\mathbf{A}$ is an $(m \times n)$ matrix, often called the coefficient matrix and *not* related to the set of actions A.

For a more complete description, see any linear programming textbook such as Luenberger (1973) or Bertsimas and Tsitsiklis (1997). The important points to note are that the objective function and the constraints are linear in $\mathbf{x}$, that $\mathbf{x}$ is a finite-dimensional vector, and that there are a finite number of

size, and the bank's decision. Some other information is reported but overall only a rather limited set of variables is collected. (Applicant credit history, for example, is not reported in the HMDA data.) If statistical analysis of an institution's HMDA data reveals a disparity in loan acceptances by race, then the analysis proceeds to a deeper and more costly review. For larger institutions, the next step in the analysis is to draw a sample of loan files from which examiners collect a much more detailed set of information on loan applicants than is available in the HMDA data. This detailed data set is then statistically analyzed, and if race appears to be statistically significant, then white-minority matches are generated and examiners look at the actual loan files to perform the deepest (and costliest) review. For more information see Avery, Beeson, and Calem (1997).

constraints. Neither the equality constraints nor the nonnegative values of **x** are critical features. Problems with inequality constraints or variables that may be negative can be easily converted to standard form.

Two classes of linear programming algorithms are presently in use. The most common are simplex-based routines. Simplex-based algorithms move along the frontier of the constraint set until an optimum is reached. The simplex algorithm was developed in the 1940s by Dantzig and has proven to be an efficient method in practice for computing solutions to linear programs. More recently, interior point algorithms have been developed. These algorithms move through the interior of the constraint set. It is commonly reported that the simplex algorithm is faster for small problems, but for large, sparse (lots of zeros in the constraint matrix) problems interior point methods can be very effective.

APPENDIX B DERIVATION OF THE LINEAR PROGRAM

With lotteries, the choice variables are $\pi(a)$ and $\pi(c|q,a)$. When placed into the standard program, the program becomes

$$\max_{\pi(a),\pi(c|q,a)} \sum_{c,q,a} \pi(c|q,a)p(q|a)\pi(a)w(q-c)$$

$$\text{s.t.} \sum_{c,q,a} \pi(c|q,a)p(q|a)\pi(a)u(c,a) \geq U, \tag{9}$$

$$\forall a \ni \pi(a) > 0, \sum_{c,q} \pi(c|q,a)p(q|a)u(c,a) \geq \sum_{c,q} \pi(c|q,a)p(q|\hat{a})u(c,\hat{a}), \forall \hat{a} \in A, \tag{10}$$

$$\forall q, \forall a \ni \pi(a) > 0, \quad \sum_{c} \pi(c|q,a) = 1, \pi(c|q,a) \geq 0, \tag{11}$$

$$\sum_{a} \pi(a) = 1, \text{ and } \pi(a) \geq 0. \tag{12}$$

This program is not a linear program because neither the objective function nor the participation constraint is linear.

To make it a linear program, we use the identity

$$\pi(c,q,a) = \pi(c|q,a)p(q|a)\pi(a) \tag{13}$$

and make the joint distribution $\pi(c,q,a)$ our choice variable.

Technology Constraints

By choosing the joint probabilities $\pi(c, q, a)$, the principal is choosing the conditional probabilities $\pi(c|q, a)$ and the unconditional probability $\pi(a)$, as is consistent with the formulation of the problem. The conditional distribution on the technology $p(q|a)$, however, is exogenous. We can keep this distribution exogenous by adding the following constraints

$$\forall \bar{q}, \bar{a}, \quad \sum_{c} \pi(c, \bar{q}, \bar{a}) = p(\bar{q}|\bar{a}) \sum_{c,q} \pi(c, q, \bar{a}). \tag{14}$$

If a joint distribution $\pi(c, q, a)$ that satisfies (14) is chosen, then the principal has only implicitly chosen the contractual terms $\pi(a)$ and $\pi(c|q, a)$.

Incentive Constraints

The incentive constraints guarantee that the agent always takes the recommended action. Thus, the constraints are for any action recommended with positive probability, so for all a such that $\pi(a) > 0$,

$$\sum_{c,q} \pi(c|q, a)p(q|a)u(c, a) \geq \sum_{c,q} \pi(c|q, a)p(q|\hat{a})u(c, \hat{a}), \ \forall \hat{a} \in A. \tag{15}$$

The left-hand side of (15) gives the utility the agent receives if he takes the recommended action. The right-hand side gives the utility the agent receives if he takes any other action. On the right-hand side, the deviating action $\hat{a}$ enters the utility function and affects the probability distribution of output. It does not affect the compensation schedule because the principal uses the recommended action in the compensation schedule.

To make these constraints linear in $\pi(c, q, a)$, we first define $\pi(c, q|a) = \pi(c|q, a)p(q|a)$ as the conditional probability of the consumption-output pair (c, q) given that action a is recommended. Next, we make the substitution $\pi(c|q, a) = (\pi(c, q|a))/(p(q|a))$ into (15) to obtain

$$\sum_{c,q} \pi(c, q|a)u(c, a) \geq \sum_{c,q} \pi(c, q|a)\frac{p(q|\hat{a})}{p(q|a)}u(c, \hat{a}), \ \forall \hat{a} \in A. \tag{16}$$

The term $\pi(c, q|a)\frac{p(q|\hat{a})}{p(q|a)}$ is the probability that the agent receives the pair (c, q) given that a was recommended but the agent instead takes action $\hat{a}$.

The final step in making the incentive constraints linear is to multiply both sides of (16) by the unconditional probability distribution $\pi(a)$ in order to express them in terms of $\pi(c, q, a)$. Rather than writing out the incentive constraints only for $\pi(a) > 0$, we write them as

$$\sum_{c,q} \pi(c, q, a)u(c, a) \geq \sum_{c,q} \pi(c, q, a)\frac{p(q|\hat{a})}{p(q|a)}u(c, \hat{a}), \forall a, (\hat{a} \neq a) \in A \times A. \tag{17}$$

This set of constraints applies not only to actions assigned positive probability but also those assigned zero probabiliy. Notice that the constraints hold trivially for actions such that $\pi(a) = 0$. This is quite convenient because normally we do not know beforehand which actions are assigned positive probability. Finally, the number of incentive constraints is finite because the number of elements in A is finite.

Probability Constraints

The last set of constraints ensures that $\pi(c, q, a)$ is a probability measure.

$$\sum_{c,q,a} \pi(c, q, a) = 1, \text{and } \forall c, q, a, \quad \pi(c, q, a) \geq 0. \tag{18}$$

Moral-Hazard Program with Lotteries

$$\max_{\pi} \sum_{c,q,a} \pi(c, q, a) w(q - c)$$

$$\text{s.t.} \sum_{c,q,a} \pi(c, q, a) u(c, a) \geq U, \tag{19}$$

$$(14), \ (17), \ \text{and} \ (18).$$

Equation (19) is the participation constraint. This program is a linear program because it has a finite number of linear constraints, a linear objective function, and a finite number of variables.

APPENDIX C RELATED LITERATURE

The moral-hazard problem is not the only private-information problem that may be formulated in terms of a linear program. Myerson (1982) contains a general environment that incorporates several types of private-information problems. Prescott and Townsend (1984), Townsend (1987a), and Townsend (1993) formulate several private-information models as linear programs. Below we list a number of private-information models and papers that set them up as linear programs.

Costly State Verification

In this model, the private information is the agent's income. The agent sends a report on his income which the principal may audit at a cost. This model has been heavily used in the auditing, macroeconomics, and finance literatures. Townsend (1988) formulates this problem as a linear program. He also shows

that the problem with the restriction that auditing be a deterministic function of the report is a mixed integer linear program. See also Boyd and Smith (1994).

Moral Hazard with a Public Input

This model is just like the moral-hazard problem except that there is also a publicly observed input into production. See Lehnert (1998) or Lehnert, Ligon, and Townsend (forthcoming 1999).

Hidden Information

In this model, a shock to income or preferences is hidden information. Several well-known models such as the Mirrlees (1971) optimal tax problem or the Mussa and Rosen (1978) monopolist problem can be worked into the linear programming formulation. Prescott and Townsend (1984) have formulated a related insurance problem as a linear program.

Hidden Information with Moral Hazard

This problem combines a shock to preferences followed by a moral-hazard problem. See Myerson (1982) or Prescott (1996).

Repeated Private Information

This problem repeats the static private-information problem over multiple periods. Phelan and Townsend (1991), Lehnert, Ligon, and Townsend (forthcoming 1999), Ligon, Thomas, and Worrall (forthcoming 1999), and Yeltekin (1998) analyze variations on the problem using a combination of linear programming methods and dynamic programming.

Limited Communication

Some problems with private information and limited communication have been formulated as linear programs. See Townsend (1987b), Townsend (1989), and Prescott (1996).

Multi-Agent Problems

Problems where multiple agents have private information can be formulated in this way. See, for example, Townsend (1993), Prescott and Townsend (1999), and Yeltekin (1998).

Limited Commitment

Lacker (1989) formulates a two-agent limited commitment with costly enforcement as a linear program. Ligon, Thomas, and Worrall (forthcoming 1999) also contain limited commitment features.

REFERENCES

Alvi, Eskander. "First-Order Approach to Principal-Agent Problems: A Generalization," *Geneva Papers on Risk and Insurance Theory,* vol. 22 (June 1997), pp. 59–65.

Arnott, Richard, and Joseph E. Stiglitz. "Randomization with Asymmetric Information," *RAND Journal of Economics*, vol. 19 (Autumn 1988), pp. 344–62.

Avery, Robert B., Patricia E. Beeson, and Paul S. Calem. "Using HMDA Data as a Regulatory Screen for Fair Lending Compliance," *Journal of Financial Services Research*, vol. 11 (April 1997), pp. 9–42.

Benston, George J., and George G. Kaufman. "Deposit Insurance Reform in the FDIC Improvement Act: The Experience to Date," Federal Reserve Bank of Chicago *Economic Perspectives*, vol. 22 (Second Quarter 1998), pp. 2–20.

Bertsimas, Dimitris, and John N. Tsitsiklis. *Introduction to Linear Optimization.* Belmont, Mass.: Athena Scientific, 1997.

Boyd, John H., and Bruce D. Smith. "How Good are Standard Debt Contracts? Stochastic versus Nonstochastic Monitoring in a Costly State Verification Environment," *Journal of Business*, vol. 67 (October 1994), pp. 539–61.

Calomeris, Charles W. "The IMF's Imprudent Role as Lender of Last Resort," *Cato Journal*, vol. 17 (Winter 1998), pp. 275–94.

Cole, Harold Linh. "Comment: General Competitive Analysis in an Economy with Private Information," *International Economic Review*, vol. 30 (February 1989), pp. 249–52.

_____, and Edward C. Prescott. "Valuation Equilibrium with Clubs," *Journal of Economic Theory*, vol. 74 (May 1997), pp. 19–39.

Dixit, Avinosh, and Timothy Besley. "James Mirrlees' Contributions to the Theory of Information and Incentives," *Scandinavian Journal of Economics*, vol. 99 (June 1997), pp. 207–35.

Dye, Ronald A. "Optimal Monitoring Policies in Agencies," *Rand Journal of Economics*, vol. 17 (Autumn 1986), pp. 339–50.

Feldman, Ron J., and Arthur J. Rolnick. "Fixing FDICIA: A Plan to Address the Too-Big-To-Fail Problem," Federal Reserve Bank of Minneapolis *The Region*, vol. 12 (March 1998), pp. 3–22.

Hansen, Gary D. "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, vol. 16 (November 1985), pp. 309–27.

Hart, Oliver. *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press, 1995.

_____, and Bengt Holmström. "The Theory of Contracts," in Truman F. Bewley, ed., *Advances in Economic Theory: Fifth World Congress.* Cambridge: Cambridge University Press, 1987.

Haubrich, Joseph G. "Risk Aversion, Performance Pay, and the Principal-Agent Problem," *Journal of Political Economy*, vol. 102 (April 1994), pp. 258–76.

Jensen, Michael C., and Kevin J. Murphy. "Performance Pay and Top-Management Incentives," *Journal of Political Economy*, vol. 98 (April 1990), pp. 225–64.

Jewitt, Ian. "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, vol. 56 (September 1988), pp. 1177–90.

Kareken, John H. "Deposit Insurance Reform or Deregulation Is the Cart, Not the Horse," Federal Reserve Bank of Minneapolis *Quarterly Review*, vol. 7 (Spring 1983), pp. 1–9.

Kupiec, Paul H., and James O'Brien. "A Pre-commitment Approach to Capital Requirements for Market Risk," Working Paper 95–36. Washington: Board of Governors of the Federal Reserve System, July 1995.

Lacker, Jeffrey M. "Limited Commitment and Costly Enforcement," Working Paper 90–2. Richmond: Federal Reserve Bank of Richmond, December 1989.

Lehnert, Andreas. "Asset Pooling, Credit Rationing, and Growth," Finance and Economics Discussion Series 1998–52. Washington: Board of Governors, December 1998.

_____, Ethan Ligon, and Robert M. Townsend. "Liquidity Constraints and Incentive Constraints," *Macroeconomic Dynamics* (forthcoming 1999).

Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall. "Mutual Insurance, Individual Savings and Limited Commitment," *Review of Economic Dynamics* (forthcoming 1999).

Luenberger, David G. *Introduction to Linear and Nonlinear Programming.* Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1973.

Marshall, David, and Subu Venkataraman. "Bank Capital Standards for Market Risk: A Welfare Analysis," *European Finance Review*, vol. 2 (1998), pp. 125–57.

Mirrlees, James A. "An Exploration in the Theory of Optimal Taxation," *Review of Economic Studies*, vol. 38 (July 1971), pp. 175–208.

Mussa, Michael, and Sherwin Rosen. "Monopoly and Product Quality," *Journal of Economic Theory*, vol. 18 (August 1978), pp. 301–17.

Myerson, Roger B. "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, vol. 10 (June 1982), pp. 67–81.

Phelan, Christopher, and Robert M. Townsend. "Computing Multi-Period, Information-Constrained Optima," *Review of Economic Studies*, vol. 58 (October 1991), pp. 853–81.

Prescott, Edward C., and Robert M. Townsend. "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard," *Econometrica*, vol. 52 (January 1984), pp. 21–54.

Prescott, Edward S. "Computing Moral-Hazard Problems Using the Dantzig-Wolfe Decomposition Algorithm," Working Paper 98–6. Richmond: Federal Reserve Bank of Richmond, June 1998.

_____. "The Pre-Commitment Approach in a Model of Regulatory Banking Capital," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 83 (Winter 1997), pp. 23–50.

_____. "Communication in Private Information Models: Theory and an Application to Agriculture." Manuscript. Richmond, December 1996.

_____, and Robert M. Townsend. "Boundaries and Connectedness of Economic Organizations." Manuscript. Richmond, February 1999.

Rogerson, Richard. "Indivisible Labor, Lotteries and Equilibrium," *Journal of Monetary Economics*, vol. 21 (January 1988), pp. 3–16.

Rogerson, William P. "The First-Order Approach to Principal-Agent Problems," *Econometrica*, vol. 53 (November 1985), pp. 1357–67.

Townsend, Robert M. *The Medieval Village Economy: A Study of the Pareto Mapping in General Equilibrium Models.* Princeton, N.J.: Princeton University Press, 1993.

_____. "Currency and Credit in a Private Information Economy," *Journal of Political Economy*, vol. 97 (December 1989), pp. 1323–44.

_____. "Information Constrained Insurance: The Revelation Principle Extended," *Journal of Monetary Economics*, vol. 21 (March/May 1988), pp. 411–50.

_____. "Arrow-Debreu Programs as Microfoundations of Macroeconomics," in Truman F. Bewley, ed., *Advances in Economic Theory: Fifth World Congress*. Cambridge: Cambridge University Press, 1987a.

_____. "Economic Organization with Limited Communication," *American Economic Review*, vol. 77 (December 1987b), pp. 954–71.

_____, and Rolf A. E. Mueller. "Mechanism Design and Village Economies: From Credit to Tenancy to Cropping Groups," *Review of Economic Dynamics*, vol. 1 (January 1998), pp. 119–72.

Yeltekin, Sevin. "Dynamic Principal-Multiple Agent Contracts." Manuscript. Stanford University, December 1998.

Gauging Manufacturing Activity: The Federal Reserve Bank of Richmond's Survey of Manufacturers

Robert L. Lacy

Midmorning on the second Tuesday of each month, the Federal Reserve Bank of Richmond posts the results of its latest survey of Fifth District manufacturers. The survey provides a comprehensive set of indicators of business conditions within the region's manufacturing sector. Survey participants share first-hand knowledge of recent changes in manufacturing activity at their companies and offer insights into expected developments six months ahead. Their compiled responses provide unique information on a broad range of manufacturing activities, including shipments, new orders, employment, and capacity utilization.

The survey of manufacturers is a valuable tool for Federal Reserve research staff responsible for monitoring the Fifth District economy.[1] It is also a source of information for analysts outside the Federal Reserve System seeking measures of the strength of manufacturing in the area. Interest in such regional economic data has grown rapidly in recent years, particularly among analysts searching regional data for clues to the future direction of the national economy. Financial press coverage of regional manufacturing and business condition polls has expanded as well. From time to time, the media cites the Federal Reserve

The author would like to thank Bruce Cox, Judy Higgins, Ray Owens, Ned Prescott, and Roy Webb for helpful comments. The views expressed in this article are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] The Fifth Federal Reserve District consists of the District of Columbia, Maryland, North Carolina, South Carolina, Virginia, and most of West Virginia.

Bank of Richmond's (Richmond Fed) survey in reports on economic activity or business developments.[2]

This article explains the techniques employed in gathering data and compiling results for the Richmond Fed's manufacturing survey. It also evaluates the survey's usefulness as a tool for economic analysis and compares survey indexes to aggregate manufacturing data and to indexes from similar surveys conducted by the Federal Reserve Bank of Philadelphia (Philadelphia Fed), the Federal Reserve Bank of Atlanta (Atlanta Fed), and the National Association of Purchasing Management (NAPM). The analysis indicates that the Richmond survey not only contributes to a better understanding of the District's manufacturing sector but may also provide timely indicators of changes in several closely watched national manufacturing data series.

1. PURPOSE AND HISTORY OF THE SURVEY

The Richmond Fed's survey of manufacturers was developed to gather timely and consistent data on manufacturing activity in the Fifth District. It is one of several research tools employed by the Bank's research staff to evaluate business conditions in the District and to collect the regional economic information needed by the Federal Reserve System to carry out effective monetary policy. Regional economic activity is tracked by each of the 12 regional Reserve Banks in the System, and their reports of changes in economic conditions around the country often receive considerable attention in monetary policy deliberations.[3]

The survey is the source of much of the manufacturing information presented by the Richmond Fed in its periodic reports on District economic conditions. These reports are prepared several weeks in advance of meetings of the Federal Open Market Committee (FOMC), which normally convenes eight times a year.[4] Each regional Reserve Bank produces such a report; these reports are compiled into a document informally referred to as the "beige book." The manufacturing survey is also a source of timely information for the development of policy recommendations made by the president of the Richmond Fed at FOMC meetings.

---

[2] The results of the Richmond Fed's manufacturing survey are reported every month by a number of prominent business news services specializing in providing economic and financial market information to their clients.

[3] References to regional economic information appear in the minutes of most Federal Open Market Committee (FOMC) meetings. The minutes from the November 17, 1998, meeting, for example, refer to anecdotal reports that "pointed to solid growth in most though not all regions of the country. . . ." Minutes of FOMC meetings are published in the *Federal Reserve Bulletin* and in annual reports of the Board of Governors of the Federal Reserve System.

[4] The voting members of the FOMC consist of the Federal Reserve's Board of Governors and five of the Reserve Bank presidents, one of whom is always the president of the Federal Reserve Bank of New York.

The survey of Fifth District manufacturers was initiated in June 1986. From 1986 to 1993, it was conducted every six or seven weeks, in a cycle directly linked to the preparation of the Federal Reserve's beige book.[5] Quantitative data from the survey enhanced the information contained in the District's beige book reports by supplementing anecdotal information acquired from manufacturing contacts through telephone conversations. Survey data also offered a timelier alternative to "official" manufacturing data released by government agencies or trade organizations. Most of these data series are available only after a lag of several months or more, thus limiting their usefulness in evaluating current economic conditions.

The survey of manufacturers took its present form in November 1993. It was converted to a monthly survey at that time and the results were made available to the general public. Its purpose, however, has remained the same over the years: to provide timely information on recent changes in manufacturing activity and changes in manufacturers' expectations about their business prospects six months ahead.

Manufacturing's share of total employment in the Fifth District, as in the nation, has declined in the 1990s as the service sector has expanded.[6] An understanding of the manufacturing sector, however, remains instrumental to grasping the dynamics of the overall economy. The sector is one of the more cyclical components of the economy, and movements in measures such as new orders and workweek are closely observed as leading indicators of economic performance. In addition, output in the manufacturing sector is easier to quantify than output in most services sectors, and changes in industry data are more readily interpreted. And, despite the steady advances of service industries in recent years, the manufacturing sector remains a sizeable component of the District economy.[7] About 1.9 million people are currently employed in manufacturing jobs in Fifth District states. These workers represent 15 percent of nonfarm employment in the region.[8]

[5] Chmura (1987/88) describes the origins of the manufacturing survey and provides charts of indexes of employment, capital expenditures, shipments, new orders, order backlogs, and inventories from 1986 to 1987.

[6] The Richmond Fed also produces a services-sector survey each month that gathers information on wholesale and retail trade, transportation, public utilities, finance, real estate, and business and health services, among other industries. The services-sector survey is similar to the manufacturing survey in size and methodology.

[7] These or similar arguments for surveying manufacturers appear in Bell and Crone's (1986) article on the Philadelphia Fed's manufacturing survey and in Rogers's (1992) article on the Atlanta Fed's survey. See Trebing (1998) for a more recent description and analysis of the Philadelphia Fed's survey.

[8] An overview of the Fifth District's manufacturing sector is provided in the Appendix.

2. MANUFACTURERS SURVEYED

Each participant in the Richmond Fed's survey is classified as a manufacturer under the Office of Management and Budget's 1987 Standard Industrial Classification (SIC) system and has production facilities located in the Fifth District. The District's manufacturing base is quite diverse and a wide variety of manufacturing firms respond to the survey. Of the 20 major manufacturing groups identified by the SIC manual, only petroleum refining and leather production are not currently represented in the survey.

The survey sample reflects both the geographic distribution of Fifth District manufacturers as well as their distribution across various types of manufacturing industries. The number of District employees in each of the 20 major manufacturing groups defined by the SIC system is a major factor in determining the mix of companies by industry included in the survey sample. Value added in manufacturing is a second consideration in the choice of industry mix. Fifth District employment and value added by industry are itemized in Table 1.[9]

As the data in Table 1 demonstrate, the two measures can yield very different results. The chemical industry (SIC 28), for example, represents only 7.8 percent of manufacturing employment in Fifth District states but contributes 19.3 percent to value added in manufacturing. Textiles, on the other hand, has an opposite distribution. The textile mill products industry (SIC 22) represents 16.6 percent of manufacturing employment in the District but only 8.3 percent of value added in manufacturing.

Table 1 also provides a distribution of survey responses by two-digit SIC code. Response percentages represent averages for the period January through April 1998; since survey participation is voluntary, percentages may vary considerably from month to month. The distribution of survey responses is within 2 percentage points of the distribution of employment for most industries. By the employment measure, primary metals manufacturers are currently the most overrepresented, since they account for 7.8 percent of survey responses but only 3.2 percent of Fifth District manufacturing employment. By the same measure, textiles and apparel manufacturers are the most underrepresented. Three industries (tobacco, textiles, and chemicals) show large differences in distributions depending on whether measured by employment or value added. In each of these cases, the survey distribution falls between the industry distribution as measured by employment and the industry distribution as measured by value added.

The survey sample also reflects the relative contribution of each state to the District's manufacturing sector. Table 2 provides a breakdown of survey

[9] The data in Table 1 reflect manufacturing employment for the entire state of West Virginia. Six counties in the northern panhandle region of the state, however, are not part of the Fifth District. Manufacturing employment in these counties represents approximately 20 percent of total West Virginia manufacturing employment.

Table 1 Distribution of Survey Responses, Employment, and Value Added by Industry

Manufacturing Industry	SIC Code	Survey Response Distribution (%)	Fifth District Employment Distribution (%)	Fifth District Value Added Distribution (%)	U.S. Employment Distribution (%)
Food & kindred products	20	5.7	8.0	8.7	9.1
Tobacco products	21	3.4	1.4	6.3	0.2
Textile mill products	22	12.8	16.6	8.3	3.3
Apparel/other fabric products	23	1.7	5.6	3.1	4.4
Lumber & wood products	24	4.1	5.6	2.9	4.3
Furniture & fixtures	25	4.1	5.8	2.8	2.7
Paper & allied products	26	6.4	4.0	4.7	3.7
Printing & publishing	27	5.7	6.7	5.0	8.3
Chemicals	28	8.4	7.8	19.3	5.5
Petroleum refining	29	0.0	N/A	0.1	0.8
Rubber & plastics	30	3.4	5.5	5.4	5.3
Leather	31	0.0	N/A	0.1	0.5
Stone, clay, & glass products	32	3.7	3.3	3.1	3.0
Primary metal industries	33	7.8	3.2	3.3	3.8
Fabricated metal products	34	5.4	5.1	4.1	7.9
Industrial machinery	35	9.8	6.7	6.6	11.6
Electronic equipment	36	10.5	7.6	7.7	9.0
Transportation equipment	37	4.7	4.8	5.2	9.9
Instruments	38	1.0	1.8	3.0	4.6
Miscellaneous manufacturing	39	1.4	0.6	0.5	2.1
TOTAL		100.0	100.0	100.0	100.0

N/A: Not available. District employment in each of these industries (SIC Codes 29 and 31) is generally less than 1 percent of manufacturing employment.

Notes: District employment and value added in manufacturing are calculated by summing state data. Two-digit SIC data are not reported by state for all manufacturing industries. Approximately 95 percent of total manufacturing employment and value added in Fifth District states are represented in the percentages shown in Table 1.

Sources: Federal Reserve Bank of Richmond, "Survey of Fifth District Manufacturing Activity," January–April 1998. Department of Labor, Bureau of Labor Statistics [1999]. Department of Commerce, "1996 Annual Survey of Manufactures," April 1998.

responses (January through April 1998) by state.[10] North Carolina, Virginia, and South Carolina account for about 85 percent of manufacturing employment

[10] Manufacturers in the District of Columbia are not surveyed. Manufacturing employment in Washington, D.C., represents less than 1 percent of total manufacturing employment in the Fifth District.

Table 2 Distribution of Survey Responses, Employment, and Value Added by State

	Survey Response Distribution (%)	Manufacturing Employment Distribution (%)	Manufacturing Value Added Distribution (%)
Maryland	13.5	9.5	9.9
Virginia	22.3	21.7	24.1
West Virginia	13.9	4.4	5.1
North Carolina	34.1	44.9	43.4
South Carolina	16.2	19.5	17.5
TOTAL	100.0	100.0	100.0

Sources: Federal Reserve Bank of Richmond, "Survey of Fifth District Manufacturing Activity," January–April 1998. Department of Labor, Bureau of Labor Statistics [1999]. Department of Commerce, "1996 Annual Survey of Manufactures," April 1998.

and value added in the Fifth District. The distributions in Table 2 suggest that North Carolina has been substantially underrepresented in survey responses while West Virginia has been overrepresented. Efforts have been undertaken recently to increase participation by North Carolina manufacturers.[11]

3. THE QUESTIONNAIRE AND DATA COLLECTED

Survey participants are asked to provide an assessment of changes in business conditions at their companies by answering two series of questions. The first series asks about changes in various measures of manufacturing activity compared to the previous month; the second series asks for changes between current activity and expected activity six months in the future. In each case, the respondent indicates only a direction of change: whether a particular activity has increased, decreased, or remained unchanged. This request for "categorical" rather than quantitative responses makes it easier to complete the questionnaire and helps keep survey response rates high.

Manufacturers are questioned about new orders, order backlogs, shipments, capacity utilization, vendor lead time, number of employees, average workweek, and wages. Responses to these questions yield insight into product demand, product flows in the manufacturing process, and use of resources. In addition, employment, workweek, and wage data contribute to a better understanding of current and future labor market conditions.

[11] Survey participation tends to decline over time and thus new participants must be recruited periodically in order to maintain an adequate sample size and mix of manufacturers. The latest recruitment of companies took place in February 1999.

The questionnaire also inquires about inventory levels and price changes. Respondents provide information regarding inventories of both raw materials and finished goods. They are asked whether inventory levels, when compared to a desired inventory level, are too high, too low, or correct. In the prices section, the questionnaire asks for an estimate of the percent change (on an annualized basis) of prices paid for raw materials and prices received for finished goods.

Survey questionnaires are mailed around the twelfth of each month and are typically addressed to plant managers or controllers—individuals with detailed knowledge of business activity at each facility. Because most of the completed questionnaires are returned within a week of receipt, their data reflect business conditions as of the middle of the month. Responses are accepted, however, up until the week before final survey results are released on the second Tuesday of the subsequent month. Typically, 60 to 75 of the approximately 175 firms that received a questionnaire each month in 1998 provided responses.

4. SURVEY RESULTS

The results of the manufacturing survey are generally expressed as diffusion indexes.[12] A separate diffusion index is determined for each question asked and is calculated as

$$\text{Index} = 100(\text{I} - \text{D})/(\text{I}+\text{N}+\text{D}),$$

where

$$\begin{aligned} \text{I} &= \text{number of respondents reporting increases,} \\ \text{N} &= \text{number of respondents reporting no change, and} \\ \text{D} &= \text{number of respondents reporting decreases.} \end{aligned}$$

Each diffusion index can range in value from minus 100, if all respondents reported decreased activity, to 100, if all respondents reported increased activity. While diffusion indexes can be scaled in other ways, this particular methodology has the appeal of assigning positive values when more respondents are reporting increased activity than decreased activity.[13]

A diffusion index is a measure of the scope of change in an activity across the firms participating in the survey.[14] A large diffusion index for employment,

[12] Changes in prices are reported as annualized percent changes rather than as diffusion indexes.

[13] In addition to the equation above, a diffusion index can be calculated as Index = 100(I+0.5N)/(I+N+D) where I = the number of respondents reporting increases, N = the number of respondents reporting no change, and D = the number of respondents reporting decreases. This approach results in an index range of 0 to 100 and a value of 50 if the number of respondents reporting increases equals the number reporting decreases.

[14] Although diffusion indexes are often associated today with the analysis of survey data, they have broader applicability. The concept of measuring the "diffusion" of economic change originated as a tool for forecasting national economic trends. Researchers at the National Bureau

for example, means that higher employment is much more widely reported than lower employment. The index does not directly measure the magnitude of changes in the levels of manufacturing activity. Survey participants report only that a particular activity has increased, decreased, or remained unchanged; they do not report the magnitude of any changes.[15]

After initial diffusion indexes are calculated from survey responses, seasonal adjustments are made. Seasonal patterns appear in responses to most of the questions. Such patterns are not unusual in manufacturing data since many goods producers have traditional periods of lower production or shutdown of operations as they retool or make other major adjustments to their production processes.

Seasonal adjustments are made using the Department of Commerce's X-11 methodology.[16] The X-11 methodology breaks time-series data into trend-cycle, seasonal, and irregular components. These components are extracted by means of a series of moving-average filters. Seasonally unadjusted diffusion indexes are reported only when historic data series are not long enough to allow adequate seasonal adjustment. At present, unadjusted series include current wages and capacity utilization as well as many of the indexes of manufacturing activity six months ahead. Most of these series begin in 1996 or 1997 and will not be seasonally adjusted until four years of historic data are available.

Final survey results are released to the public at 10:00 a.m. on the second Tuesday of each month. A summary of recent trends is provided along with three months of index data and a three-month moving average of index data. Table 3 provides the manufacturing indexes and price trends reported in a recent release of survey results.

5. INTERPRETATION OF SURVEY RESULTS

Manufacturing survey indexes represent a compilation of the perceptions of a representative sample of District manufacturers regarding current and future business activity. They are unique data that reveal a great deal about manufacturing activities at firms participating in the survey. The indexes can also provide insight into the strength of the overall manufacturing sector in the

of Economic Research calculated diffusion indexes as early as 1950 to measure the extent of change in component series of leading and coincident U.S. economic indicators.

[15] In addition, the diffusion index is a summary statistic; information about the number of increases, decreases, and unchanged responses is not conveyed. There may be times when it would be useful to know, for example, whether a diffusion index of zero meant that all respondents reported "no change"; that a large number of respondents reporting "increase" had been offset by an equally large number of respondents reporting "decrease"; or, perhaps most likely, that most respondents reported no change while a fairly small number of reported increases were offset by an equal number of reported decreases.

[16] See Shiskin, Young, and Musgrave (1967) for a detailed description of the X-11 technique.

Table 3 Survey of Fifth District Manufacturing Activity (Information Reported January 12, 1999)

	December 1998	November 1998	October 1998	3-Month Average
Business Activity Indexes				
Compared to Previous Month				
Shipments	4	8	−4	3
New orders	−7	0	4	−1
Backlog of orders	−9	−6	−5	−7
Capacity utilization*	−16	−7	6	−6
Vendor lead time*	3	5	0	3
Number of employees	5	−14	5	−1
Average workweek	−2	1	−7	−3
Wages*	3	9	11	8
Six Months from Now				
Shipments	27	31	18	25
New orders*	33	19	13	22
Backlog of orders*	16	18	−3	10
Capacity utilization*	31	15	9	18
Vendor lead time*	7	2	−6	1
Number of employees	12	−7	−2	1
Average workweek	6	5	−4	2
Wages*	42	40	42	41
Capital expenditures*	13	18	22	18
Inventory Levels				
Finished goods inventories*	25	31	36	31
Raw materials inventories*	19	18	29	22
Prices (Percent Change, Annualized)				
Current				
Prices paid	0.76	0.49	0.60	
Prices received	0.03	−0.04	−0.74	
Expected (Next Six Months)				
Prices paid	0.84	0.74	0.78	
Prices received	0.27	0.45	0.07	

*Indicators not seasonally adjusted due to insufficient historical data.

region. However, care must be taken in interpreting the survey and extending survey results to the District or nation as a whole.

The first and perhaps most obvious caution is that survey indexes can be volatile from month to month, making them difficult to interpret. Some of this volatility is undoubtedly due to the survey's relatively small size; a larger number of responses would likely decrease monthly fluctuations. Furthermore, since the survey is voluntary, the group of respondents changes somewhat

from month to month, which may also contribute to volatility. The three-month moving average of data reported with survey results smooths the series and allows longer-term trends to be more readily identified.

A second caveat relates more broadly to survey sampling. While the companies participating in the survey are selected to be representative of Fifth District manufacturers, they are not selected based on statistical criteria for sampling. Therefore, one cannot conclude with any specific level of confidence that a change in an index number necessarily implies a change in the corresponding activity for the manufacturing sector in Fifth District states.

Statistical analysis indicates, however, that trends in key survey indexes are generally consistent with trends in manufacturing data series available from other sources. In particular, key indexes from the Richmond Fed's survey are positively correlated with comparable indexes from the manufacturing surveys of the Philadelphia Fed, the Atlanta Fed, and NAPM. Furthermore, a number of indexes from the Richmond Fed survey are positively correlated with changes in comparable official data series, suggesting that the Richmond Fed survey may be of value as a timely indicator of changes in the Fifth District and national economies.

6. COMPARISON TO OTHER MANUFACTURING SURVEYS

Figures 1–3 show selected Richmond Fed survey diffusion indexes plotted with comparable diffusion indexes from NAPM's "Report on Business" survey. The NAPM survey is widely regarded as one of the most reliable indicators of manufacturing activity in the country.[17] With the exception of a four-year break during World War II, the "Report on Business" has been published every month since 1931. NAPM's survey covers the entire country and receives considerable attention from the financial press when it is released on the first business day of each month.

Diffusion indexes in the NAPM survey are calculated by adding the percent of respondents reporting an increase to half the percent of respondents reporting no change. An index reading above 50 indicates an activity is expanding. An index of 50 on NAPM's scale, therefore, is comparable to an index of zero on the Richmond Fed's scale. Employment, new orders, and shipments/production indexes from both surveys for the period November 1993 to December 1998 are graphed.[18]

Richmond Fed survey indexes can also be compared to the manufacturing surveys conducted by two other Federal Reserve banks. The Philadelphia and

[17] See, for example, Klein and Moore (1988).

[18] The NAPM production index is used as a proxy for shipments in Figure 3.

Figure 1 Manufacturing Employment
Richmond Fed and NAPM

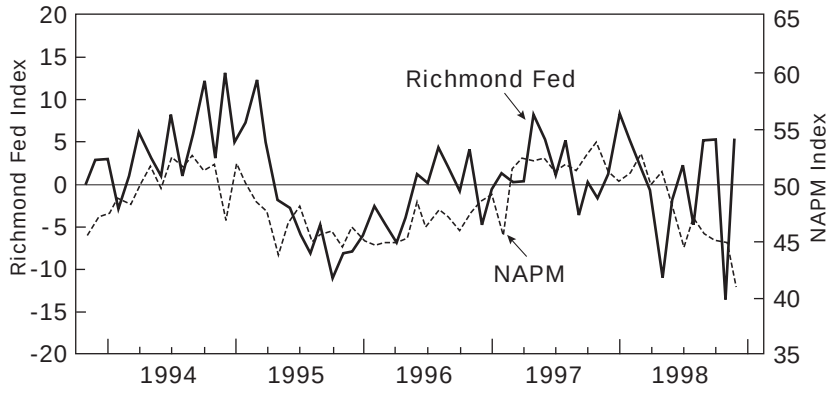
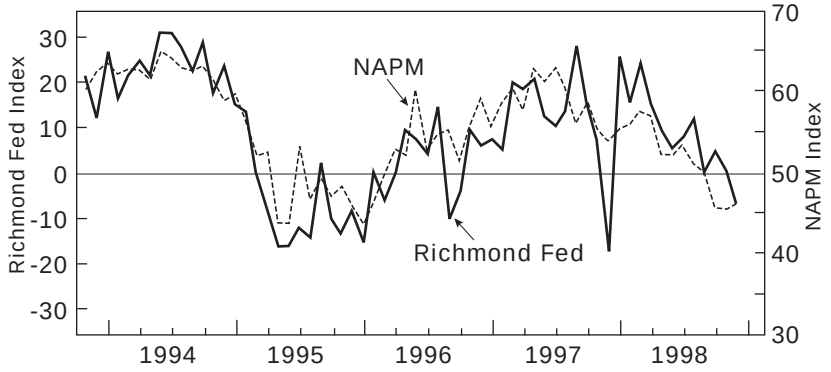


Figure 2 New Orders
Richmond Fed and NAPM



Atlanta Feds produce their surveys of District manufacturers on a monthly basis and calculate their diffusion indexes the same way that Richmond does.[19] The Philadelphia survey has been conducted since 1968, while the Atlanta survey

[19] The Federal Reserve Bank of Kansas City also conducts a manufacturing survey, but it is administered on a quarterly rather than monthly basis.

Figure 3 Shipments/Production
Richmond Fed and NAPM

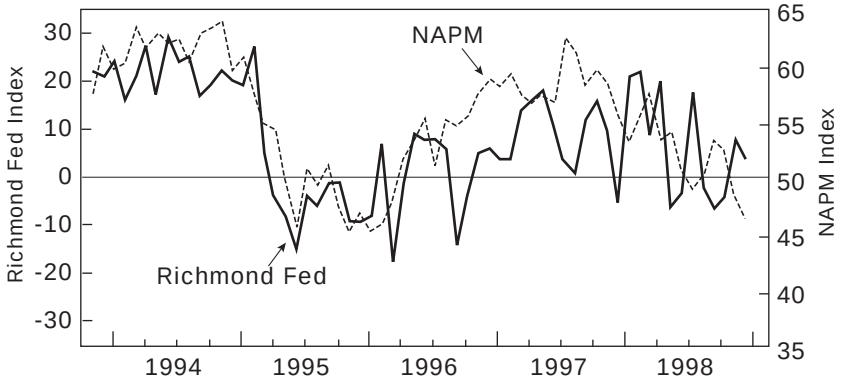
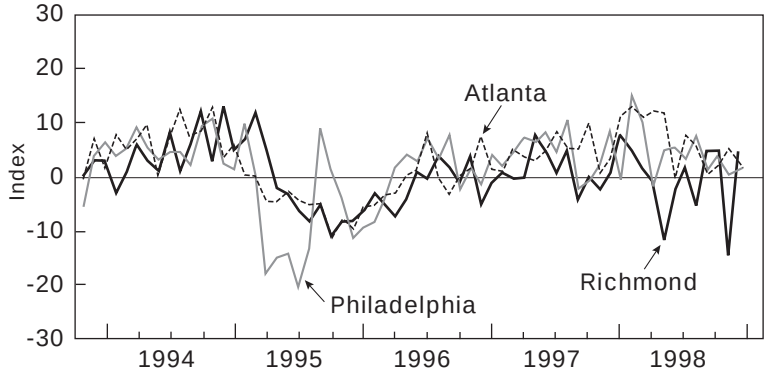


Figure 4 Manufacturing Employment
Federal Reserve Surveys



was initiated in 1991. Philadelphia releases its results on the third Thursday of the month, Atlanta on the second business day after the tenth of the month. Graphs of selected diffusion indexes from the Richmond, Philadelphia, and Atlanta manufacturing surveys are provided in Figures 4–6. Indexes for employment, new orders, and shipments are plotted.

Table 4 provides correlation statistics from a comparison of Richmond

Figure 5 New Orders
Federal Reserve Surveys

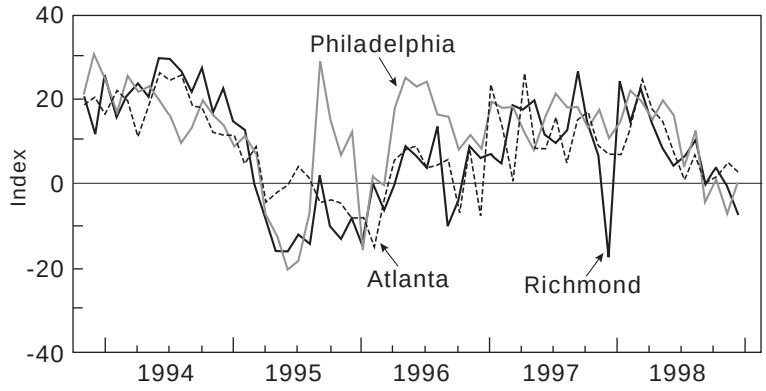
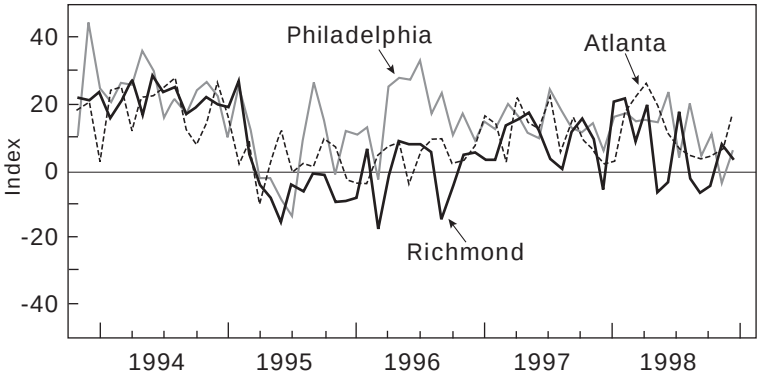


Figure 6 Shipments
Federal Reserve Surveys



survey indexes to selected indexes from NAPM and the Philadelphia and Atlanta Feds based on the study period of November 1993 to December 1998. The data in Figures 1–6 and Table 4 indicate that the Richmond Fed's survey indexes for these three major measures of manufacturing activity tend to track the indexes of the NAPM survey and those of other Federal Reserve surveys during most of the study period.

Table 4 Correlations between Richmond Fed Survey and Other Manufacturing Surveys (Major Indexes)

	Employment Index	New Orders Index	Shipments/ Production Index
Richmond Fed survey Compared to			
NAPM survey	0.358	0.812	0.674
Philadelphia Fed survey	0.369	0.628	0.507
Atlanta Fed survey	0.376	0.727	0.513

Notes: Numbers represent Pearson correlation coefficients. All correlations are significant at a 95 percent confidence level. NAPM's production index is used as a proxy for shipments. Indexes are seasonally adjusted.

7. DISTRICT AND NATIONAL ECONOMIC MEASURES

While a comparison of Richmond Fed survey indexes to comparable indexes from other manufacturing surveys provides useful information, a more direct test of the survey as a tool for gauging manufacturing activity is to compare survey indexes to "official" manufacturing statistics from government agencies or industry sources.[20] Directly comparable data in the form of diffusion indexes are not available, but month-to-month changes in some aggregate data series can be used as a proxy for the purpose of comparison to survey indexes. Since survey questions ask about changes in business activity, survey indexes are more directly related to changes in official data series than absolute levels of such data.

Efforts to "benchmark" Richmond manufacturing survey indexes against changes in official monthly statistics at a District level are limited by the paucity of regional manufacturing data. Manufacturing employment and workweek data, however, are available by state from the Labor Department's Bureau of Labor Statistics (BLS) and were aggregated to develop Fifth District totals. Month-to-month percent changes in these District totals were then compared to diffusion indexes from the manufacturing survey. Correlation coefficients based on a three-month moving average of data for the period November 1993 through December 1998 are included in Table 5. Employment data are compared on a seasonally adjusted and nonadjusted basis. While correlation coefficients are positive for both employment and workweek, they are significant only for the employment measures.

[20] The approach taken in comparing survey indexes to official statistics follows, in general, that taken by Bell and Crone (1986).

Table 5 Richmond Fed Manufacturing Survey Correlations of Survey Indexes with Changes in District Aggregate Data

Employment Index (SA)	0.639*
Employment Index (NSA)	0.613*
Workweek (NSA)	0.131

*Variables significant at a 95 percent confidence level.

SA: Seasonally adjusted.

NSA: Not seasonally adjusted.

Notes: Numbers represent Pearson correlation coefficients. Comparisons are based on a three-month moving average of variables.

Source: Department of Labor, Bureau of Labor Statistics: employment in manufacturing, average workweek.

Further analysis of survey data suggests that several of the indexes may also provide a timely indication of changes in national manufacturing activity. While the survey is designed to collect information solely from Fifth District manufacturers, District manufacturing activity often tracks national activity, since the economic and business conditions that prevail in the region often prevail throughout the country. Moreover, Fifth District states account for a considerable portion, about 10 percent, of total national manufacturing employment. As shown in Table 1, the distribution of manufacturing employment across industries in the Fifth District, with the exception of textiles, is similar to the distribution of manufacturing employment in the United States.

Relationships between survey index data and national manufacturing statistics were explored by comparing survey indexes for employment, shipments, new orders, workweek, and prices paid to changes in comparable aggregate measures of national manufacturing activity from the Labor Department and Department of Commerce. Correlation coefficients based on a three-month moving average of seasonally adjusted data are included in Table 6. NAPM survey indexes for employment, production, and new orders are also compared to changes in aggregate measures of national manufacturing activity. Since the NAPM surveys companies nationwide, its indexes are more directly comparable to national economic measures. Correlation coefficients from a comparison of NAPM indexes to changes in aggregate measures of manufacturing activity provide a rough benchmark of the degree of correlation that might reasonably be expected in the Richmond survey.

The employment indexes in both surveys are highly correlated with official employment data. The Richmond Fed's survey indexes for shipments, new orders, and prices paid are also correlated with comparable national

Table 6 Correlations of Survey Indexes with Changes in National Aggregate Data

Index	Richmond Fed Survey	NAPM Survey
Employment	0.552*	0.623*
Shipments	0.432*	0.486*
New orders	0.338*	0.403*
Prices paid	0.851*	–
Workweek	0.062	–

*Variables significant at a 95 percent confidence level.

Notes: Numbers represent Pearson correlation coefficients. Comparisons are based on a three-month moving average of variables. NAPM's production index is used as a proxy for shipments.

Sources: Department of Labor, Bureau of Labor Statistics: employment in manufacturing, average workweek; PPI for commodities, materials and components for manufacturing; Department of Commerce: manufacturers' shipments and orders.

manufacturing statistics.[21] As was the case in the comparison to District data, the workweek index correlation statistic is not significant.

Do the survey's "expectation" indexes reliably foretell changes in District or national manufacturing activity? The results to date are mixed. Table 7 contains correlation statistics from comparisons of various expectation indexes to corresponding changes in actual manufacturing data six months later.[22] The workweek and prices-paid variables show significant positive correlations with "official" data, while the employment and shipments indexes do not. The relatively high correlation of the prices-paid index is consistent with the strong correlation shown for prices paid in Table 6. Prices are reported on an annualized percentage-change basis; the index is an average of the percent changes reported and thus not a diffusion index. The fact that more information is captured in the prices-paid index may explain part of the high correlations for this index. The high correlations for expected workweek are in contrast to the low correlation statistics for workweek reported in Tables 5 and 6.

[21] Balke and Petersen (1998) explore a broader but related issue. They examine how well beige book descriptions of economic activity by Federal Reserve Banks match national economic activity as measured by real GDP growth. The authors find that their quantified measures of beige book descriptions track current GDP growth "quite well." The Richmond Fed was one of several Federal Reserve Banks whose regional description, when quantified, was statistically significant in predicting current-quarter real GDP growth.

[22] Questions regarding employment, shipments, workweek, and prices paid have appeared in the survey since November 1993. Most of the other "six-month-ahead" variables were not added until May 1997. Thus, sufficient data do not exist to evaluate these more recent indexes.

Table 7 Richmond Fed Manufacturing Survey Correlations of Expectation Indexes with Changes in Regional and National Data

Index	District Comparison	National Comparison
Employment	−0.041	−0.210
Shipments	–	−0.033
Workweek	0.546*	0.386*
Prices paid	–	0.557*

*Variables significant at a 95 percent confidence level.

Notes: Numbers represent Pearson correlation coefficients. Comparisons are based on a three-month moving average of data. All results based on seasonally adjusted data, except for workweek comparison at the District level.

8. CONCLUSIONS

The Federal Reserve Bank of Richmond's manufacturing survey provides a practical set of indexes for tracking changes in production activities within a representative group of Fifth District manufacturers. These indexes are timely and comprehensive, covering all major facets of operations, from new orders received to the volume and prices of products shipped. Collectively, they paint a composite picture of manufacturing activity in the region. Because the survey has enabled the systematic compilation of such data as a complement to anecdotal manufacturing information gathered from other sources, it has proven to be a valuable tool for regional economic analysis.

While periodic surveys of Fifth District manufacturers have been conducted since the mid-1980s, the monthly manufacturing survey in its current form dates back only to November 1993. It is thus just beginning to develop a track record long enough to allow statistical testing and comparison to other manufacturing data series. The analysis conducted to date indicates that several major indexes, including employment, shipments, and new orders, are consistent with similar indexes from other surveys and with manufacturing data available from government or industry sources. A longer study period, however, is needed before one can draw firm conclusions regarding how well the Richmond survey tracks other data series.[23] Survey performance over periods of both economic expansion and contraction, in particular, needs to be evaluated. The U.S. economy

[23] A longer study period would allow for improved statistical analysis of the survey indexes included in Tables 4–7. The survey indexes that have thus far not been subject to any statistical analysis because of limited historic data could also begin to be evaluated. These indexes include wages, vendor lead time, and capacity utilization as well as many of the expectation indexes.

overall has not contracted during any quarter since the initiation of the monthly manufacturing survey in late 1993.

The passage of time will allow the collection of additional data and further statistical analysis. The results of such analysis will provide additional information regarding how well survey indexes track other measures of manufacturing activity and, perhaps, how much confidence can be placed in survey results. But, to a large extent, the manufacturing survey indexes stand on their own. They are unique diffusion indexes of Fifth District manufacturing activity; no directly comparable data exist.

APPENDIX : MANUFACTURING IN FIFTH DISTRICT STATES

The lion's share of manufacturing output in the Fifth District—about 85 percent—comes from facilities located in North Carolina, South Carolina, and Virginia. This area has a long, rich manufacturing history dating back to the earliest years of English settlement in North America. Two industries, cotton textiles and tobacco, have anchored the manufacturing base in the region for over a hundred years; while they no longer dominate the region's manufacturing sector, they remain vital industries. Today the manufacturing sector in the Carolinas and Virginia is a diverse mix of many types of firms, large and small, traditional and cutting edge.

North Carolina is the largest manufacturing state in the Fifth District and the eighth largest industrial state in the country. Over 800,000 people are employed in the manufacturing sector in North Carolina; twice as many as any other state in the District. Its preeminence as the leading industrialized state in the South was established early in the century as its textiles, tobacco, and furniture industries flourished.

More than one-quarter of manufacturing workers in North Carolina are employed in the textile and apparel industries. Output from these industries represents approximately 16 percent of the state's manufacturing product. North Carolina has been a leading textile-producing state in the country since the 1920s, when the center of the cotton textile industry in the nation began shifting from New England to the South. The textile industry remains by far the largest manufacturing employer in the state.

Tobacco has roots that extend even deeper into North Carolina's past. By the 1880s, rapidly expanding production of a new strain of milder, "bright leaf" tobacco in North Carolina boosted the state's tobacco fortunes and allowed North Carolina to begin to rival Virginia in tobacco production. Approximately

40 percent of value added in manufacturing U.S. tobacco products today comes from North Carolina manufacturers. The tobacco industry represents 13 percent of the state's manufacturing output.

Furniture making is the second-largest manufacturing industry in terms of employment in North Carolina. The state has been a leading producer of wood household furniture since the 1930s. The success of the furniture industry has also stimulated related industries, including logging and lumber operations. Employment in the lumber and furniture-making industries makes up about 14 percent of manufacturing employment and 8 percent of manufacturing output in North Carolina.

Textiles and apparel manufacturing, tobacco production, and furniture making collectively represent one-third of North Carolina's manufacturing output. Other large sectors include chemicals, electronic equipment manufacturing, and industrial machinery. The state's textiles and apparel manufacturers, as well as its tobacco producers, have seen their shares of state manufacturing output decline in the 1990s, while the chemical industry has grown.

South Carolina's manufacturing sector is less than half as large as North Carolina's. Over one-quarter of manufacturing employment (17 percent of output) is in the textile and apparel industries. Chemical industries contribute the largest share of manufacturing output, about 21 percent, and 10 percent of the sector's employment. Paper production and industrial machinery manufacturing are also primary industries. Manufacturing output has grown faster in South Carolina than in any other state in the District since 1990, in part because of expanding automobile manufacturing.

Virginia is the second-largest industrial state in the District with a little over 400,000 manufacturing employees. Manufacturing in Virginia has traditionally been more diverse than in North and South Carolina; it has not been as heavily concentrated in textile manufacturing as states further south. The tobacco industry remains Virginia's largest manufacturing industry as measured by value of product: 15 percent of gross state product in manufacturing comes from tobacco. The textiles and apparel industries are also major industries in the state, employing approximately 47,000 people. The chemical industry produces 12 percent of manufacturing output, while 10 percent of employment and output is in food production. Automobile manufacturing and shipbuilding facilities in the state give it a larger presence in transportation industries than any other Fifth District state.

West Virginia and Maryland contribute about 15 percent to the District's manufacturing output. Chemical and primary metals firms dominate West Virginia's manufacturing sector. These two industries account for one-third of manufacturing employment and over 60 percent of manufacturing output. Substantial lumber operations also exist in the state. Maryland's manufacturing sector is larger than West Virginia's and much more diverse. Food

production is another major industry; large poultry operations thrive on Maryland's Eastern Shore. The state also has substantial printing and publishing, industrial machinery, instrument, and chemical industries.

REFERENCES

Balke, Nathan S., and D'Ann Petersen. "How Well Does the Beige Book Reflect Economic Activity? Evaluating Qualitative Information Quantitatively," Working Paper 98–02. Dallas: Federal Reserve Bank of Dallas, June 1998.

Bell, John, and Theodore Crone. "Charting the Course of the Economy: What Can Local Manufacturers Tell Us?" Federal Reserve Bank of Philadelphia *Business Review* (July/August 1986), pp. 3–16.

Board of Governors of the Federal Reserve System. *Annual Reports,* 1993–1997. Washington: Board of Governors.

Chmura, Christine. "New Survey Monitors District Manufacturing Activity," Federal Reserve Bank of Richmond *Cross Sections,* vol. 4 (Winter 1987/88), pp. 9–11.

Federal Reserve Bank of Richmond. "Survey of Fifth District Manufacturing Activity," December 1993–January 1999.

Klein, Philip A., and Geoffrey H. Moore. "N.A.P.M. Business Survey Data: Their Value as Leading Indicators," *Journal of Purchasing and Materials Management* (Winter 1988), pp. 32–40.

National Association of Purchasing Management. "Report on Business," November 1993–January 1999.

Rogers, R. Mark. "Tracking Manufacturing: The Survey of Southeastern Manufacturing Conditions," Federal Reserve Bank of Atlanta *Economic Review,* vol. 77 (September/October 1992), pp. 26–33.

Shiskin, Julius, Allan H. Young, and John C. Musgrave. "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper No. 15. Washington: U.S. Department of Commerce, Bureau of the Census, February 1967.

Trebing, Michael E. "What's Happening in Manufacturing: 'Survey Says . . . ,' " Federal Reserve Bank of Philadelphia *Business Review* (September/October 1998), pp. 15–29.

United States Department of Commerce. "1996 Annual Survey of Manufactures," April 1998.

United States Department of Labor, Bureau of Labor Statistics. Employment, Hours, and Earnings for 1997. Available: http://stats.bls.gov/sahome.html [February 1999].