Central Banking: Then and Now

J. Alfred Broaddus, Jr.

Thank you very much for that kind introduction. It is very nice indeed to be here in the Shenandoah Valley. Like Woodrow Wilson, I'm a native son of Virginia. Unlike him, I can't claim to have begun life in this magnificent Valley. But I did have the pleasure of completing my undergraduate studies not too far from here at Washington and Lee. So I feel very much at home here, and I appreciate your invitation to be with you.

The theme of this conference is "Facing Economic Issues: Clinton and Wilson." And this is a quite appropriate theme, because there are obvious parallels. President Clinton and the country face pressing economic problems today, many of which have been discussed by previous speakers at this conference. President Wilson also faced substantial economic challenges in his Administrations. One of President Wilson's greatest achievements—which occurred in his first year in office—was his orchestration of the difficult compromise, among a number of powerful and conflicting groups in the country, that culminated in passage of the Federal Reserve Act in December 1913 and the creation of our central bank, including its regional arms, the Federal Reserve Banks, the following year.

Against this background, what I would like to do this morning is to tell you a story: an historical story, if I may, which seems appropriate in this setting. It is the story of the Federal Reserve, inflation, deflation and the relationship

This article is adapted from an address given by J. Alfred Broaddus, Jr., president of the Federal Reserve Bank of Richmond, to the Woodrow Wilson Forum sponsored by the Woodrow Wilson Birthplace Foundation, Staunton, Virginia, on April 2, 1993. Marvin Goodfriend, senior vice president and director of research at the Richmond Fed, contributed substantially to the preparation of the article.

between the three. Probably everyone here knows that the Fed is supposed to maintain the purchasing power of the American dollar and to prevent inflation. It is also supposed to prevent *de*flation, which is not much on people's minds today, but was at times in Wilson's day and certainly in the 1930s. Another way of saying this is that the Fed is supposed to keep the aggregate level of prices—not the individual prices of particular goods and services, but the *aggregate* price level—reasonably stable over time. A stable price level by definition implies the absence of both persistent inflation and persistent deflation.

That the Fed is in some sense responsible for stabilizing the price level presupposes some benefit from doing so. As many of you know, there has been far less than complete agreement in the United States, both in the distant past and more recently, on the desirability of price-level stability—particularly the desirability of controlling inflation. Some people, especially those who borrow money regularly, benefit from inflation, at least temporarily and partially. But I think it's fair to say that a majority of Americans value a stable price level and a sound dollar, even if they don't think about it a lot. In general they don't want the frequently high and typically variable inflation rates that have plagued so many other countries in the past and now. Americans sense that stable prices and stable money prevent the arbitrary redistributions of real income and wealth that accompany inflation and weaken societies. They sense also that stable prices and stable money eliminate the confusion, uncertainty, risk and inefficiency that inflation introduces into the nation's free market system.

Now while most Americans believe that the Fed is supposed to "fight" inflation and deflation in some general sense, they are also aware that the fight has been an uneven one and by no means fully successful in all periods of our history. On the contrary, the country went through a cataclysmic *de*flation in the 1930s. Subsequently it went through a substantial *in*flation in the late 1970s and early 1980s—not as traumatic and damaging as the experience in the '30s, but a very bad time nonetheless.

What's the problem? Why hasn't the Fed done a better job? I am going to argue today that one reason—and maybe the *main* reason—is that the Fed does not now have, and it never has had, a clear congressional mandate to stabilize the price level. Consequently, the Fed's success in stabilizing the price level in at least some periods of its history has been and continues to be a function largely of (1) prevailing general economic conditions, (2) the strength of the Federal Reserve's leaders, and (3) old-fashioned luck. The implication, of course, is that something probably should be done to strengthen the Fed's hand so that its performance would be less dependent on fortuitous circumstances. And let me make it clear that I personally feel strongly that something should be done. I am well aware that in today's relatively low inflation climate, many people do not see this as a pressing issue, such as the federal budget deficit or health care reform, that requires immediate attention. I disagree for reasons I hope to make clear in the remainder of my comments.

There are probably several ways one could make this argument, but, as I suggested earlier, I want to take an historical approach, which seems appropriate here. I'll proceed as follows. First, I want briefly to describe the monetary conditions that led to the creation of the Fed—with the assistance of Woodrow Wilson. Then I will try to indicate exactly what the framers of the Federal Reserve Act expected the Fed to do—its mandate in 1913 and 1914, as it were. A particular point here is that the mandate did not include, in any explicit way, stabilizing the price level. Next, I'll indicate how the rapid and substantial change in circumstances during and after World War I, shortly after the Fed was created, forced the Fed to confront the problem of stabilizing the price level early on in the 1920s, a challenge it met with some success until the stock market crash in 1929. All of this is the "then" part of my remarks. Finally, I'll skip over to the late 1970s and 1980s—the "now" part—and show that the Federal Reserve has faced many of the same inflation challenges in recent years that it faced in the 1920s. It has achieved some success in this later period also, for remarkably similar reasons. However, the absence of a clear price-level stability mandate today presents the nation with some—not all, but some—of the same kinds of risks it faced in its early years. We are much better equipped to deal with these risks now than we were then. But we can and should reduce them by clarifying the Fed's price stability mandate, preferably through legislation.

That's my introduction, and it has been a long one. But let me proceed, and I will try to make the remainder of my points as compactly as I can.

1. THE GOLD STANDARD AND PRICE STABILITY BEFORE THE FEDERAL RESERVE

As I suggested a minute ago, the Federal Reserve was established in 1914 to remedy banking and currency problems that had been recurring since the Civil War. The country had no central bank during this period, which is known to economic historians as the National Banking Era. The United States left the gold standard to help finance the Civil War, but returned to it in 1879. Thereafter, monetary conditions were largely governed by the flow of gold to and from the United States as part of the international balance of payments adjustment mechanism under the international gold standard.

Under the gold standard, the national money supply was closely linked to the nation's stock of monetary gold, which included gold coin, Treasury currency backed by gold, and gold reserves held by banks. When the country ran a balance of trade surplus, for example, the excess of foreign receipts over expenditures was received in gold. The gold inflow set in train a multiple expansion of deposits that increased the money supply. The increase in the money supply then increased domestic demand for goods and services and put upward pressure on domestic prices. The reverse occurred when the country ran a trade deficit. For our purposes, the point is that under a gold standard without a central bank, the nation's stock of money was automatically regulated by conditions in world markets.

This system had good features and not-so-good features. On the good side, the gold standard did keep the aggregate price level under control over the very long run. The aggregate level of prices in 1914, for example, was not very different from the level 30 years before in the early 1880s. By comparison, the price level rose 270 percent between 1960 and 1985. So the gold standard provided an anchor for the price level over the long run—that is, it provided a means of stabilizing the price level over the long run. Moreover, it was a *credible* anchor; the public understood the mechanism and knew it worked.

But the gold standard had significant limitations in the short and intermediate terms. First, while the gold standard anchored the price level over the very long run, it nonetheless allowed it to drift upward and downward by significant amounts over fairly long periods. For example, slow growth in the world gold supply caused the price level to decline at over 1 percent per year from 1879 to 1897, which provoked William Jennings Bryan's famous plea not to crucify mankind on a cross of gold. Subsequently, new gold discoveries and improved mining techniques caused the metal's supply to *increase* rapidly in the late 1890s and early 1900s. Consequently, the price level *rose* at over 2 percent per year from about 1897 to 1914. A second limitation was that the strict discipline of the gold standard did not allow the money supply to increase rapidly in response to domestic disturbances such as a banking panic or a stock market crash.

2. SHORT-TERM INTEREST RATE BEHAVIOR BEFORE THE FEDERAL RESERVE

Let me expand just a little on that last point and shift the focus temporarily from prices to interest rates, since it was really a concern about financial problems and sharp interest rate movements under the gold standard that led to the Federal Reserve Act. Because the nation's monetary gold stock was relatively unresponsive to domestic economic conditions in the short run, the National Banking Era was characterized by considerable short-term interest rate variability. Sudden sustained short-term interest rate spikes of over 10 percentage points occurred on eight occasions during this period. Some, though not all, of these spikes were associated with banking panics, which involved a loss of confidence in the banking system and a rush to convert bank deposits into currency. Since banks held only a fractional reserve of coin and currency in their vaults, "bank runs" generated a scramble for liquidity that could not be satisfied in the short run. Major banking panics occurred in 1873, 1884, 1890, 1893 and 1907.

In addition to the recurring interest rate spikes, there was a pronounced seasonal pattern in short-term interest rates. This pattern resulted from the relatively strong demand for currency during the fall harvest and Christmas holiday seasons. It was exacerbated by the reserve requirement provisions of the National Bank Act, which led to a phenomenon known as "pyramiding"—the concentration of reserves in big-city banks. The practice of counting correspondent balances as legal reserves, combined with the payment of interest on interbank balances, caused reserves to concentrate in the larger cities, especially in New York. The withdrawal of interbank balances in peak agricultural and holiday periods tended to exacerbate seasonal pressures on the banking system. Consequently, short-term interest rates varied seasonally by as much as 6 percentage points over the course of a year.

3. THE FEDERAL RESERVE'S MANDATE IN 1914

This background information is essential in understanding what President Wilson and the Congress had in mind when they passed the Federal Reserve Act. *The Federal Reserve was established in 1914 in large part to alleviate the two main problems of the National Banking Era: (1) recurrent interest rate spikes associated with liquidity crises and banking panics, and (2) interest rate seasonals exacerbated by reserve pyramiding.* Specifically, as stated in its preamble, the purposes of the Federal Reserve Act were "to provide for the establishment of Federal Reserve banks, to furnish an elastic currency, to afford means of rediscounting paper, to establish a more effective supervision of banking in the United States, and for other purposes."

Under the Act, 12 Federal Reserve Banks (including ours in Richmond) were established around the country as depositories for the required reserves that previously had been held at correspondent banks in New York City and elsewhere. By requiring that private banks hold reserves directly in a Federal Reserve Bank, the Act eliminated reserve pyramiding and eased the seasonal strain on the banking system.

The most important power given the new central bank, however, was the authority to issue currency and to create bank reserves at least partly independently of the nation's monetary gold reserves. The Fed could create currency and reserves as long as the Federal Reserve Banks kept a minimum 40 percent gold reserve against Federal Reserve notes, which were paper currency, and a 35 percent gold reserve against deposits held by private banks at Federal Reserve Banks. These minimum gold reserve ratios made the Fed respect the discipline of the gold standard; however, *the monetary gold stock was so large during the Fed's early years that these requirements were not "binding."* In other words, they did not constrain the volume of Federal Reserve notes that could be issued nor the volume of bank reserve deposits that could be created by Reserve Bank discount window lending. The power to create currency and bank reserves enabled the Fed to do what it had been established to do: eliminate both the seasonal in interest rates and the periodic spikes in rates that had plagued the country during the National Banking Era.

4. PRICE STABILITY IN THE FED'S EARLY YEARS

The Expectation

As we have just seen, the new central bank was well equipped to deal with both seasonal and special liquidity pressures and their effects on interest rates. But we need now to shift our focus back to the price level and ask: What did the Federal Reserve System and its ability to create currency and bank reserves imply for the stability of the price level—that is, the stability of the purchasing power of money? The answer is that it was taken for granted that the minimum gold reserve ratio under the gold standard would continue to provide what economists call a nominal anchor for the monetary system, which is a fancy way of saying that it would provide for a reasonably stable price level over time. (As a footnote, I should point out here that the framers of the Federal Reserve Act apparently did not give much attention to the *intermediate* drift of the price level upward and downward which, as I mentioned earlier, can and did occur under the gold standard.) The clear presumption underlying the Act was that the new central bank would concern itself mainly with making liquidity available on a timely basis to smooth short-term movements in interest rates. Any discretionary injection of currency or bank reserves for this purpose, however, was expected to be only temporary, so that the nation's money supply and price level would, over the long term, be governed by the nation's stock of monetary gold, much as it had been before the establishment of the Fed.

Given this presumption—and this is a crucially important point about the history of central banking in the United States—the Federal Reserve Act did not include a mandate for price stability because everyone expected that the price level in fact would be stable over time as long as the Federal Reserve respected its minimum gold reserve ratio. The gold standard would guarantee price stability and the new central bank could focus on stabilizing the banking system and interest rates. No separate mandate to resist inflation or deflation was needed.

Federal Reserve Policy in the Aftermath of World War I

This was the expectation. Let me turn now to the *reality* of the early years of the Fed—more specifically, the period between 1914 and 1929. The presumptions about the gold standard and price-level stability implicit in the Federal Reserve Act were tested swiftly and severely during these years. In one of the great ironies of monetary history, by the time the Federal Reserve Banks actually opened for business in 1914, the outbreak of World War I in Europe had brought about widespread suspensions of national commitments to maintain the fixed currency price of gold. Because the United States remained neutral until 1917, it was able to remain on the gold standard throughout the War, and, although it embargoed gold exports, it continued to fix the dollar price of gold at $20.67 per ounce.

As it turned out, United States participation in the War and the large federal deficits that accompanied it—yes, there were deficits back then too—occasioned the first major use of the fledgling central bank's power to create currency and bank reserves. Most of the federal deficit was covered by sales of U.S. government bonds to the public. The additional supply of bonds, naturally, put upward pressure on interest rates, which would have greatly increased the cost of financing the War had the pressures been allowed to persist. Consequently, the Reserve Banks held short-term interest rates down by keeping their discount rates low and accommodating credit demand at these rates—which they were able to do because of the excess gold reserves I mentioned earlier. The discount window lending by Federal Reserve Banks, in turn, increased the supply of bank reserves and caused the U.S. money supply to rise.

Now, as you are no doubt aware, rapid money growth produces inflation over time. Consequently, the highly accommodative monetary policy during the War caused the U.S. price level approximately to double. Although the War ended in 1918, Federal Reserve policy remained accommodative in 1919 in an effort to cushion the negative economic impact of demobilization. The continued rapid growth in Federal Reserve notes and in bank reserves that resulted from this policy, along with the lifting of the wartime gold embargo that allowed gold to flow abroad again, finally mopped up the excess gold and caused the Federal Reserve's gold reserve ratio to become binding in mid-1920, toward the end of President Wilson's second term.

At this point, the Fed finally had to confront the constraints of the gold standard, and it responded affirmatively and aggressively. Faced with the need to defend its gold reserve ratio, the Fed raised its discount rate from 4 percent to 7 percent in 1920, a near doubling. In today's terminology this constituted a sharp "tightening" of monetary policy, and it was strong medicine. The deflationary impact was swift and extraordinary. Prices fell precipitously, and by June 1921 about half of the earlier wartime increase in the price level had been reversed. Unfortunately, the sharp decline in the price level was accompanied by a severe economic contraction and rising unemployment lasting from early 1920 to mid-1921. But by acting as it did, the Fed essentially validated the implicit assumption underlying the Federal Reserve Act—that the country would remain on the gold standard, which would maintain a stable price level over the long run if not the shorter run. To use some current jargon, the Fed attained *credibility* for its commitment to the gold standard and price stability by its stiff tightening of policy in 1920. As a postscript, many monetary historians would argue that the Fed could have achieved *greater* credibility with less economic disruption if it had tightened policy sooner. Regrettably, the cost of failure to resist inflation promptly and decisively when it arises is a lesson the nation has had to learn repeatedly.

Price Stability in the 1920s

After validating the country's commitment to the gold standard in the early '20s, and once it had obtained a cushion of gold reserves above its legal minimum, the Fed began to use its monetary policy powers to achieve a greater degree of short-term price-level stability. Under the able leadership of Benjamin Strong, Governor of the Federal Reserve Bank of New York, the Fed deliberately began to offset the effect of temporary gold inflows on the U.S. money supply by selling equivalent amounts of securities from its portfolio. Likewise, temporary short-term outflows of gold were offset by security purchases. Such "sterilization" insulated the U.S. economy from the money supply and aggregate demand instability that gold flows would have caused had they been allowed to affect currency and bank reserves.

Aggregate economic conditions were favorable during most of the period from 1922 to 1929, in my view, partly because the Fed recently had won at least belated credibility for its commitment to price stability by defending the gold reserve ratio in 1920 and 1921, partly because of Strong's extraordinarily skillful discretionary containment of inflation, and partly because of the absence of severe economic shocks. Unfortunately, at the end of the decade, these foundations began to crumble. After having been partially restored in the '20s, the international gold standard became increasingly fragile and deflationary. Moreover, Governor Strong died an untimely death in 1928, which robbed the Fed of strong leadership. Thus the Fed—bereft of any *explicit* price stability mandate—was simply unable to maintain a discretionary monetary policy aimed at price stability. The consequence was a 30 percent decline in prices in the early 1930s and the most terrible economic depression in American history.

Before moving to my concluding comments about the "now" period in the title of this talk, it may be helpful to summarize briefly the principal points about the "then" period. The main point is that the Federal Reserve Act did *not* mandate the Federal Reserve System to control inflation or stabilize the price level; instead, it instructed the Fed, in effect, to smooth interest rates. The reason for the omission of a price stability mandate was that it was assumed that the gold standard would produce long-run price stability because the Fed would adhere to its minimum gold reserve ratio over time. The Federal Reserve was successful in pursuing price stability in the 1920s in part because of favorable underlying economic and financial conditions in the period between 1921 and 1929. But prices were also stable because the Fed had made its price stability objective credible by strongly defending its minimum gold reserve ratio early in the decade. Subsequently, the Fed reinforced its commitment by sterilizing gold inflows under the skillful leadership of Benjamin Strong. Once Strong and the favorable economic climate were removed, however, the absence of a price stability mandate led inexorably to the debacle of the 1930s.

While what happened during the Depression decade of the 1930s obviously is very important in U.S. monetary history, I must move on now from the "then" part of my talk to the concluding "now" part. We shall see that at least some of the deficiencies in the institutional structure of American monetary policymaking that existed in 1929 *still* exist, and that they present some risks, although the risks are different from those of the earlier period.

5. INFLATION IN THE 1970s AND 1980s

We pick up our story a half-century later in the mid-1970s. At the time, inflation had been rising slowly but steadily since the early 1960s. The U.S. dollar and, through it, the world's other major currencies, had been linked to gold under an arrangement known as the Bretton Woods System after the town in New Hampshire where the agreement had been forged at the end of World War II. Under the arrangement, the U.S. had pledged to maintain convertibility of the dollar into gold at $35 per ounce. *But when excessively accommodative monetary policy and gold outflows caused the Federal Reserve's then 25 percent gold reserve ratio to become binding in the mid-'60s, in sharp contrast to the Fed's behavior in 1920 and 1921, the gold reserve requirement was eliminated.* After some attempts to repair the Bretton Woods System, it finally collapsed in 1973.

The year 1973 is generally remembered as the year of the first oil price shock, but it was also a watershed in U.S. monetary history. Before 1973 there was a sense that both the domestic and international monetary systems should retain at least some link to gold, even though the country had not really permitted the gold standard rules to constrain monetary policy for some time. Since 1973, however, there has been a general—although not universal—belief that the gold standard is a thing of the past. Consequently, for the last 20 years the Fed has lacked even the weak Bretton Woods commitment to gold that would have anchored the price level at least over the very long run and helped it deliver price stability. Since the Federal Reserve was originally designed to operate in an institutional environment with at least some such commitment, one might have expected Congress, as a matter of logic, to give the Fed an explicit price stability mandate when the Bretton Woods System fell apart. Unfortunately, no clear mandate has been forthcoming, although Congressman Stephen Neal of North Carolina introduced an amendment to the Federal Reserve Act in 1989 and has reintroduced it every year since that would provide such a mandate. The Neal Amendment (sometimes referred to as the "zero inflation amendment") would require the Fed, over a period of time, to eliminate inflation as a significant factor in economic and business decisions. The Fed supports this

Amendment, and I personally believe its passage would benefit the American economy enormously.

As you probably know, Congress *did* pass legislation in the late 1970s that requires the Fed to set and report targets for the growth of the U.S. money supply. Many people, including your speaker, were hopeful at the time that this legislation would yield more stable and noninflationary money growth rates, and, hence, a more stable price level. But, frankly, it did not work well in this period. As measured by the Consumer Price Index, the inflation rate rose from 4.9 percent in 1976 to 13.3 percent in 1979 and 12.5 percent in 1980. To be sure, the higher inflation partly reflected the continued sharp increases in oil prices in this period. But it is also true that money supply growth exceeded its targets almost continuously throughout the late 1970s. This performance created doubts about the Fed's commitment to the targets, which encouraged inflationary price- and wage-setting behavior even before the oil price shock. Congress' willingness to accept the inflationary money growth rates, and its failure to mandate the Federal Reserve to stabilize prices, further undermined the public's confidence that inflation would be resisted. In short, by the late 1970s the Fed had little if any credibility as an inflation fighter or as a defender of the purchasing power of the dollar.

6. AGGRESSIVE INFLATION FIGHTING IN THE 1980s

By the time Paul Volcker became Federal Reserve Chairman in August 1979, the inflation outlook had begun to deteriorate rapidly. The widely publicized announcement on October 6, 1979, of the Federal Reserve's intention to control money growth more closely inaugurated a period of aggressive inflation fighting. The announcement signaled financial markets and the country that the Fed was prepared to take responsibility for delivering low inflation, even without an explicit mandate for price stability from Congress.

But the announcement was just the beginning. Because the Fed's credibility as an inflation fighter had been so badly compromised, the System had to follow the announcement with strong actions to demonstrate its intent, much as the Fed had had to do in the early 1920s. And strong action *was* taken in the form of a severe tightening of policy that took short-term interest rates from around 11 percent in late 1979 to 17 percent by April 1980 and ultimately to around 20 percent by early 1981. This was the sharpest tightening the Federal Reserve had ever engineered in so short a time. The action succeeded in bringing inflation down to around 4 percent in 1982. In addition, in a manner similar to the early 1920s, it greatly enhanced the Fed's credibility as a defender of the purchasing power of the dollar, although—in another parallel to the '20s—it was accompanied by a sharp and costly contraction. This credibility, combined with (in yet another parallel to the '20s) the able leadership of Chairman Volcker and his successor, Alan Greenspan, has enabled the Fed to maintain the low inflation rate in subsequent years and, indeed, to reduce it somewhat further to a trend rate currently of approximately 3 percent.

7. IMPLICATIONS OF THE PARALLELS BETWEEN THE ’20s AND THE ’80s: A CONCLUDING COMMENT

As we have seen, Federal Reserve policy in the early 1980s had much in common with that of the 1920s. Both decades opened with periods of exceedingly tight monetary policy in response to earlier accelerations of inflation, and the restrictive policies succeeded in bringing inflation sharply downward in both periods. Beyond this, the Fed’s strong actions in each instance conferred upon it an enhanced credibility that helped keep inflation low for the remainder of the decade. Moreover, unusually capable central bankers in both periods took advantage of this credibility to pursue price stability with essentially discretionary actions, even though Strong was acting within the overall framework of the gold standard in the earlier period.

There is one final, less comforting comparison between the two periods, however, that needs to be drawn. As I have indicated, the Fed entered the 1930s without Benjamin Strong, with an eroding and exceedingly deflationary gold standard, and with no alternative, explicit price stability mandate. Currently, the Fed is moving toward the end of this century and the beginning of the next in a stronger and qualitatively different condition. Inflation, rather than deflation, is the current concern. Economic conditions are more tranquil now than they were at the end of 1929, despite the many problems we still face. Further, in my opinion the Fed currently enjoys energetic and very capable leadership. *However, as in 1929,* there is no clear mandate for the Fed to pursue price-level stability. This makes many of us who work at the Fed uneasy, and it explains why the Federal Reserve supports Congressman Neal’s Amendment, which, as I noted earlier, would provide us with such a mandate.

In short, ladies and gentlemen, under present institutional arrangements surrounding the conduct of American monetary policy, maintenance of a sound dollar in the longer-term future will require continued strong leadership at the Fed, an absence of major destabilizing economic shocks like the oil shocks of the 1970s and, ultimately, a measure of good luck. The continuation of all these circumstances indefinitely would be fortuitous. I don’t feel very comfortable in this situation, and you shouldn’t feel comfortable either—especially the younger people in the audience. This economic issue may seem less immediate and pressing than some of the others you’ve faced over the last day and a half. But I can assure you that it is no less important. We need to resolve it promptly.

Unit Roots in Macroeconomic Time Series: Some Critical Issues

Bennett T. McCallum

An enormous amount of analytical literature has recently appeared on the topic of “unit roots” in macroeconomic time series. Indeed, tests for the presence of unit roots and techniques for dealing with them have together comprised one of the most active areas, over the past decade, in the entire field of macroeconomics. The issues at hand have involved substantive questions about the nature of macroeconomic growth and fluctuations in developed economies and technical questions about model formulation and estimation in systems that include unit-root variables. The present paper attempts to describe several of the main issues and to evaluate alternative positions. It does not pretend to be a comprehensive survey of the literature or to provide an “even-handed” treatment of issues, however.[1] Instead, it attempts to develop a convincing perspective on the topic, one that is consistent with the views of many active researchers in the area but that may nevertheless be somewhat idiosyncratic.

The exposition that is presented below is designed to be predominantly nontechnical in nature. Indeed, it takes a rather old-fashioned approach to

This paper has been prepared for the Research Department of the Federal Reserve Bank of Richmond. The author, H. J. Heinz Professor of Economics at Carnegie Mellon University and research associate at the National Bureau of Economic Research, is indebted to John Campbell, David DeJong, Marvin Goodfriend, Robert King, Peter Schmidt, and Mark Watson for suggestions and useful criticism. The views expressed do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] For other recent survey articles, see Stock and Watson (1988) and Campbell and Perron (1991).

econometric issues and uses recently developed concepts only sparingly. It does, however, rely extensively on notational conventions involving the time series "lag operator," L. Under these conventions the symbol L may be manipulated as if it were an algebraic symbol while its effect, when applied to a time series variable, is to shift the variable's date back in time by one period. Thus $Lx_t = x_{t-1}$ while $bLLx_t = bL^2x_t = bx_{t-2}$, etc. In addition, the notation $\alpha(L)$ will denote a polynomial expression in the lag operator as follows: $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \alpha_3 L^3 + \ldots$. Therefore, $\alpha(L)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots$. Using this notation, then, a distributed-lag regression relation of y_t on current and lagged values of x_t could be written as $y_t = \alpha(L)x_t + \epsilon_t$, with ϵ_t a stochastic disturbance term. Furthermore, polynomials in L, which are often restricted to have only a finite number of terms, may be "multiplied" as in the following example:[2] if $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2$ and $\beta(L) = \beta_0 + \beta_1 L$, then $\alpha(L)\beta(L) = \beta_0\alpha_0 + \beta_0\alpha_1 L + \beta_0\alpha_2 L^2 + \alpha_0\beta_1 L + \alpha_1\beta_1 L^2 + \alpha_2\beta_1 L^3$. Finally, "division" by a lag polynomial means that the implied inverse, $\alpha^{-1}(L)$, is a polynomial such that $\alpha^{-1}(L)\alpha(L) = 1$. Thus $\alpha(L)\beta^{-1}(L)$ yields a polynomial $\gamma(L)$ such that $\beta(L)\gamma(L) = \alpha(L)$. It should be mentioned that the first coefficient of a lag polynomial, such as α_0, is often normalized so as to equal one.

A brief outline of our discussion is as follows. In Section 1, the distinction between trend stationarity and difference stationarity of time series is introduced. That distinction is then related to the "unit root" concept in Section 2, which is primarily devoted to a description of attempts by researchers to determine whether the time series of real GNP values for the United States is difference or trend stationary (i.e., does or does not have an autoregressive unit root). Two approaches, involving different strategies for the specification of maintained and tested hypotheses, are discussed. Then in Section 3 a third approach, which presumes that the real GNP series is a sum of trend-stationary and difference-stationary components, is considered. From the discussion in Sections 2 and 3 together, it is concluded that the relevant question is not whether the GNP series is difference stationary, but what is the relative contribution of the two components. It is also concluded that an accurate answer is not obtainable with the amount of data available.

In Section 4 the topic changes to the question of how to process trending data before conducting regression studies relating two or more variables. The answer that is developed is that the choice between differencing and deterministic trend removal is normally of secondary importance, the principal relevant consideration being the serial correlation properties of the regression residuals. This regression context is continued in Section 5, which concerns the topic of cointegration. It is argued that strict cointegration is probably rather rare, since relationship disturbances will usually be—like shocks to univariate series—

[2] The proper term is "convolution." Any reader who desires a more rigorous description of lag operators may consult Dhrymes (1971).

a sum of stationary and difference-stationary processes. Examples relating to money demand and purchasing-power-parity studies are provided. In Section 6, finally, some conclusions are tentatively put forth.

1. STOCHASTIC VS. DETERMINISTIC TRENDS

As most readers are well aware, many macroeconomic data series display upward tendencies or "trends" when observations are plotted against time. For many purposes it is useful and/or conventional to work with detrended values of these variables—i.e., versions from which trend components have been removed. Traditionally, most researchers would effect this detrending step by subtracting from the raw numbers (or their logs) a deterministic[3] trend expression such as $\alpha_0 + \alpha_1 t$, where t is a time index. For various reasons it is often useful to express the basic series in terms of logarithms of the raw data, in which case α_1 becomes a measure of the per-period growth rate of the variable in question. Thus if y_t is the basic variable, the traditional detrending procedure implicitly splits y_t into two components, one representing the trend and the other a cyclical or non-trend component.[4] With y_t the basic variable and ϵ_t a white-noise[5] disturbance, we have

$$y_t = \alpha_0 + \alpha_1 t + \gamma(L)\epsilon_t, \tag{1}$$

where $\alpha_0 + \alpha_1 t$ is the trend component and $\gamma(L)\epsilon_t$ is the non-trend component (or the detrended series). In this traditional decomposition, it is assumed that the detrended series $\gamma(L)\epsilon_t$ is a stationary stochastic process, which requires (among other things) that the population means $E[\gamma(L)\epsilon_t]$, variances $E[\gamma(L)\epsilon_t]^2$, and autocovariances $E[\gamma(L)\epsilon_t\gamma(L)\epsilon_{t-j}]$ are the same for all t. (Here the variance and covariance expressions are written under the presumption that the means equal zero.) Accordingly, y_t is said to be a *trend-stationary* variable; it may have a trend component but its deviations from a deterministic trend are stationary. A variable's status with regard to stationarity is of importance in its own right, as we shall see in a moment, and also because there is a large body of statistical techniques whose validity depends upon stationarity of the variables being analyzed.

At least since 1982,[6] however, many researchers have preferred an alternative model of the trend vs. non-trend decomposition. Instead of (1), they use

[3] That is, non-stochastic.

[4] Throughout, our discussion will ignore *seasonal* components.

[5] A white-noise random variable is one generated by a process that specifies that each period's value, ϵ_t, is drawn from a population with mean zero and finite variance σ^2_ϵ, and is not dependent on previous values.

[6] This was the year in which the article by Nelson and Plosser (1982) was published. The popularity of differencing had been growing gradually, however, at least since the much earlier publication of Box and Jenkins (1970).

a formulation such as (2),

$$\Delta y_t = \beta + A(L)\epsilon_t, \tag{2}$$

where $A(L)\epsilon_t$ is stationary and β represents the average per-period change (or growth rate) of the variable y_t (or the variable whose log is y_t). In this formulation y_t is said to be a *difference-stationary* variable, i.e., one that is generated by a difference-stationary time series. Such a variable cannot in general be made stationary by the removal of a deterministic trend; instead, the series needs to be first-differenced prior to processing.

The basic distinction between trend-stationary (TS) and difference-stationary (DS) variables is that the former do, and the latter do not, tend to return to a fixed deterministic trend function. Since the non-trend component $\gamma(L)\epsilon_t$ in (1) is stationary with mean zero, the process is such that y_t tends to fluctuate about the fixed trend function $\alpha_0 + \alpha_1 t$. In formulation (2), by contrast, the tendency is for y_t to grow at the rate β from its *current* position, whatever that might be. There is, except in a special limiting case, no tendency for y_t to return to any fixed trend path.

The distinction between TS and DS series was emphasized in a highly influential paper by Nelson and Plosser (1982). In this paper, the authors clearly described the TS vs. DS distinction and also discussed the undesirable statistical consequences of detrending by the traditional technique of removing a deterministic time function when in fact the series is generated by a DS process. In addition, Nelson and Plosser (1982) presented evidence suggesting that many important U.S. time series are of the DS class and went on to argue that evidence indicates that U.S. business cycles are largely real as opposed to monetary in nature, i.e., that real shocks have been the principal sources of cyclical variability with the contribution of monetary shocks being entirely of secondary importance. The last of these arguments was not found convincing by the macroeconomics profession—see, e.g., McCallum (1986) and West (1988)—but the hypothesis that many important series (including real GNP) reflect DS processes became quite widely accepted. More recently, opinion has partially reversed itself—as we shall see below—but for the past eight to ten years the idea that real GNP is not trend stationary has been viewed by a large number of researchers as true (and important). It will be useful, consequently, to devote some attention to the logic of the statistical tests that led researchers to that position. In the process of presenting this logic, several relevant points of importance will be brought out—including the meaning of the term "unit root."

2. A UNIT ROOT IN U.S. GNP?

Consider now the TS representation (1) with the lag polynomial $\gamma(L)$ written as the ratio of two other polynomials $\theta(L)$ and $\phi(L)$, both assumed with little loss of generality to be finite[7] of order q and p, respectively. Thus we have

$$y_t = \alpha_0 + \alpha_1 t + \theta(L)\phi^{-1}(L)\epsilon_t. \tag{3}$$

Now suppose that $1/\rho$ is the smallest root of the polynomial $\phi(L)$, i.e., is the smallest number[8] that satisfies the equation $1 + \phi_1 z + \cdots + \phi_p z^p = 0$.[9] Then $\phi(L)$ could be written as $(1 - \rho L)\tilde{\phi}(L)$ and multiplication of (3) by $(1 - \rho L)$ would give[10]

$$(1 - \rho L)y_t = \alpha_0(1 - \rho) + \rho\alpha_1 + \alpha_1(1 - \rho)t + \theta(L)\tilde{\phi}^{-1}(L)\epsilon_t. \tag{4}$$

And in the special case in which $\rho = 1$, the latter collapses to

$$(1 - L)y_t = \alpha_1 + \theta(L)\tilde{\phi}^{-1}(L)\epsilon_t. \tag{5}$$

Since $(1 - L)y_t$ equals Δy_t, then, the latter is of the same form as (2). Consequently, when there is a unit root to $\phi(L)$—when $1/\rho = 1.0$—representation (1) yields, as a special case, the DS formulation (2).[11]

In light of the foregoing result, a natural test procedure is suggested for determining whether "y_t has a unit root"—i.e., whether the AR polynomial $\phi(L)$ has a unit root so that y_t is DS. What is involved is that the researcher maintains the hypothesis that (1) is true, represents it as in equation (4), and then tests the ("null") hypothesis that ρ in (4) is equal to one. If the latter hypothesis is rejected, then one concludes that y_t is not a DS series. But if the hypothesis $\rho = 1.0$ is not rejected, then one can in a sense conclude that y_t is a DS variable—or that the behavior of y_t is not significantly different from that of a DS variable. Because ordinary asymptotic distribution theory breaks down in the case in which ρ is precisely equal to one, a consistent test requires that the relevant "t-statistic" on the coefficient of y_{t-1} be compared with a critical value taken from a nonstandard distribution. But this can readily be done, since Dickey and Fuller (1979) have provided the profession with the pertinent tables.

The foregoing procedure was in fact employed by Nelson and Plosser (1982) to test for unit roots in over a dozen important U.S. time series. In only one of these could the tested hypothesis that $\rho = 1.0$ be rejected at a

[7] That is, to include only a finite number of terms with nonzero coefficients. Any stationary stochastic process can be closely approximated by an expression of the form $\theta(L)\phi^{-1}(L)\epsilon_t$.

[8] Perhaps a complex number.

[9] Consider, for example, the second-order case. Then $1 + \phi_1 z + \phi_2 z^2 = 0$ could equivalently be written as $(1 - \alpha_1 z)(1 - \alpha_2 z) = 0$, where $\phi_1 = -(\alpha_1 + \alpha_2)$ and $\phi_2 = \alpha_1\alpha_2$. But the latter equation is satisfied by the two values $z^1 = 1/\alpha_1$ and $z^2 = 1/\alpha_2$. So the lag polynomial $1 + \phi_1 L + \phi_2 L_1^2$ could as well be expressed as $(1 - \alpha_1 L)(1 - \alpha_2 L)$. The roots of the polynomial $\phi(L)$ are said to be $1/\alpha_1$ and $1/\alpha_2$.

[10] Here $(1 - \rho L)(\alpha_0 + \alpha_1 t) = \alpha_0 - \rho\alpha_0 L + \alpha_1 t - \rho\alpha_1(t - 1) = \alpha_0(1 - \rho) + \alpha_1 t - \rho\alpha_1 t + \rho\alpha_1 = \alpha_0(1 - \rho) + \alpha_1(1 - \rho)t + \rho\alpha_1$.

[11] If $\rho > 1.0$, then y_t will have explosive behavior of a type that seems unlikely and that will become easily detectable after a few years.

conventional significance level (i.e., 0.01 or 0.05), so the authors' tentative conclusion was that most U.S. macroeconomic data series are of the DS class, i.e., are unit-root variables.

There was, however, one rather obvious difficulty with this tentative conclusion,[12] as follows: while it was not possible to reject the hypothesis that the series' roots like ρ were equal to one, it would also have been impossible to reject hypotheses asserting that these roots equaled 0.99, for example, or even 0.95. But with ρ equal to 0.99 or 0.95, the model would be one of the TS class. Continuing with this perspective, it might be argued that it is highly implausible that the tested hypothesis of ρ equal to unity would hold *precisely,* as opposed to approximately. The data, that is, could do no more than show that the value of ρ is *close to* one. Consequently, this entire testing approach, which begins with a TS representation and posits the DS model as a special case, seemed highly unconvincing to a number of analysts.[13]

An alternative approach would be to begin with a maintained hypothesis implying difference stationarity and then express trend stationarity—the absence of an AR unit root—as a special case. Thus the time series process for y_t could be written as in (2) but with $A(L) = \theta(L)\phi^{-1}(L)$:

$$\Delta y_t = \beta + \theta(L)\phi^{-1}(L)\epsilon_t. \tag{6}$$

Then if the moving-average lag polynomial $\theta(L)$ were to have a unit root so that $\theta(L) = (1 - L)\tilde{\theta}(L)$, expression (6) could be operated on by $(1 - L)^{-1}$ to yield

$$y_t = \beta_0 + \beta t + \tilde{\theta}(L)\phi^{-1}(L)\epsilon_t. \tag{7}$$

(That $[1-L]^{-1}\beta$ equals $\beta_0+\beta t$ can be justified by multiplying each by $[1-L]$.) Consequently, it would be possible to express (6) as

$$\phi(L)\Delta y_t = \beta\phi(L) + (1 - \gamma L)\tilde{\theta}(L)\epsilon_t, \tag{8}$$

estimate the latter, and test the hypothesis that $\gamma = 1$. If it were possible to *reject* the latter, then the outcome might be viewed as providing more convincing evidence in favor of the DS view.[14]

In fact, the influential paper by Campbell and Mankiw (November 1987) proceeded in precisely this fashion, using quarterly postwar data for the United States, 1947–1985. So what did these authors find? As it happens, the answer is not straightforward because it is unclear how many terms should be included in estimation of the $\phi(L)$ and $\theta(L)$ polynomials in (8). In their paper, Campbell and Mankiw reported results for 16 different cases representing all possible

[12] The difficulty was recognized, but not emphasized, by Nelson and Plosser (1982).

[13] See, e.g., McCallum (1986, pp. 405–6).

[14] It would, however, be possible to object that expressing trend stationarity as a zero-measure special case effectively biases the procedure in favor of a DS finding. Note, incidentally, that a unit root in the MA polynomial does not imply a process of the "unit root" type.

Table 1 Test Statistics from Campbell and Mankiw (November 1987, Table I)

Number of AR Parameters	Number of MA Parameters		
	1	2	3
1	22.96*	11.73*	0.00
2	2.06	4.02*	0.00
3	0.95	1.31	0.00

Notes: Tabulated entries are values of 2 log (SSE^0/SSE), where SSE denotes the sum of squared residuals and SSE^0 indicates imposition of the constraint that makes $A(1) = 0$. The ARMA models are estimated for Δy_t where y_t is the log of U.S. real GNP, seasonally adjusted, quarterly for 1947:1–1985:4. Asterisks indicate values that are significantly different from zero (0.05 significance level) under the usual test, but this test is inappropriate (as discussed in the text).

combinations of zero to three AR parameters and zero to three MA parameters. Of these, it is arguable that only those with at least one AR and one MA term should be seriously entertained. The usual test statistics for those nine cases are given in Table 1. For each case, the reported number is the likelihood ratio statistic for a test of the hypothesis that $\theta(L)$ has a unit root—i.e., that the TS hypothesis is true. In most cases this statistic has asymptotically,[15] under the null hypothesis, a chi-square distribution with one degree of freedom, so that the critical value is 3.84 for a test with significance level 0.05 (or 6.63 for a 0.01 level). Based on these values, the table indicates that in three of the nine cases—i.e., for three of the nine specifications—the null TS hypothesis can be rejected at the 0.05 level. Described in this fashion, then, the Campbell and Mankiw (November 1987) results did not provide strong evidence against the TS hypothesis (or, in favor of the unit root hypothesis). But under the particular hypothesis of concern in this case, that $\theta(L)$ has a unit root, the usual asymptotic distribution theory breaks down—as it does when testing for a unit root in the AR polynominal $\phi(L)$. This breakdown tends to reduce the critical level for the likelihood ratio statistics and to produce an excessive number of extreme values such as those in the final column of Table 1. Thus the figures in that table are actually more unfavorable for the TS hypothesis than they appear to be at first glance.

Furthermore, Campbell and Mankiw did not describe the results as in the previous paragraph. Instead, they suggested that the ARMA (2,2) model[16]—the case with two autoregressive and two moving average parameters—commands special attention because it is not significantly worse than the (2,3) or (3,2)

[15] I.e., in the limit as the sample size grows without bound.

[16] An ARMA model is one that admits both autoregressive and moving average polynomials. The notation (p, q) indicates how many terms (p and q) are included in the AR and MA polynomials. Sometimes the number of times d that the basic variable has been differenced is included in a (p, d, q) notation.

models and is significantly better than the (2,1) model (and somewhat better than the [1,2] specification).[17] But in this case, the results (see Table 1) call for rejection of the TS null hypothesis, even given the test's bias toward acceptance. The suggestion of Campbell and Mankiw, consequently, was that postwar quarterly evidence supports the notion that real GNP for the U.S. is *not* trend stationary, but instead is generated by a DS (or unit root) process. We shall return to the persuasiveness of this suggestion below. But first it will be useful to discuss a different aspect of the Campbell and Mankiw analysis, which their discussion emphasized.

In particular, a notable feature of the Campbell-Mankiw (November 1987) paper is its presentation of an attractive measure of the ultimate or "long-run" response of y_t to a unit shock, i.e., a 1.0 realization of the disturbance ϵ_t. To define this measure, consider again the DS formulation (2), $\Delta y_t = \beta + A(L)\epsilon_t$, and write it out as

$$y_t = y_{t-1} + \beta + \epsilon_t + A_1\epsilon_{t-1} + A_2\epsilon_{t-2} + \cdots . \tag{9}$$

From the latter expression, it can be seen that the per-unit effect of ϵ_t on y_t is 1.0 (in the sense that if ϵ_t were to equal some positive value instead of its mean zero, then y_t would be higher by the same amount.) But then the per-unit effect of ϵ_t on y_{t+1} would be 1 + A_1, with the part A_1 occurring "directly" and the remainder through its effect on y_t. Continuing with this line of reasoning, it is found that the (per-unit)[18] effect on y_{t+k} would be $1+A_1+A_2+\cdots+A_k$. In the limit as $k \to \infty$, then, we would have $1+A_1+A_2+\cdots$, which may be denoted $A(1)$. (That expression arises from writing $A(L) = 1 + A_1L + A_2L^2 + \cdots$ and inserting 1 wherever L appears.) Thus the measure $A(1)$ reflects the ultimate or long-run effect of ϵ_t on y_t when the process generating y_t is of form (2).

An important property of the measure $A(1)$ is that its value is zero for any TS process. To see that, write $A(L) = \theta(L)\phi^{-1}(L)$ and recall that for a TS variable the MA polynomial can be written as $\theta(L) = (1 - L)\tilde{\theta}(L)$. Thus we have

$$A(L) = (1 - L)\tilde{\theta}(L)\phi^{-1}(L) \equiv (1 - L)a(L) = a(L) - La(L), \tag{10}$$

where $a(L) \equiv \tilde{\theta}(L)\phi^{-1}(L)$. But then we obtain

$$A(1) = a(1) - La(1) = a(1) - a(1) = 0 \tag{11}$$

since $La(1) = L(1 + a_1 + a_2 + \cdots) = 1 + a_1 + a_2 + \cdots$. Thus if $\theta(L)$ can be written as $(1 - L)\tilde{\theta}(L)$, as it can when the process at hand is TS, it is true that $A(1) = 0$.

---

[17] Here the meaning of "model A is better than B" is that B is nested in A and can be rejected with a significance level of 0.05.

[18] Henceforth the words "per unit" will typically be omitted.

Table 2 Estimates of $A(1)$ from Campbell and Mankiw (November 1987, Table IV)

Number of AR Parameters	Number of MA Parameters 1	2	3
1	1.72	1.73	0.03
2	1.77*	1.52	0.00
3	1.36*	1.60*	0.00

Notes: See Table 1.

What about the values of $A(1)$ implied by various DS processes? For each of these $A(1)$ will be nonzero, but will take on various values depending on the response pattern. In particular, $A(1)$ will exceed one if the ultimate impact on y_t of a shock is greater than the first-period impact. An important special case is provided by the random-walk process in which $\Delta y_t = \beta + \epsilon_t$. In this case $A(L) = 1 + 0L + 0L^2 + \cdots = 1$ so $A(1) = 1$. Next, the first-order MA case has $\Delta y_t = \beta + \epsilon_t + \theta\epsilon_{t-1}$ so $A(L) = 1 + \theta L$ and $A(1) = 1 + \theta$. Then $A(1)$ is greater than or smaller than one depending on whether θ is positive or negative.

A somewhat more general process is the ARMA (1,1) model for Δy_t, namely,

$$(1 - \phi L)\Delta y_t = (1 + \theta L)\epsilon_t. \qquad (12)$$

In this case $A(L) = (1+\theta L)(1-\phi L)^{-1}$ so $A(1) = (1+\theta)/(1-\phi)$. An example application is provided by the Campbell-Mankiw (November 1987) estimates with the U.S. GNP series. Their point estimates of ϕ and θ are 0.522 and -0.179, respectively, so that $A(1) = (1 - .179)/(1 - .522) = 0.821/0.478 = 1.717$. Thus the ARMA (1,1) model for Δy_t suggests that the long-run response of y_t (log of GNP) to a shock will be about 1.7 times as large as the immediate (within one quarter) response.

In sum we see that the measure $A(1)$ provides an attractive way of expressing the magnitude of the "long-run response" of a variable (y_t) to a unit shock (ϵ_t).[19] And in their study of postwar U.S. GNP, Campbell and Mankiw (November 1987) find that for all but three of their nine cases[20] $A(1)$ is substantially larger than one, implying that the impact of shocks is to cumulate, rather than dissipate, as time passes. Their values are reported in Table 2, where it may be noted that for the ARMA (2,1), (3,1) and (3,2) cases (marked with asterisks), the point estimates of $A(1)$ are large even though the $A(L)$ polynomials are not significantly different from ones with $A(1) = 0$ according to the

[19] Another useful measure has been featured in the work of Cochrane (1988). It is described below in text attached to footnote 26.

[20] Recall that they actually reported 16 cases, but that we are focusing on 9.

usual test (compare Table 1). Consequently, although they are guarded in their statements, Campbell and Mankiw seem to conclude that there are grounds for being reasonably confident that the long-run response of a unit shock to U.S. GNP is substantially greater than one. In this sense, shocks to GNP have no trend-reversion tendency at all.[21]

This conclusion has not, however, held up to subsequent criticism. One major reason for skepticism was provided in a study by Christiano and Eichenbaum (1990). Basically, their study emphasized that the different ARMA specifications considered by Campbell and Mankiw give rise to quite different values of $A(1)$—as is evident in Table 2—and that there is very little reason to conclude that any one of them is appropriate—or even that one of those with $A(1) > 1$ is. The Christiano-Eichenbaum argument is that relevant inferences are sensitive to the choice of ARMA specification employed, even within the set of those that provide approximately equally good fits to the data.

One of the experiments conducted by Christiano and Eichenbaum will illustrate their results. In this experiment they conducted simulations with a model with parameters matching those estimated by Campbell and Mankiw in the ARMA (3,3) case. In other words, they pretended that this case—which implies $A(1) = 0$—is true, and then considered what would happen if it were studied under the assumption that the ARMA (2,2) specification were correct. For each simulation they would generate 150 "data" points using the (3,3) parameters, then estimate a (2,2) model and test the hypothesis that $A(1) = 0$. They conducted 2,000 such simulations and found that the hypothesis $A(1) = 0$, which was true in the process studied, was nevertheless rejected in 74 percent of the simulations.[22] Similarly, in 2,000 more simulations, based on the ARMA (1,3) parameter estimates from Campbell and Mankiw, it was found that the true hypothesis $A(1) = 0$ was rejected in 38 percent of the simulations.

The conclusion reached by Christiano and Eichenbaum was as follows: on the basis of 150 observations, about the number of quarterly postwar data periods, it is not possible to make accurate inferences about the long-run response measure $A(1)$. Equivalently, it is not possible to determine with high reliability whether the stochastic process generating real GNP observations is of the TS or DS class.

During the last few years, numerous additional papers on the topic have appeared; only a few can be mentioned. Sims (1988) has suggested that Bayesian techniques of statistical inference are more appropriate than classical in this particular context and DeJong and Whiteman (1989, 1991) have presented Bayesian results that provide support for the view that the U.S. GNP process is actually of the TS class. That conclusion has been strongly challenged by Phillips (1991), in a paper that provided the basis for a symposium occupying an

[21] It is sometimes said that they are highly "persistent," but that terminology is inappropriate for reasons clearly described by Cochrane ("Comments," 1991, pp. 206–7).

[22] With a test statistic designed to have a 0.05 significance level.

entire issue of the *Journal of Applied Econometrics*. The symposium includes rejoinders by Sims and DeJong-Whiteman; it would be difficult to identify any clear outcome. Others, including Stock (1991), Cochrane (April 1991), Sowell (1992) and Rudebusch (1993), have reached the Christiano-Eichenbaum (1990) conclusion—i.e., that it is not possible with existing data to settle the issue—by alternative means. In my opinion, this last conclusion seems generally appropriate, but there is another way of approaching the issue that is conceptually rather simple and perhaps illuminating.

3. THE UNOBSERVED COMPONENTS APPROACH

In the previous section two approaches were mentioned, based on equations (3) and (6). In the first of these, the maintained hypothesis is trend stationarity with difference stationarity viewed as a zero-measure[23] special case, whereas in (6) the DS hypothesis is maintained and TS is treated as the (zero-measure) special case. Let us now consider an alternative approach that proceeds within a framework in which both TS and DS components are presumed to play a role, the implied statistical problem being to determine how much weight to give to each. Aspects of this "unobserved components" approach have been developed by Harvey (1985), Watson (1986), Clark (1987), and Cochrane (1988).

The analysis presented by Clark (1987) provides a useful introduction and perspective. It begins by writing the observable variable under study, y_t, as the sum of a DS "trend" term z_t and a stationary "cycle" term x_t :

$$y_t = z_t + x_t. \tag{13}$$

Although a more general specification would be possible, Clark assumes that the cyclical component is a pure AR process so that $\phi(L)x_t = v_t$, with v_t white noise. Indeed, in his empirical implementation with U.S. GNP data Clark makes $\phi(L)$ a second-order polynomial, so that x_t is an ARMA (2,0). The trend component is assumed to obey

$$z_t = z_{t-1} + d + w_t, \tag{14}$$

where w_t is white noise, independent of v_t. Actually Clark takes the drift term d to be itself a random walk: $d_t = d_{t-1} + u_t$ with u_t white. But empirically he finds the variability of u_t to be very small, so we shall for simplicity view d_t as a constant, as in (14). The model at hand for y_t is therefore

$$y_t = (1 - \phi_1 L - \phi_2 L^2)^{-1} v_t + (1 - L)^{-1}(d + w_t). \tag{15}$$

Let us consider, then, how (15) fits the U.S. quarterly postwar GNP data.

[23] I.e., a case represented by parameter values that would be graphically represented as a point (with zero area) in a region depicting all possible parameter values.

Table 3 Estimates of ARMA (2,2) Models from Clark (1987)

Parameters and Statistics	Constrained	Unconstrained
ϕ_1	1.548	0.658
ϕ_2	−0.601	−0.420
θ_1	−1.214	−0.335
θ_2	0.248	0.529
SE	0.0103	0.0103
Q(10)	7.9	4.8
$A(1)$	0.64	1.57

Notes: ARMA models estimated for Δy_t (see Table 1) for 1948:1–1985:4. Q(10) denotes the Box-Pierce Q-statistic for ten autocorrelations of the residuals; under the null hypothesis of white noise it has asymptotically a chi-square distribution with ten degrees of freedom.

As a preliminary, it will be instructive to consider a comparison that begins by expressing (15) as

$$(1-\phi_1 L-\phi_2 L^2)\Delta y_t=(1-L)v_t+(1-\phi_1 L-\phi_2 L^2)(d+w_t). \tag{16}$$

Here the right-hand side is the sum of two independent MA processes, with the higher-order one being an ARMA (0,2). Using Granger's Lemma,[24] then, we can write (16) as

$$(1-\phi_1 L-\phi_2 L^2)\Delta y_t=\delta+(1+\theta_1 L+\theta_2 L^2)\epsilon_t, \tag{17}$$

where ϵ_t is an implied, constructed white-noise disturbance and where $\delta = d(1-\phi_1-\phi_2)$. But the representation in (17) has six parameters ($\phi_1, \phi_2, \theta_1, \theta_2$, σ_ϵ^2, and δ) whereas the basic model (15) has only five ($\phi_1, \phi_2, \sigma_v^2, \sigma_w^2$, and d). So the particular components model at hand, which sums an AR (2) component and a random-walk component, can be viewed as a *constrained* version of an ARMA (2,2) model for Δy_t.

It is of course true that the unconstrained model (17) must fit the data at least slightly better than the constrained version (15). But Clark's estimates, reported in Table 3, indicate that in the case at hand there is almost no difference, i.e., almost no deterioration in fit, from imposing the constraint. In particular, the estimated residual variance for (15) is essentially the same as with (17) and the Box-Pierce Q(10) statistic is not much worse. So the constrained version—the components model (15)—could as well be the preferred choice.[25]

[24] Granger's Lemma says that the sum of two independent ARMA processes, one ARMA (p_1, q_1) and the other ARMA (p_2, q_2), is an ARMA (p^*, q^*) where $p^* \leq p_1 + p_2$ and $q^* \leq \max(p_1+q_2, p_2+q_1)$. For pure MA processes, then, $q^* \leq \max(q_1, q_2)$.

[25] Both Clark (1988) and Cochrane (1988) have developed arguments suggesting that unconstrained ARMA models with difference series tend to be poor at the job of estimating long-run properties such as $A(1)$.

But although the constrained and unconstrained ARMA models fit the data about the same, they yield very different $A(1)$ measures. Whereas the unconstrained version gives $\hat{A}(1) = 1.57$, virtually the same as estimated by Campbell and Mankiw, for the (unconstrained) components model the estimate is 0.64.

In two diagrams, Clark (1987) presented evidence apparently suggesting that for U.S. quarterly GNP the unobserved components model may provide a better estimate than the unconstrained ARMA of the long-run response statistic $A(1)$. The first of these, denoted Figure V in Clark's paper, plots the implied autocorrelations at various lags for the two models (plus one more, an ARMA [0,2]) and for the Δy_t sample. In that plot it will be seen that the unconstrained ARMA (denoted ARIMA 212) matches the sample somewhat better at short lags (e.g., 1–5 quarters) but that the components model provides a better match at lags of 5-20 quarters. More striking are the related results shown in Clark's Figure VI, which plots the variance ratios V_k/V_1, where $V_k \equiv (1/k)\,\mathrm{Var}(y_t - y_{t-k})$, for lag lengths k up to 60 quarters.[26] In this case, the apparent superiority of the components model's match to the sample data is striking. But, as Campbell and Mankiw (May 1987) point out, sample values of V_k provide biased estimates of their population counterparts. Accordingly, Campbell and Mankiw suggest that the sample values should be multiplied by $T/(T - k)$, where T is the sample size. Here $T = 148$, so the adjusted sample values of V_k are considerably larger than the unadjusted values for $k \geq 20$.

With this bias adjustment incorporated, the match between sample and components-model values of V_k would continue to be somewhat better than between sample and unconstrained ARMA values, but not nearly to the same extent as in Clark's Figure VI. The same point applies, but with less force, to his Figure V.

More generally, the unobserved components approach to modeling the trend vs. cyclical decomposition seems conceptually attractive, in part because it does not treat either TS or DS processes as (zero-measure) special cases. The implied question is not whether one of these two possibilities can be rejected, but instead is "How important quantitatively is the z_t as opposed to the x_t component?" That question cannot be answered in precisely the stated form, since the variance of the DS component z_t depends on the horizon considered and goes to infinity in the limit. But one type of answer is provided by the $A(1)$ measure itself and another by a comparison of the variances of v_t and w_t, i.e., the shocks to x_t and z_t. In the case at hand, Clark's estimates are $\hat{\sigma}_v = 0.0072$ and $\hat{\sigma}_w = 0.0066$.

An objection to the components approach as implemented by Clark (1987) and Watson (1986) was expressed by Campbell and Mankiw (May 1987, p. 115). This objection is that with the DS component z_t modeled as a random walk, the estimated value of $A(1)$ must lie between zero and one; thus values

[26] As $k \to \infty$, the limit V of the V_k sequence is the long-run response measure proposed by Cochrane (1988) mentioned above in footnote 19. Its relation to $A(1)$ is $V = (1 - R^2)[A(1)]^2$, where R^2 is $1 - (\sigma_\epsilon^2/\mathrm{Var}\,\Delta y_t)$.

greater than one are ruled out a priori. But while this important objection is applicable to the Clark and Watson studies, it is not applicable to the approach in general, for the latter can accommodate other DS processes for z_t. Instead of a random walk, for example, the z_t process could be specified as a first-order MA: $\Delta z_t = d + w_t + \theta w_{t-1}$. For the z_t component alone, $A(1)$ would then equal $1+\theta$ so values in excess of one will be obtained if $\theta > 0$. And if the variability of z_t is large in relation to that for x_t, then the $A(1)$ value for y_t could easily exceed one.[27]

Another objection is that it is unreasonable to assume that x_t and z_t components are independent, as the approach presumes. There is undoubtedly some merit to this point, since technology shocks will presumably have both cyclical and long-lasting effects. But the Campbell-Mankiw ARMA approach amounts to use of an unobserved components model in which the shocks (like v_t and w_t in [15]) are *perfectly* correlated,[28] which property seems even less desirable.

Perhaps the most important objection to the unobserved components modeling of trend vs. cycle is that it is computationally much more difficult than estimation of ARMA models, the necessary steps involving application of Kalman filter techniques. For a discussion of such techniques, the reader is referred to Harvey (1981).

On the basis of the foregoing discussion, it would seem reasonable to conclude that the postwar U.S. quarterly real GNP process is most likely of the DS class, since a sum of DS and TS components is itself a DS process.[29] But it is far from clear that the long-run impact of shocks exceeds that of the random-walk case in which $A(1) = 1.0$. Instead, a measure such as 0.6, which attributes a substantial share of GNP variability to a stationary component, is just as plausible. What does seem clear is that it is not possible, on the basis of currently available data, to estimate $A(1)$ with much accuracy or reliability.

Conceptually, the basic components-approach idea, of viewing a time series as the sum of DS and TS processes, seems attractive as a framework for thinking about the properties of univariate time series. In many cases, both components would be presumed to be of non-negligible importance so many series will be of the DS class. That does not imply, however, that any particular method can be relied upon for trend vs. cyclical decomposition of time series data.

---

[27] But with more parameters in the DS component, the components model may become equivalent to an unconstrained ARMA.

[28] See Watson (1986, p. 53).

[29] Quite recently, Kwiatkowski, Phillips, Schmidt, and Shin (1992) have conducted tests of the hypothesis that the DS component is of negligible importance in a components formulation. For the real GNP series this hypothesis was found to be of borderline significance at the 0.05 level.

4. DETRENDING PRIOR TO ECONOMETRIC ANALYSIS

In this section we switch our attention away from trend estimation, conducted for the purpose of isolating trend and cyclical components of a series, and toward trend removal (or "detrending"), conducted for the purpose of obtaining series suitable for econometric analysis of relationships among variables. In this context, then, the issue is whether to process variables prior to (say) regression analysis by removal of an estimated deterministic trend[30] or by differencing of the series. A major reason for detrending is that the standard formulae for standard errors, test statistics, etc., are in most cases based on asymptotic distribution theory that assumes stationarity of the regressor variables.[31] Belief that some variable is generated by a process of the DS type—i.e., one with a unit root—might then lead to the presumption that data differencing would be preferable for that variable prior to its use in a regression study.

Other influential arguments for differencing of data prior to time series econometric analysis were put forth by Granger and Newbold (1974) and Nelson and Kang (1984). In the earlier of these papers it was shown that a regression relating y_t to x_t would spuriously tend to find a relationship when in fact y_t and x_t are generated independently but by random-walk processes. The Nelson-Kang piece emphasized a tendency for trendless random-walk variables to be spuriously related to time trends in estimated regressions.

As a result of these and other studies, considerable support developed during the mid-1980s for the position that differencing should routinely be carried out prior to regression analysis involving time series data.[32] The case for such a practice was succinctly summarized by Plosser and Schwert (1978, p. 653) as follows: "Ignoring the effects of underdifferencing can be a far more costly error than ignoring the effects of overdifferencing." More recently, there has been significant counter-movement based on phenomena related to the concept of "cointegration." A consideration of that position will be presented below, but it will be useful first to consider the merits of routine differencing, rather than detrending, of variables with an apparent trend component.

[30] In least-squares regression analysis the inclusion of a time trend among the regressors is equivalent to the use of variables detrended by prior regression on the same time variable (i.e., using residuals, from these prior regressions on time, as the detrended variables).

[31] The standard formulae do not rely on asymptotic distribution theory in cases in which there are no lagged dependent variables in the system under study, but such cases are the exception in applied macroeconomics.

[32] A contrary argument is that differencing sacrifices information pertaining to levels or to long-run relationships. Estimation of a levels relationship after differencing will not, of course, provide any information about the constant term, but that is usually of little importance. The argument developed below suggests that little is lost with regard to long-run multipliers unless the variables are cointegrated, a topic that is taken up briefly in Section 6.

The issues at hand can be usefully introduced and illustrated in an example similar to that used by Plosser and Schwert (1978). Consider a linear regression relationship that is (by assumption) correctly specified in first differences, viz.,

$$\Delta y_t = \beta \Delta x_t + \epsilon_t, \tag{18}$$

where ϵ_t is a white-noise disturbance with variance σ_ϵ^2 and where x_t is exogenous, generated by a process independent of the process generating ϵ_t. Now, if instead of (18) the investigator estimates by ordinary least squares (OLS) the relationship between x_t and y_t in levels, he is in effect applying OLS to

$$y_t = \alpha + \beta x_t + \eta_t, \tag{19}$$

in which the disturbance term $\eta_t = \epsilon_t + \epsilon_{t-1} + \cdots$ is serially correlated and nonstationary. In this *underdifferenced* case, as Plosser and Schwert point out, the OLS estimator of β could be inconsistent, depending on the process generating x_t. In any event, whether or not the OLS estimator is consistent, its sampling distribution does not have finite moments. Inferences based on the usual OLS formulae are likely, accordingly, to be highly inappropriate.

Next, suppose that instead the investigator applies OLS to the second differences of y_t and x_t, estimating

$$\Delta(\Delta y_t) = \beta \Delta(\Delta x_t) + \Delta \epsilon_t. \tag{20}$$

In this case with *overdifferencing* the disturbance $\Delta\epsilon_t$ is again serially correlated but now its distribution is stationary. The OLS estimator of β will be unbiased and consistent, but will be inefficient and its sampling variance will (except in special cases) not be consistently estimated by the usual formulae.

These foregoing considerations, discussed by Plosser and Schwert (1978), are of some interest but are actually relevant only under the presumption that the investigator is wrong about the appropriate degree of differencing *and* makes use of OLS estimators even though the implied disturbances are serially correlated. Of considerably greater interest, it would seem, are the consequences of estimating β with underdifferenced or overdifferenced data when the investigator recognizes the presence of serial correlation in the OLS residuals and responds by utilizing an estimator designed to take account of autocorrelated disturbances in the appropriate manner. In the overdifferenced case, for example, the true relation can be written as

$$\Delta(\Delta y_t) = \beta \Delta(\Delta x_t) + \epsilon_t + \theta \epsilon_{t-1}, \tag{21}$$

with $\theta = -1.0$. The interesting question, then, is whether the investigator will be led seriously astray if he regresses $\Delta(\Delta y_t)$ on $\Delta(\Delta x_t)$ using an estimation procedure designed for cases in which the disturbance process is a first-order MA.

Now precisely this last question has been investigated via Monte Carlo experimentation[33] by Plosser and Schwert (1977). They find that even though the absolute value of θ tends to be somewhat underestimated—with a sample size of $T = 100$ the mean across 1,000 replications of the estimates of θ is about -0.94—the estimates of β are not appreciably biased and the experimental sampling distribution is not such as to lead frequently to incorrect inference. Specifically, the frequency of rejection of a true hypothesis with a nominal significance level of 0.05 is 0.063 in one experiment and 0.081 in the other. Plosser and Schwert conclude, appropriately, that “the cost associated with overdifferencing may not be large when care is taken to analyze the properties of regression disturbances” (1978, p. 643).

The corresponding case of an investigation with underdifferencing arises if we write the true relation as

$$y_t = \alpha + \beta x_t + (1 - \rho L)^{-1}\epsilon_t, \tag{22}$$

with $\rho = 1.0$, and ask whether the investigator will be led seriously astray (regarding β) if he regresses y_t on x_t under the assumption that the disturbance process is a first-order AR. With respect to this possibility, Plosser and Schwert (1978, p. 643) recognize that “if the resulting estimate of ρ is close to one, as it should be in this case, differencing would be indicated leading to the correct model:..” They do not, however, consider the effects on the estimation of β of concluding one’s investigation with the estimate provided by the levels regression that takes account of AR disturbances—which is the situation corresponding to the presumed behavior of the investigator in the overdifferencing case. This asymmetry in discussion prevents them from giving a comparison of the relative costs of underdifferencing vs. overdifferencing when the investigator is intelligently taking account of the serial correlation properties of the disturbances.

Some Monte Carlo results relevant to this type of procedure have, however, been obtained by Harvey (1980) and Nelson and Kang (1984). The latter of these papers is devoted primarily to emphasizing various ways in which investigators could be led to misleading results if they estimate underdifferenced relationships and do *not* correct for serially correlated residuals, but it briefly reports (on pp. 79–80) results of testing a true hypothesis analogous to $\beta = 0$ in (22) with β and ρ estimated jointly. With $T = 100$ and a significance level of 0.05, the frequency of rejection in 1,000 replications is 0.067, which compares favorably with the Plosser-Schwert results for the case with overdifferencing. The study by Harvey (1980) compares mean-squared-error (MSE) values[34] for

[33] This approach is used because the usual asymptotic distribution theory breaks down in cases with unit roots in either the MA or AR polynomial.

[34] Across 200 replications.

estimates of β in (22) with $\rho = 1.0$ when estimated with first differences and when estimated jointly with ρ using levels data (i.e., with underdifferencing and an autocorrelation correction). Two specifications regarding the behavior of the exogenous variable x_t are considered by Harvey. In one of these the x_t process is stationary; in that case the MSE value for the estimator of β is 0.310 with the (correct) first-difference specification and 0.309 with underdifferencing (and autocorrelation correction).[35] In the other case, which features strongly trending x_t data, the analogous MSE figures are 0.061 and 0.078.[36]

Also of relevance, though not conforming to the symmetric contrast provided by our specifications (18) and (20), is evidence provided by Harvey relating to the estimation of a relation like (22) but with $\rho = 0.9$. The alternative estimators are based on application of maximum likelihood to the (correct) levels specification and OLS to the first-differenced specification, the latter amounting to estimation with $\rho = 1.0$ by constraint.[37] The $T = 100$ MSE values are 0.263 and 0.262 for the two estimators with stationary x_t's, and 0.009 vs. 0.018 with trending x_t's.

On the basis of the described Monte Carlo experiments, the appropriate conclusion would seem to be that neither overdifferencing nor underdifferencing leads to serious estimation or testing mistakes in regression models with exogenous regressors, provided that the investigator takes intelligent account of serial correlation present in the regression residuals. It is perhaps worth noting, given the tenor of their discussion, that this conclusion is not contradicted in the least by the four studies involving actual data (and unknown specifications) that are explored by Plosser and Schwert (1978).

Specifically, in each of these four cases the authors conclude that first differencing is probably appropriate, but the point estimates and standard errors (for the parameter analogous to β) that are provided by regressions with undifferenced data are virtually the same when the Cochrane-Orcutt procedure is used to account for $\rho \neq 0$. In their Table 1 regression of (log) income on the (log) money stock, for example, the slope coefficient (and standard error) values are 1.127 (0.122) for the Cochrane-Orcutt levels regression and 1.141 (0.126) in the differenced case. The OLS regressions with data that have been differenced twice give estimates that do not agree as well, but in each of these cases there is evidence of *uncorrected* serial correlation in the residuals. In Table 1, for example, the first residual autocorrelation is -0.36.

It is additionally worth noting that Plosser and Schwert (1978, p. 638) also take the view that "the real issue is not differencing, but an appropriate appreciation of the role of the error term in regression models." Thus our

[35] Actually the estimator "with autocorrelation correction" involves full maximum likelihood estimation of (22).

[36] These values are for sample size of $T = 100$; Harvey also gives results for $T = 20$ and $T = 50$.

[37] In the levels formulation, ρ is estimated jointly with β.

disagreement with Plosser and Schwert seems to be whether the "representative investigator" will, or will not, recognize and take steps in response to the presence of autocorrelated residuals.

The foregoing evidence relates, however, principally to relations with exogenous regressors. In practice, it is much more common for equations of interest to include one or more lagged endogenous variables. But if Δy_{t-1} were to appear as an additional regressor in (18), then the situation regarding estimation of β would be quite different. In order to obtain a bit of evidence as to the validity of the suggestion—that the presence or absence of differencing is not crucial when serial correlation corrections are applied—in situations in which lagged endogenous variables are present, let us consider results pertaining to two example relationships (that may be of some substantive interest).

Specifically, let us first consider estimation of the rudimentary single-equation model of aggregate demand utilized in McCallum (1987), that is, an equation relating growth of nominal GNP to growth rates of the monetary base. Notationally, let x_t and b_t denote logarithms of nominal GNP and the base, respectively, for period t and consider quarterly observations, seasonally adjusted, for the sample period 1954:1–1991:3.[38]

As a starting point, consider the following updated version of the specification emphasized in McCallum (1987):

$$\begin{array}{l} \Delta x_t = \underset{(.002)}{0.0078} + \underset{(.073)}{0.3248}\,\Delta x_{t-1} + \underset{(.104)}{0.3190}\,\Delta b_t. \\ R^2 = 0.196 \qquad SE = 0.0097 \qquad DW = 2.12 \qquad Q(10) = 8.3 \end{array} \tag{23}$$

Here parameter standard errors are shown in parentheses while the reported statistics are the unadjusted R^2, the estimated standard deviation of the disturbance term, the Durbin-Watson statistic, and the Box-Pierce Q-statistic based on the first ten autocorrelation terms.[39] These statistics give no evidence of residual autocorrelation and it is the case that Δx_{t-2} would not provide additional explanatory power. As it happens, however, inclusion of Δb_{t-1} would provide additional explanatory power and would make Δb_t insignificant. Accordingly, let us switch our attention to the variant of (23) in which Δb_t is replaced by Δb_{t-1}, a variant also used in McCallum (1987). The 1954:1–1991:3 estimates are as follows:

$$\begin{array}{l} \Delta x_t = \underset{(.002)}{0.0076} + \underset{(.075)}{0.2845}\,\Delta x_{t-1} + \underset{(.105)}{0.3831}\,\Delta b_{t-1}. \\ R^2 = 0.215 \qquad SE = 0.0096 \qquad DW = 2.07 \qquad Q(10) = 8.0 \end{array} \tag{24}$$

[38] Data for 1953-1990 are taken from the Citibase data set, while 1991 values come from the *Survey of Current Business* (GNP) and the Federal Reserve Bank of St. Louis (adjusted monetary base). Calculations are performed with version 7.0 of Micro TSP.

[39] Under the assumption that the disturbances are white noise, Q(10) has asymptotically a chi-squared distribution with eight degrees of freedom; its critical value for a 0.05 significance level is therefore 18.3.

Here there is no evidence of residual autocorrelation and additional lagged values of Δx_t and Δb_t would not enter significantly. The important properties of the estimated relation are that Δx_t is mildly autoregressive and is positively related to Δb_t, with a moderately large elasticity value that is not significantly different from 0.5.

The first question to be answered, then, is "What would we have found if we had estimated this same relation in (log) levels, using series with deterministic trends removed?" To develop an answer, first consider equation (25), where the detrending is effected by inclusion of time as an additional regressor:

$$\begin{aligned} x_t = \underset{(.072)}{0.0273} + \underset{(.0002)}{0.00021}\, t + \underset{(.020)}{1.0160}\, x_{t-1} - \underset{(.014)}{0.0321}\, b_{t-1}. \\ R^2 = 0.9999 \quad SE = 0.0104 \quad DW = 1.40 \quad Q(10) = 23.1 \end{aligned} \tag{25}$$

Here the results are entirely different from those in (24), but there is distinct evidence of residual autocorrelation. Re-estimation with the disturbance term assumed to follow an AR(1) process yields

$$\begin{aligned} x_t = \underset{(34.1)}{5.857} + \underset{(.060)}{0.0067}\, t + \underset{(.081)}{0.2763}\, x_{t-1} + \underset{(.150)}{0.592}\, b_{t-1} + \underset{(.023)}{0.996}\, u_{t-1}, \\ R^2 = 0.9999 \quad SE = 0.0095 \quad DW = 2.14 \quad Q(10) = 9.0 \end{aligned} \tag{26}$$

where u_t is defined as $(1 - \rho L)^{-1}\epsilon_t$. Now, with the AR(1) disturbance specification, we estimate the autocorrelation parameter to be very close to one and the magnitude of the coefficients attached to x_{t-1} and b_{t-1} revert to the neighborhood of the corresponding values in the differenced relation (24).[40] The trend term is insignificant, as was the constant in (24), and qualitatively the relation in (26) is quite similar to the version estimated in differences.

Next, we move in the opposite direction by differencing the variables one more time than in the reference case (24). Let $\Delta\Delta x_t \equiv \Delta(\Delta x_t)$ for brevity. Then with the disturbance treated as white noise, the result is

$$\begin{aligned} \Delta\Delta x_t = \underset{(.0009)}{0.0002} - \underset{(.074)}{0.3993}\, \Delta\Delta x_{t-1} + \underset{(.158)}{0.363}\, \Delta\Delta b_{t-1}. \\ R^2 = 0.182 \quad SE = 0.0110 \quad DW = 2.12 \quad Q(10) = 18.3 \end{aligned} \tag{27}$$

Here the estimated parameter on the lagged GNP variable is entirely unlike that in (24), but the Q-statistic gives borderline evidence of serial correlation. Estimated with a MA(1) specification for the disturbance, the results change to:

$$\begin{aligned} \Delta\Delta x_t = \underset{(.0008)}{0.00001} + \underset{(.065)}{0.1666}\, \Delta\Delta x_{t-1} + \underset{(.139)}{0.3571}\, \Delta\Delta b_{t-1} - \underset{(.032)}{0.946}\, e_{t-1}. \\ R^2 = 0.370 \quad SE = 0.0097 \quad DW = 1.89 \quad Q(10) = 9.5 \end{aligned} \tag{28}$$

[40] The coefficient on the base variable is now somewhat larger than 0.5, rather than smaller, but the difference is less than two standard errors.

Now the figures are again quite close to those in the once-differenced specification (24). Not only the estimated parameter values, but also the standard errors are approximately the same—and there is no evidence of serial correlation. Thus the results are similar for regressions using detrended levels, differences, and second differences of the x_t and b_t variables, provided that autocorrelation corrections are used.

A second example concerns spot and forward exchange rates. In a recent paper (McCallum 1992), I have summarized some empirical regularities for the post-1973 floating rate period, focusing on \$/£, \$/DM, and \$/Yen rates over the time span 1978:01–1990:07. Letting s_t and f_t denote logs of the spot and 30-day forward rates at the end of month t, one of the observed regularities is that OLS regression of s_t on f_{t-1} provides a tight fit with a slope coefficient very close to one—see the estimates reported in the first panel of Table 4. When Δs_t is regressed on Δf_{t-1}, however, the relationship disappears and the estimated slope coefficient becomes insignificantly different from zero—see the second panel of Table 4. That contrast would seem to contradict the argument of the preceding paragraphs since there is little indication of serial correlation in the residuals in either case.

The results in panel three, however, support our previous argument. There the levels equation relating s_t and f_{t-1} is reestimated with an AR(1) specification for the disturbance process, even though the DW and Q-statistics in the top panel do not clearly call for such a step. And for all three exchange rates the result is the same—the AR parameter ϕ is estimated to be close to one with the slope coefficient on f_{t-1} becoming indistinguishable from zero. The results in panel three, in other words, are essentially equivalent to those in panel two, even though differenced data are used in the latter and not in the former.

In addition, the specification using second differences together with a MA(1) disturbance is implemented in the fourth panel of Table 4. There the DM case differs slightly from the previous results, the slope coefficient on f_{t-1} being estimated as about 0.21 and significant, but for both the £ and Yen rates the previous results are obtained again—the slope coefficient is close to zero with the MA parameter being estimated in the vicinity of -1.0. For five out of the six comparisons with the reference case of panel two, then, the results are in strong agreement despite contrasting treatment in terms of differencing. And even in the sixth case, the extent of disagreement is relatively minor. Thus the evidence is again supportive of the general argument that the extent of differencing is not crucial, in the context of detrending of variables prior to econometric analysis, provided that residual autocorrelation corrections are utilized.[41]

[41] It should be said explicitly that this argument is not being made with regard to Granger causality tests or variance decomposition statistics in vector-autoregression studies. It is my impression that these results are rather sensitive to the detrending procedure. But such results are, I believe, of less importance than impulse response patterns.

Table 4 Spot on Forward Exchange Rate Regressions, Sample Period 1978:01–1990:07

		Estimates (std. errors)				Statistics			
Rate	**Variables**	**Const.**	**Slope**	**AR(1)**	**MA(1)**	R^2	**SE**	**DW**	**Q(10)**
\$/DM	s_t on f_{t-1}	−0.009	0.990			0.963	0.0362	2.05	9.6
		(.012)	(.016)						
\$/£	" "	0.014	0.977			0.960	0.0359	1.82	5.9
		(.009)	(.016)						
\$/Yen	" "	−0.046	0.991			0.975	0.0380	1.84	7.7
		(.067)	(.013)						
\$/DM	Δs_t on Δf_{t-1}	0.002	−0.063			0.004	0.0358	1.96	7.5
		(.003)	(.081)						
\$/£	" "	0.000	0.024			0.000	0.0352	1.99	3.4
		(.003)	(.082)						
\$/Yen	" "	0.003	0.038			0.001	0.0374	1.99	4.8
		(.003)	(.082)						
\$/DM	s_t on f_{t-1}	−0.575	−0.057	0.991		0.964	0.0359	1.96	8.1
		(.510)	(.083)	(.015)					
\$/£	" "	0.509	0.035	0.979		0.962	0.0351	1.99	3.8
		(.143)	(.084)	(.017)					
\$/Yen	" "	−4.743	0.046	0.989		0.976	0.0374	2.00	5.0
		(.681)	(.083)	(.0l3)					
\$/DM	$\Delta\Delta s_t$ on $\Delta\Delta f_{t-1}$	0.000	−0.206		−0.887	0.526	0.0361	1.86	4.1
		(.003)	(.057)		(.039)				
\$/£	" "	0.000	−0.032		−0.933	0.472	0.0357	1.96	3.1
		(.003)	(.060)		(.035)				
\$/Yen	" "	0.000	−0.010		−0.956	0.468	0.0379	1.94	4.0
		(.003)	(.060)		(.027)				

Data source: Bank for International Settlements.

5. COINTEGRATION

Now suppose that x_t and y_t are two time series variables generated by DS processes—i.e., their univariate series have AR unit roots—that are dynamically related by a distributed-lag relation with a stationary disturbance. In (29), for example, we assume u_t to be stationary:[42]

$$y_t = \alpha + \beta(L)x_t + u_t. \tag{29}$$

[42] It is not being assumed that x_t is necessarily a predetermined variable, i.e., that u_t is uncorrelated with $x_t, x_{t-1}, \cdots$.

Under these conditions y_t and x_t are said to be *cointegrated*, the term arising because DS variables are referred to by many time series analysts as "integrated."[43] Now it is a striking fact that when y_t and x_t are cointegrated, then an OLS regression of y_t on x_t alone—with no lags—will yield a slope coefficient b that is a consistent estimator of the "long-run" effect $\beta(1) = \beta_0 + \beta_1 + \cdots$. This result would appear to be of practical importance, as it promises to provide a simple way of discovering features of long-run relationships between variables. To demonstrate the result, let us express the residual $e_t = y_t - bx_t$ as

$$e_t = \alpha + \beta(L)x_t + u_t - bx_t = \alpha + [\beta(L) - b]x_t + u_t. \tag{30}$$

But with x_t an integrated (DS) variable, e_t will then be integrated unless $\beta(1) - b = 0$. And if e_t were integrated, then the sum of squared e_t values would increase without limit as the sample size goes to infinity, so the OLS criterion of picking b to minimize this sum forces b toward $\beta(1)$.

There are numerous additional theoretical results concerning cointegrated variables including extension to multivariate settings[44] and close connections between cointegration and the existence of "error correction" forms of dynamic models.[45] For present purposes, however, the main item of interest concerns the frequently expressed contention that if two (or more) DS variables are not cointegrated, then there exists no long-run relationship between (or among) them. On the basis of this notion, various researchers have concluded that purchasing-power-parity fails even as a long-run tendency (see, e.g., Taylor [1988] and McNown and Wallace [1989]) whereas others have drawn analogous conclusions regarding traditional money demand relations—see, e.g., Engle and Granger (1987).[46] Cuthbertson and Taylor (1990, p. 295) have stated the matter thusly: "If the concept of a stable, long-run money demand function is to have any empirical content whatsoever, then m_t [log money]... must be cointegrated" with log prices, log income, and interest rates.

Now clearly there is a technical sense in which these suggestions are correct: if y_t and x_t are both DS but not cointegrated, then the disturbance entering any linear relationship between them must (by definition) be nonstationary. So they can drift apart as time passes. I would argue, however, that it is highly misleading to conclude that in any practical sense long-run relationships are

[43] If a variable must be differenced d times to render it stationary, it is said to be integrated of order d, abbreviated $I(d)$. The term "integrated" was popularized by Box and Jenkins (1970), its genesis being that a random-walk variable is at any time equal to the infinite sum ("integral") of all past disturbances. Cointegration analysis was developed by Granger (1983) and Engle and Granger (1987).

[44] See, for example, the expository piece by Dickey, Jansen, and Thornton (1991).

[45] See Hendry (1986).

[46] Other writers have apparently accepted this characterization prior to reaching the opposite empirical conclusion. A few examples are Mehra (1989), Hoffman and Rasche (1991), Miller (1991), Hafer and Jansen (1991), and Diebold, Husted, and Rush (1991).

therefore nonexistent. My argument is entirely interpretive; it includes no suggestion of technical error in the literature criticized. But its importance is not thereby diminished.

To develop the argument at hand, let us take the example of a traditional money demand function of the form

$$m_t - p_t = \beta_0 + \beta_1 y_t + \beta_2 R_t + \eta_t, \tag{31}$$

where $m_t - p_t$ is the log of real money balances, y_t the log of a real transactions variable (such as GDP), and R_t is an opportunity-cost variable relevant to the measure of money being used. Let us suppose for the purpose of the argument that $m_t - p_t$, y_t, and R_t are all DS variables. And let us suppose that $m_t - p_t$, y_t, and R_t have all been processed by removal of a deterministic trend.[47] Then the cointegration status of the relationship depends upon the properties of the disturbance η_t—if its process is of the DS type, the variables in (31) will not be cointegrated.

It is my contention that the traditional view of money demand theory, represented for example by the *New Palgrave* entry by McCallum and Goodfriend (1987), would actually suggest that the variables in (31) are *unlikely* to be cointegrated. The reason is that the rationale for (31) depends upon the transactions-facilitating function of money, but the technology for effecting transactions is constantly changing. And since technical progress cannot be well represented by measurable variables, the effects of technical change not captured by a deterministic trend show up in the disturbance term, η_t. But the nature of technological progress is such that changes (shocks) are typically *not* reversed. Thus one would expect there to be an important permanent component to the η_t process, making it one of the DS type.

In such a situation, however, the "long-run" messages of traditional money demand analysis may continue to apply. Provided that the magnitude of the variance to the innovation in η_t is not large in relation to potential magnitudes of Δm_t values, it will still be true that inflation rates will be principally determined by money growth rates, that long-run monetary neutrality will prevail, that superneutrality will be approximately but not precisely valid, etc. That the disturbance term in the money demand relationship is of the DS class is simply not a source of embarrassment or special concern for supporters of the traditional theory of money demand.[48]

Much the same can be said, furthermore, in the context of PPP doctrine. Nominal exchange rates are probably not cointegrated with relative price levels

[47] This step should not be at issue; the existence of technological change in the payments industry is widely accepted.

[48] Many of these supporters have been willing to estimate money demand functions in first-differenced form, thereby implicitly assuming a DS disturbance process.

because technological and taste shocks affecting real exchange rates have permanent components.[49] But major differences among nations in money growth and inflation rates may nevertheless dominate other effects on nominal exchange rates over long spans of time, leaving the practical messages of the PPP doctrine entirely valid as a long-run matter.[50] That such is the case in actuality is indicated by the evidence collected by Gailliot (1970) and Officer (1980).

In both of the preceding examples, it was argued that one should expect the disturbance term in a relation among levels of economic variables to include both permanent and transitory components, and therefore to possess an autoregressive unit root. This argument—which is an application to disturbance terms of the unobserved-components perspective put forth in Section 3—would seem to be applicable quite broadly; indeed, to the disturbances of most behavioral relations. That point of view implies, unfortunately, that cointegrating relationships will be rare[51] and so the potentially beneficial estimation result mentioned in the first paragraph of this section will not be forthcoming.[52]

The argument of the present section has a natural counterpart, it might be added, in the context of debates concerning non-trend stationarity of the price level. Some commentators, including Barro (1986) and Haraf (1986), have emphasized uncertainty concerning future values of the price level and have accordingly suggested that it is highly undesirable for p_t (log of the price level) to be generated by a unit-root process. The point of view expressed here emphasizes, by contrast, the relative unimportance of p_t nonstationarity per se, given the existing magnitude of the disturbance variance for the p_t process, in comparison with recent values of the *trend* growth rate. One way to express the point is to hypothesize that citizens and policymakers in the United States would view price-level performance as highly satisfactory if it were generated (in quarterly terms) as

$$p_t = \delta + p_{t-1} + \epsilon_t \qquad (32)$$

if $\delta = 0$ and ϵ_t were white noise with $\sigma_\epsilon^2 = 0.00002$. (The latter figure approximately equals the one-quarter forecast variance over 1954–1991.) Looking 20 years ahead, the forecast variance of p_t would be $80(0.00002) = 0.0016$, so a 95 percent confidence interval would be the current value plus or minus 0.08

[49] As suggested, for example, by Stockman (1987).

[50] Here the interpretation of PPP is taken to agree with popular usage, although a good case can be made for an alternative interpretation that expresses PPP as a form of a neutrality proposition.

[51] The s_t, f_t example in Section 5 is, however, a case in which cointegration evidently does obtain.

[52] Campbell and Perron (1991, pp. 218-19) argues against this suggestion by means of a reductio ad absurdum. The latter is not actually applicable, however, as my argument is directed only toward variables that enter agents' utility functions or budget constraints.

(or ± 8 percent in terms of the price level). That figure pales into insignificance in comparison with the *expected* change in p_t over 20 years if δ were nonzero and equal to (e.g.) 0.011, a figure that corresponds to a 4.5 percent annual rate of inflation.

6. CONCLUSIONS

In this final section we shall conclude the arguments. The discussion will not be a summary of what has gone before—which is itself largely a condensation of other work—but instead will attempt to reach *conclusions* in the sense of "logical consequences" of what has gone before. In developing our arguments it will be useful to distinguish the two different purposes of trend analysis that were mentioned above: (i) isolating trend from cyclical components and (ii) trend removal for the purpose of obtaining series suitable for econometric analysis. We begin with subject (ii).

In the context of removing trends from time series so that relationships among these series can be studied by conventional econometric methods, we have seen that there is a tendency for similar results to be obtained from the two methods, provided that serial correlation corrections are applied to the residuals of the relationship being studied. This suggests that it is not crucial whether the analyst differences the data or removes deterministic trends. The recommended course of action would then be, evidently, to estimate the econometric model both ways—with differenced and (deterministically) detrended data—and hope that similar results will in fact be obtained. But emphasis in presentation will usually be given to one set of results or the other, and in some cases the results will not be similar. A natural basis for choice would then be to feature the results that require the smaller amount of correction to remove autocorrelation of the residuals. In the case of the GNP-monetary base example of Section 4, for example, the preferred results would be those in equation (24), rather than (26) or (28). And in the exchange rate example of Table 4, the results in the second panel would be preferred, according to this criterion.

Now consider purpose (i), the estimation of trends so as to isolate trend from cyclical components of a series. In Sections 2-4 above we have reviewed various results all of which indicate that there is no reliable method for distinguishing among alternative trend/cycle decompositions even when these have entirely different long-run response characteristics and different implications about the relative importance of the two components. This seems, at first glance, a discouraging conclusion.

Reflection on the issue suggests, however, that it actually makes very little sense even to attempt to distinguish between trend and cycle on the basis of a variable's univariate time series properties alone. The reason is that the separation of trend and cycle will in most cases be desired because the analyst believes that the two components have different economic properties or significance. With regard to real GNP, for example, Nelson (1989, p. 74) emphasizes that analysts "tend to think of the processes generating the two components as quite different," one being "due to growth in labor force and capital stock and to technological change" and the other "arising largely from monetary [and fiscal] disturbances." But such components will be neither independent nor perfectly correlated, as presumed by the two main trend estimation procedures described above. And without knowledge of the extent of correlation, they are not identified even under the assumption that the trend component is a random walk. This latter assumption, moreover, is itself rather unsatisfactory.

More generally, the distinction between trend and cycle is by many economists viewed as pertaining to movements that are socially desirable and undesirable, respectively. But whether such is the case clearly depends upon the economist's theory of cyclical fluctuations, for some of these—the real business cycle hypothesis, for example—will not view cyclical movements as something that policy should attempt to mitigate. The nature of the cycle vs. trend distinction, in other words, depends upon the theory of macroeconomic fluctuations adopted. But if that is the case, then it makes little sense to attempt to separate out the cyclical component by means of a procedure that takes no account of alternative theories but relies merely on a variable's time series properties.[53]

The reader may have noticed that the remarks in this concluding section have pertained exclusively to trend analysis, with the term "unit roots" failing to appear. More generally, it may have been noted that there is no inevitable connection between the two concepts—unit roots may be present in a series that is entirely trendless (and vice versa). But the presence of trends is a constant source of practical difficulties in the analysis of time series data, and the recent interest in unit roots has stemmed largely from the notion of stochastic trends. It is then for reasons of practicality that emphasis has here been given to the topic of trends. Our principal messages regarding unit roots per se are implicit in our conclusions regarding trends. But since those messages are somewhat negative concerning the value of unit root testing, it needs to be mentioned explicitly that introduction of the unit root concept, together with recognition that series are likely to include DS components, has been a valuable corrective to the earlier habit of presuming trend stationarity and has led to several analytical insights.[54]

[53] It should be noted that this argument does not imply that it is pointless to try to attempt to reach substantive macroeconomic conclusions on the basis of analyses such as that of Blanchard and Quah (1989), which utilizes multiple time series and relies upon explicit substantive assumptions for identification.

[54] A recent example is provided by the related analyses of Fisher and Seater (1993) and King and Watson (1992).

REFERENCES

Barro, Robert J. "Recent Developments in the Theory of Rules versus Discretion," *Economic Journal,* vol. 96 (Supplement, 1986), pp. 23–37.

Blanchard, Olivier J., and Danny Quah. "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review,* vol. 79 (September 1989), pp. 655–73.

Box, George E. P., and Gwilym M. Jenkins. *Time Series Analysis.* San Francisco: Holden-Day, 1970.

Campbell, John Y., and N. Gregory Mankiw. "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics,* vol. 102 (November 1987), pp. 857–80.

__________. "Permanent and Transitory Components in Macroeconomic Fluctuations," *American Economic Review Papers and Proceedings,* vol. 77 (May 1987), pp. 111–17.

Campbell, John Y., and Pierre Perron. "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," *NBER Macroeconomics Annual 1991.* Cambridge, Mass.: MIT Press, 1991.

Christiano, Lawrence J., and Martin Eichenbaum. "Unit Roots in Real GNP: Do We Know, and Do We Care?" *Carnegie-Rochester Conference Series on Public Policy,* vol. 32 (Spring 1990), pp. 7–62.

Clark, Peter K. "Nearly Redundant Parameters and Measures of Persistence in Economic Time Series," *Journal of Economic Dynamics and Control,* vol. 12 (June/September 1988), pp. 447–61.

__________. "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics,* vol. 102 (November 1987), pp. 797–814.

Cochrane, John H. "Comment," *NBER Macroeconomics Annual 1991.* Cambridge, Mass.: MIT Press, 1991.

__________. "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control,* vol. 15 (April 1991), pp. 275–84.

__________. "How Big Is the Random Walk in GNP?" *Journal of Political Economy,* vol. 96 (October 1988), pp. 893–920.

Cuthbertson, Keith, and Mark P. Taylor. "Money Demand, Expectations, and the Forward-Looking Model," *Journal of Policy Modelling,* vol. 12 (Summer 1990), pp. 289–315.

DeJong, David N., and Charles H. Whiteman. "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series,'" *Journal of Monetary Economics,* vol. 28 (October 1991), pp. 221–54.

———. "Trends and Cycles as Unobserved Components in Real GNP: A Bayesian Perspective," *Proceedings of the American Statistical Association Business and Economics Section*, 1989, pp. 63–70.

Dhrymes, Phoebus J. *Distributed Lags.* San Francisco: Holden-Day, 1971.

Dickey, David A., and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, June 1979, pp. 423–31.

Dickey, David A., Dennis W. Jansen, and Daniel L. Thornton. "A Primer on Cointegration with an Application to Money and Income," Federal Reserve Bank of St. Louis *Review*, vol. 73 (March/April 1991), pp. 58–78.

Diebold, Francis X., Stephen Husted, and Mark Rush. "Real Exchange Rates Under the Gold Standard," *Journal of Political Economy*, vol. 99 (December 1991), pp. 1252–71.

Engle, Robert F., and C. W. J. Granger. "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, vol. 55 (March 1987), pp. 251–76.

Fisher, Mark, and John J. Seater. "Long-Run Neutrality and Superneutrality in an ARIMA Framework," *American Economic Review*, vol. 83 (forthcoming 1993).

Gailliot, Henry J. "Purchasing Power Parity as an Explanation of Long-Term Changes in Exchange Rates," *Journal of Money, Credit, and Banking*, vol. 2 (November 1970), pp. 348–57.

Goodfriend, Marvin. "Interest Rate Smoothing and Price Level Trend-Stationarity," *Journal of Monetary Economics*, vol. 19 (May 1987), pp. 335–48.

Granger, C. W. J. "Co-Integrated Variables and Error-Correcting Models," Working Paper 83-13. San Diego: University of California, 1983.

———, and Paul Newbold. "Spurious Regressions in Econometrics," *Journal of Econometrics*, vol. 2 (July 1974), pp. 111–20.

Hafer, R. W., and Dennis W. Jansen. "The Demand for Money in the United States: Evidence from Cointegration Tests," *Journal of Money, Credit, and Banking*, vol. 23 (May 1991), pp. 155–68.

Haraf, William S. "Monetary Velocity and Monetary Rules," *Cato Journal*, vol. 6 (Fall 1986), pp. 641–62.

Harvey, Andrew C. "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, vol. 3 (1985), pp. 216–27.

———. *Time Series Models.* Oxford: Phillip Allen, 1981.

———. "On Comparing Regression Models in Levels and First Differences," *International Economic Review*, vol. 21 (October 1980), pp. 707–20.

Hendry, David. "Econometric Modelling with Cointegrated Variables: An Overview," *Oxford Bulletin of Economics and Statistics,* vol. 48 (August 1986), pp. 201–12.

Hoffman, Dennis L., and Robert H. Rasche. "Long-Run Income and Interest Elasticities of Money Demand in the United States," *Review of Economics and Statistics,* vol. 78 (November 1991), pp. 665–74.

King, Robert, and Mark W. Watson. "Testing Long Run Neutrality," NBER Working Paper No. 4156. September 1992.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics,* vol. 54 (October-December 1992), pp. 159–78.

McCallum, Bennett T. "A Reconsideration of the Uncovered Interest Parity Relationship," NBER Working Paper No. 4113. July 1992.

__________. "The Case for Rules in the Conduct of Monetary Policy: A Concrete Example," Federal Reserve Bank of Richmond *Economic Review,* vol. 73 (September/October 1987), pp. 10–18.

__________. "On 'Real' and 'Sticky-Price' Theories of the Business Cycle," *Journal of Money, Credit, and Banking,* vol. 18 (November 1986), pp. 397–414.

__________, and Marvin S. Goodfriend. "Demand for Money: Theoretical Analysis," in *The New Palgrave: A Dictionary of Economics.* London: Macmillan Press, 1987, reprinted in Federal Reserve Bank of Richmond *Economic Review,* vol. 74 (January/February 1988), pp. 16–24.

McNown, Robert, and Myles S. Wallace. "National Price Levels, Purchasing Power Parity, and Cointegration: A Test of Four High Inflation Economies," *Journal of International Money and Finance,* vol. 8 (December 1989), pp. 533–45.

Mehra, Yash P. "Some Further Results on the Sources of Shift in M1 Demand in the 1980s," Federal Reserve Bank of Richmond *Economic Review,* vol. 75 (September/October 1989), pp. 3–13.

Miller, Stephen M. "Monetary Dynamics: An Application of Cointegration and Error-Correction Modelling," *Journal of Money, Credit, and Banking,* vol. 23 (May 1991), pp. 139–54.

Nelson, Charles R. "Implicit Estimates of the National and Cyclical Components of Japan's Real GNP," *Bank of Japan Monetary and Economic Studies,* vol. 7 (August 1989), pp. 73–91.

__________, and Heejoon Kang. "Pitfalls in the Use of Time as an Explanatory Variable in Regression," *Journal of Business & Economic Statistics,* vol. 2 (January 1984), pp. 73–82.

Nelson, Charles R., and Charles I. Plosser. "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics,* vol. 10 (September 1982), pp. 139–62.

Officer, Lawrence H. "Effective Exchange Rates and Price Ratios Over the Long Run: A Test of the Purchasing-Power-Parity Theory," *Canadian Journal of Economics,* vol. 13 (May 1980), pp. 206-30.

Phillips, P. C. B. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics,* vol. 6 (October/December 1991), pp. 333–64.

Plosser, Charles I., and William G. Schwert. "Money, Income, and Sunspots: Measuring Economic Relationships and the Effects of Differencing," *Journal of Monetary Economics,* vol. 4 (November 1978), pp. 637–60.

__________. "Estimation of a Noninvertible Moving Average Process: The Case of Overdifferencing," *Journal of Econometrics,* vol. 6 (August 1977), pp. 199–224.

Rudebusch, Glenn D. "The Uncertain Unit Root in Real GNP," *American Economic Review,* vol. 83 (March 1993), pp. 264–72.

Sims, Christopher A. "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control,* vol. 12 (June/September 1988), pp. 463–74.

Sowell, Fallaw. "Modelling Long-Run Behavior with the Fractional ARIMA Model," *Journal of Monetary Economics,* vol. 29 (April 1992), pp. 277–302.

Stock, James H. "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics,* vol. 28 (December 1991), pp. 435–59.

__________, and Mark W. Watson. "Variable Trends in Economic Time Series, *Journal of Economic Perspectives,* vol. 2 (Summer 1988), pp. 147–74.

Stockman, Alan C. "The Equilibrium Approach to Exchange Rates," Federal Reserve Bank of Richmond *Economic Review,* vol. 73 (March/April 1987), pp. 12–30.

Taylor, Mark P. "An Empirical Examination of Long-Run Purchasing Power Parity Using Cointegration Techniques," *Applied Economics,* vol. 20 (October 1988), pp. 1369–81.

Watson, Mark W. "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics,* vol. 18 (July 1986), pp. 49–75.

West, Kenneth D. "On the Interpretation of the near Random-Walk Behavior in GNP," *American Economic Review,* vol. 78 (March 1988), pp. 202–11.

Commercial Paper

Thomas K. Hahn

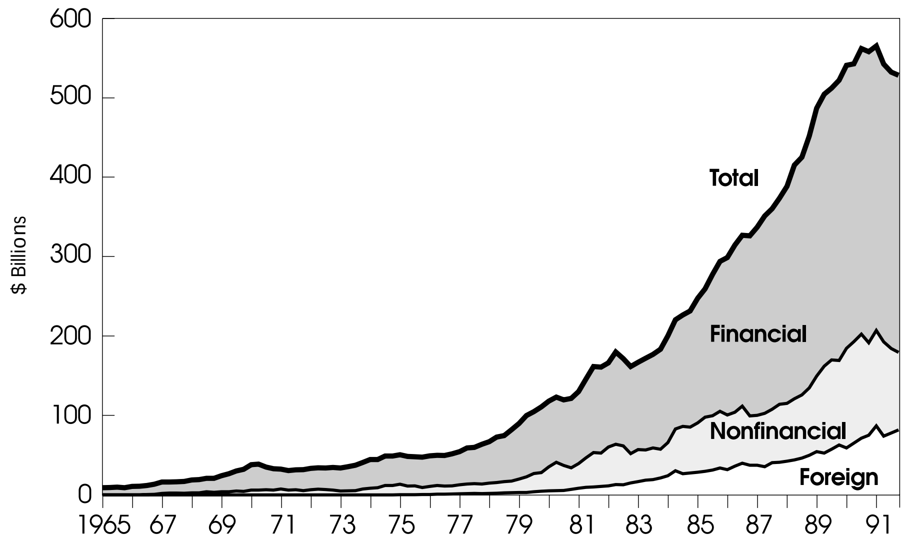
Commercial paper is a short-term unsecured promissory note issued by corporations and foreign governments. For many large, creditworthy issuers, commercial paper is a low-cost alternative to bank loans. Issuers are able to efficiently raise large amounts of funds quickly and without expensive Securities and Exchange Commission (SEC) registration by selling paper, either directly or through independent dealers, to a large and varied pool of institutional buyers. Investors in commercial paper earn competitive, market-determined yields in notes whose maturity and amounts can be tailored to their specific needs.

Because of the advantages of commercial paper for both investors and issuers, commercial paper has become one of America's most important debt markets. Commercial paper outstanding grew at an annual rate of 14 percent from 1970 to 1991. Figure 1 shows commercial paper outstanding, which totaled $528 billion at the end of 1991.

This article describes some of the important features of the commercial paper market. The first section reviews the characteristics of commercial paper. The second section describes the major participants in the market, including the issuers, investors, and dealers. The third section discusses the risks faced by investors in the commercial paper market along with the mechanisms that are used to control these risks. The fourth section discusses some recent innovations, including asset-backed commercial paper, the use of swaps in commercial paper financing strategies, and the international commercial paper markets.

The author, a consultant with TKH Associates and former assistant economist at the Federal Reserve Bank of Richmond, would like to thank Timothy Cook, Bob LaRoche, Jerome Fons, and Mitchell Post for comments. The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

Figure 1 Commercial Paper Outstanding



Source: Board of Governors of the Federal Reserve System.

1. CHARACTERISTICS OF COMMERCIAL PAPER

The Securities Act of 1933 requires that securities offered to the public be registered with the Securities and Exchange Commission. Registration requires extensive public disclosure, including issuing a prospectus on the offering, and is a time-consuming and expensive process.[1] Most commercial paper is issued under Section 3(a)(3) of the 1933 Act which exempts from registration requirements short-term securities as long as they have certain characteristics.[2] The exemption requirements have been a factor shaping the characteristics of the commercial paper market.

[1] Registration for short-term securities is especially expensive because the registration fee is a percent of the face amount at each offering. Thirty-day registered notes, rolled over monthly for one year, would cost 12 times as much as a one-time issuance of an equal amount of one-year notes.

[2] Some commercial paper is issued under one of the two other exemptions to the Securities Act. Commercial paper which is guaranteed by a bank through a letter of credit is exempt under Section 3(a)(2) regardless of whether or not the issue is also exempt under Section 3(a)(3). Commercial paper sold through private placements is exempt under Section 4(2). See Felix (1987) for more information on the legal aspects of commercial paper issuance.

One requirement for exemption is that the maturity of commercial paper must be less than 270 days. In practice, most commercial paper has a maturity of between 5 and 45 days, with 30–35 days being the average maturity. Many issuers continuously roll over their commercial paper, financing a more-or-less constant amount of their assets using commercial paper. Continuous rollover of notes does not violate the nine-month maturity limit as long as the rollover is not automatic but is at the discretion of the issuer and the dealer. Many issuers will adjust the maturity of commercial paper to suit the requirements of an investor.

A second requirement for exemption is that notes must be of a type not ordinarily purchased by the general public. In practice, the denomination of commercial paper is large: minimum denominations are usually $100,000, although face amounts as low as $10,000 are available from some issuers. Because most investors are institutions, typical face amounts are in multiples of $1 million. Issuers will usually sell an investor the specific amount of commercial paper needed.

A third requirement for exemption is that proceeds from commercial paper issues be used to finance "current transactions," which include the funding of operating expenses and the funding of current assets such as receivables and inventories. Proceeds cannot be used to finance fixed assets, such as plant and equipment, on a permanent basis. The SEC has generally interpreted the current transaction requirement broadly, approving a variety of short-term uses for commercial paper proceeds. Proceeds are not traced directly from issue to use, so firms are required to show only that they have a sufficient "current transaction" capacity to justify the size of the commercial paper program (for example, a particular level of receivables or inventory).[3] Firms are allowed to finance construction as long as the commercial paper financing is temporary and to be paid off shortly after completion of construction with long-term funding through a bond issue, bank loan, or internally generated cash flow.[4]

Like Treasury bills, commercial paper is typically a discount security: the investor purchases notes at less than face value and receives the face value at maturity. The difference between the purchase price and the face value, called the discount, is the interest received on the investment. Occasionally, investors request that paper be issued as an interest-bearing note. The investor pays the

---

[3] Some SEC interpretations of the current transaction requirement have been established in "no-action" letters. "No-action" letters, issued by the staff of the SEC at the request of issuers, confirm that the staff will not request any legal action concerning an unregistered issue. See Felix (1987, p. 39).

[4] Past SEC interpretations of Section 3(a)(3) exemptions have also required that commercial paper be of "prime quality" and be discountable at a Federal Reserve Bank (Release No. 33-4412). The discounting requirement was dropped in 1980. An increased amount of commercial paper in the later 1980s was issued without prime ratings.

face value and, at maturity, receives the face value and accrued interest. All commercial paper interest rates are quoted on a discount basis.[5]

Until the 1980s, most commercial paper was issued in physical form in which the obligation of the issuer to pay the face amount at maturity is recorded by printed certificates that are issued to the investor in exchange for funds. The certificates are held, usually by a safekeeping agent hired by the investor, until presented for payment at maturity. The exchanges of funds for commercial paper first at issuance and then at redemption, called "settling" of the transaction, occur in one day. On the day the commercial paper is issued and sold, the investor receives and pays for the notes and the issuer receives the proceeds. On the day of maturity, the investor presents the notes and receives payment. Commercial banks, in their role as issuing, paying, and clearing agents, facilitate the settling of commercial paper by carrying out the exchanges between issuer, investor, and dealer required to transfer commercial paper for funds.

An increasing amount of commercial paper is being issued in book-entry form in which the physical commercial paper certificates are replaced by entries in computerized accounts. Book-entry systems will eventually completely replace the physical printing and delivery of notes. The Depository Trust Company (DTC), a clearing cooperative operated by member banks, began plans in September 1990 to convert most commercial paper transactions to book-entry form.[6] By May 1992, more than 40 percent of commercial paper was issued through the DTC in book-entry form.

The advantages of a paperless system are significant. The fees and costs associated with the book-entry system will, in the long run, be significantly less than under the physical delivery system. The expense of delivering and verifying certificates and the risks of messengers failing to deliver certificates on time will be eliminated. The problem of daylight overdrafts, which arise from nonsynchronous issuing and redeeming of commercial paper, will be reduced since all transactions between an issuing agent and a paying agent will be settled with a single end-of-day wire transaction.

2. MARKET PARTICIPANTS

Issuers and Uses of Commercial Paper

Commercial paper is issued by a wide variety of domestic and foreign firms, including financial companies, banks, and industrial firms. Table 1 shows

[5] The Federal Reserve publishes in its H.15 statistical release daily interest rates for dealer-offered and directly placed commercial paper of one-month, three-month and six-month maturities. All rates are based on paper with relatively low default risk. Commercial paper rates of various maturities for select finance issuers and a dealer composite rate are also published daily in *The Wall Street Journal.*

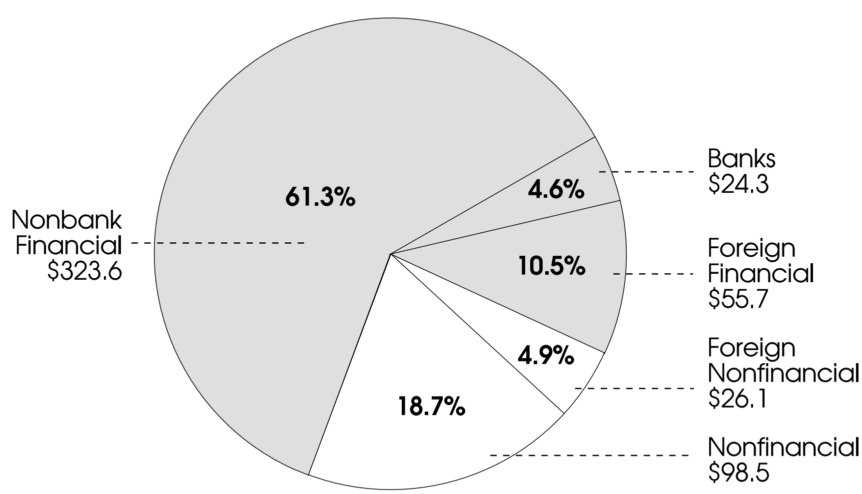
[6] See The Depository Trust Company (1990).

Table 1 Commercial Paper Outstanding by Major Issuer
Billions of dollars

Category	**Major Issuer**	**Average Amount Outstanding**	**Dealer**
Finance	General Electric Capital (subsidiary of GE)	$36.9	Direct, Multiple
Auto Finance	General Motors Acceptance (subsidiary of GM)	$23.6	Direct
Investment Banking	Merrill Lynch	$ 7.5	Dealer is subsidiary
Commercial Banking	J.P. Morgan	$ 4.4	Multiple
Industrial	PepsiCo	$ 3.4	Multiple
Foreign	Hanson Finance	$ 3.5	Multiple
Asset-Backed	Corporate Asset Funding	$ 5.3	Goldman Sachs

Note: Quarterly Average Commercial Paper is for the first quarter of 1992, except GE, GMAC, and PepsiCo, which are for the fourth quarter of 1991.
Source: *Moody's Global Short Term Record*, June 1992.

Figure 2 Commercial Paper Outstanding by Issuer Type
End of 1991 Total $528.1 Billion
Billions of dollars



Source: Board of Governors of the Federal Reserve System.

examples of the largest commercial paper issuers. Figure 2 shows outstanding commercial paper by type of issuer.

The biggest issuers in the financial firm category in Figure 2 are finance companies. Finance companies provide consumers with home loans, retail automobile loans, and unsecured personal loans. They provide businesses with a variety of short- and medium-term loans including secured loans to finance purchases of equipment for resale. Some finance companies are wholly owned subsidiaries of industrial firms that provide financing for purchases of the parent firm's products. For example, a major activity of General Motors Acceptance Corporation (GMAC) is the financing of purchases and leases of General Motor's vehicles by dealers and consumers. The three largest issuers—GMAC, General Electric Capital, and Ford Motor Credit—accounted for more than 20 percent of the total nonbank financial paper outstanding at the end of 1991.

The financial issuer category also includes insurance firms and securities firms. Insurance companies issue commercial paper to finance premium receivables and operating expenses. Securities firms issue commercial paper as a low-cost alternative to other short-term borrowings such as repurchase agreements and bank loans, and they use commercial paper proceeds to finance a variety of security broker and investment banking activities.

Commercial bank holding companies issue commercial paper to finance operating expenses and various nonbank activities. Bank holding companies have recently decreased their commercial paper issues following declines in the perceived creditworthiness of many major domestic bank issuers.

More than 500 nonfinancial firms also issue commercial paper. Nonfinancial issuers include public utilities, industrial and service companies. Industrial and service companies use commercial paper to finance working capital (accounts receivable and inventory) on a permanent or seasonal basis, to fund operating expenses, and to finance, on a temporary basis, construction projects. Public utilities also use commercial paper to fund nuclear fuels and construction. Figure 3 shows that commercial paper as a percent of commercial paper and bank loans for nonfinancial firms rose from just 2 percent in 1966 to over 15 percent at the end of 1991.

The domestic commercial paper issuers discussed above include U.S. subsidiaries of foreign companies. Foreign corporations and governments also issue commercial paper in the U.S. without use of a domestic subsidiary and these foreign issues have gained increased acceptance by U.S. investors. Foreign financial firms, including banks and bank holding companies, issue almost 70 percent of foreign commercial paper (Federal Reserve Bank of New York 1992). Industrial firms and governments issue the remainder. Japan, the United Kingdom, and France are among the countries with a significant number of issuers.

Investors

Money market mutual funds (MMFs) and commercial bank trust departments are the major investors in commercial paper. MMFs hold about one-third of the outstanding commercial paper, while bank trust departments hold between 15

Figure 3 Commercial Paper as a Percent of Commercial Paper and Bank Loans, Nonfinancial Firms

Source: Board of Governors of the Federal Reserve System.

and 25 percent.[7] Other important investors, holding between 5 and 15 percent, are nonfinancial corporations, life insurance companies, and private and government pension funds. Other mutual funds, securities dealers, and banks also hold small amounts of commercial paper. Individuals hold little commercial paper directly because of the large minimum denominations, but they are large indirect investors in commercial paper through MMFs and trusts.

There have been major shifts in ownership of commercial paper during the post-World War II period. Prior to World War II, the most important investors in commercial paper were banks, which used commercial paper as a reserve asset and to diversify their securities portfolios. In the fifties and sixties, industrial firms began to hold commercial paper as an alternative to bank deposits, which had regulated interest rates that at times were significantly below the market-determined rates on commercial paper. Historically high and variable interest rates during the 1970s led households and businesses to hold more

[7] Precise data on holdings of commercial paper by investor type, except by MMFs, are not available. Some estimates are provided in Board of Governors of the Federal Reserve System (1992, p. 52), Stigum (1990, p. 1027), and Felix (1987, p. 13).

Table 2 Money Market Mutual Funds and Commercial Paper

End of	MMF Assets ($ billions)	Commercial Paper Outstanding ($ billions)	MMF Holdings of CP ($ billions)	CP as Percent of MMF Assets	Percent of CP Held by MMFs
1975	3.7	47.7	0.4	11	1
1980	74.5	121.6	25.0	33	21
1985	207.5	293.9	87.6	42	30
1990	414.8	557.8	199.1	48	36
1991	449.7	528.1	187.6	42	36

Note: MMFs exclude tax-exempt funds.
Source: Board of Governors of the Federal Reserve System.

of their funds in short-term assets and to transfer funds from bank deposits with regulated interest rates to assets like MMF shares with market-determined rates. At the same time, many large businesses found that they could borrow in the commercial paper market at less expense than they could borrow from banks. MMFs demanded the short-term, large-denomination, relatively safe, and high-yield characteristics offered by commercial paper and hence absorbed a major portion of new commercial paper issues. Table 2 shows that both the commercial paper market and MMFs have experienced very rapid growth since 1975. By the end of 1991, MMFs held 36 percent of the commercial paper outstanding and commercial paper composed 42 percent of their total assets.

Placement and Role of the Dealer

Most firms place their paper through dealers who, acting as principals, purchase commercial paper from issuers and resell it to the public. Most dealers are subsidiaries of investment banks or commercial bank holding companies. A select group of very large, active issuers, called direct issuers, employ their own sales forces to distribute their paper. There are approximately 125 direct issuers, most of which are finance companies or bank holding companies. These issuers sell significant amounts of commercial paper on a continuous basis.

When an issuer places its commercial paper through a dealer, the issuer decides how much paper it will issue at each maturity. The dealer is the issuer's contact with investors and provides the issuer with relevant information on market conditions and investor demand. Dealers generally immediately resell commercial paper purchased from issuers and do not hold significant amounts of commercial paper in inventory. Dealers will temporarily hold commercial paper in inventory as a service to issuers, such as to meet an immediate need for a significant amount of funds at a particular maturity.

The difference between what the dealer pays the issuer for commercial paper and what he sells it for, the "dealer spread," is around 10 basis points on an annual basis. A large commercial paper program with $500 million in paper outstanding for one year would cost the issuer $500,000 in dealer fees.

Because independent dealers are relatively inexpensive, only large and well-recognized issuers distribute their own commercial paper. Direct issuers are typically committed to borrowing $1 billion or more in the commercial paper market on a continuous basis (Felix 1987, p. 20). Partly as a result of the decline in dealer spreads over the last ten years, the percentage of total commercial paper issued directly fell from almost 55 percent in 1980 to just 35 percent at the end of 1991. An additional factor in the growth of dealer-placed commercial paper has been the entry into the market of smaller issuers who do not have borrowing needs large enough to justify a direct sales force.

Competition among dealers significantly increased in the late 1980s after the entrance into the market of bank dealers, which are subsidiaries of bank holding companies. Prior to the mid-1980s, commercial banks mainly acted as agents who placed commercial paper without underwriting and who carried out the physical transactions required in commercial paper programs, including the issuing and safekeeping of notes and the paying of investors at maturity. Bank dealers entered the market after legal restrictions on underwriting by bank holding companies were relaxed, and the increased competition led to declines in profit margins and the exit from the market of some major investment bank dealers. Salomon Brothers closed its dealership and Paine Webber sold its dealership to CitiCorp. Goldman Sachs, another important dealer, responded to increased competition by rescinding its longstanding requirement that it be the sole dealer for an issuer's commercial paper. Issuers have increased their use of multiple dealers for large commercial paper programs, frequently including a bank dealer in their team of dealers.

The largest commercial paper dealers are still the investment banks, including Merrill Lynch, Goldman Sachs, and Shearson Lehman. Commercial bank holding companies with large commercial paper dealer subsidiaries include Bankers Trust, CitiCorp, BankAmerica, and J.P. Morgan. Some foreign investment and commercial bank holding companies have also become significant dealers.

The secondary market in commercial paper is small. Partly the lack of a secondary market reflects the heterogeneous characteristics of commercial paper, which makes it difficult to assemble blocks of paper large enough to facilitate secondary trading. Partly it reflects the short maturity of the paper: investors know how long they want to invest cash and, barring some unforseen cash need, hold commercial paper to maturity. Dealers will sometimes purchase paper from issuers or investors, hold the paper in inventory and subsequently trade it. Bids for commercial paper of the largest issuers are available through brokers.

Some direct issuers offer master note agreements which allow investors, usually bank trust departments, to lend funds on demand on a daily basis at a rate tied to the commercial paper rate. Each day the issuer tells the investor the rate on the master note and the investor tells the issuer how much it will deposit that day. At the end of 1991, approximately 10 percent of GMAC's short-term notes outstanding were master notes sold to bank trust departments (GMAC 1992, p. 13).

3. RISK IN THE COMMERCIAL PAPER MARKET

Ratings

Since 1970, when the Penn Central Transportation Co. defaulted with $82 million of commercial paper outstanding, almost all commercial paper has carried ratings from one or more rating agency. Currently, the four major rating agencies are Moody's, Standard & Poor's, Duff & Phelps, and Fitch. An issuer's commercial paper rating is an independent "assessment of the likelihood of timely payment of [short-term] debt" (Standard & Poor's 1991, p. iii). Table 3 lists the four rating agencies, the rating scales they publish, and the approximate number of commercial paper ratings issued at the end of 1990. The ratings are relative, allowing the investor to compare the risks across issues. For example, Standard & Poor's gives an A-1 rating to issues that it believes have a "strong" degree of safety for timely repayment of debt, an A-2 rating to issues that it believes have a degree of safety that is "satisfactory," and an A-3 rating to issues that it believes have a degree of safety that is "adequate." Below these

Table 3 Rating Agencies and Commercial Paper Ratings

	Higher A/Prime	Lower A/Prime	Speculative Below Prime	Defaulted	Approx. # of CP Ratings	Major Publication Listing CP Ratings
Moody's	P-1	P-2, P-3	NP	NP	2,000	Moody's Global Short-Term Market Record
Standard & Poor's	A-1+, A-1	A-2, A-3	B, C	D	2,000	S&P Commercial Paper Ratings Guide
Duff & Phelps	Duff 1+, Duff 1, Duff 1-	Duff 2, Duff 3	Duff 4	Duff 5	175	Short-Term Ratings and Research Guide
Fitch	F-1+, F-1	F-2, F-3	F-5	D	125	Fitch Ratings
Range of Likely S&P Long-Term Bond Rating	AAA, AA, A	A, BBB	BB, B, CCC, CC, C			

three categories are the speculative grades in which the capacity for repayment is small relative to the higher-rated issues. Finally, a D rating indicates the issuer has defaulted on its commercial paper. Almost all issuers carry one of the two highest Prime or A ratings.

Issuers hire the rating agencies to rate their short-term debt and pay the agencies an annual fee ranging from $10,000 to $29,000 per year. For an additional fee the agencies will also rate other liabilities of the issuer, including their long-term bonds. The ratings are provided to the public, generally by subscription, either through publications, computer databases, or over the phone. Major announcements by the rating agencies are also reported on news wire services. Table 3 lists each agency's major publication in which commercial paper ratings appear.

Rating agencies rely on a wide variety of information in assessing the default risk of an issuer. The analysis is largely based on the firm's historical and projected operating results and its financial structure. Relevant characteristics include size (both absolute and compared to competitors), profitability (including the level and variation of profits), and leverage. Table 4 shows the means of selected historical characteristics of a sample of publicly traded nonfinancial issuers by commercial paper rating category. The table shows that higher-rated issuers are on average more profitable than lower-rated issuers and, with some exceptions, larger. Additionally, higher-rated issuers rely less heavily on debt financing than lower-rated issuers and have stronger interest-coverage and

Table 4 Characteristics of Industrial Commercial Paper Issuers by Rating, Three-Year Averages

Standard & Poor's Commercial Paper Rating	Number of Companies	Assets (millions)	Interest Coverage	Debt Coverage	Leverage	Profitability
A-1+	91	$4,547	8x	.7x	27%	18%
A-1	102	$2,924	5x	.5x	35%	16%
A-2	97	$1,866	4x	.4x	36%	14%
A-3	9	$5,252	2x	.2x	52%	10%

Notes: Sample consists of nonfinancial commercial paper issuers required to file with the SEC.

Interest coverage is defined as the ratio of income available for interest to interest expense. Income available for interest is defined as pre-tax income less special income plus interest expense.

Debt coverage is defined as the ratio of cash flow to short- and long-term debt. Cash flow is income plus preferred dividends plus deferred taxes.

Leverage is defined as the ratio of total debt to invested capital. Invested capital is the sum of short- and long-term debt, minority interest, preferred and common equity, and deferred taxes.

Profitability is defined as the ratio of income available for interest to invested capital.

Source: Standard & Poor's Compustat Services.

debt-coverage ratios.[8] In addition to evaluating the firm's operating results and financial structure, rating agencies also evaluate more subjective criteria like quality of management and industry characteristics. The same factors influence the issuer's short-term and long-term debt rating so there is generally a close correspondence between the commercial paper rating and the bond rating.

Ratings are crucially important in the commercial paper market. Ratings are useful as an independent evaluation of credit risk that summarizes available public information and reduces the duplication of analysis in a market with many investors (Wakeman 1981). Ratings are also used to guide investments in commercial paper. Some investors, either by regulation or choice, restrict their holdings to high-quality paper and the measure of quality used for these investment decisions is the rating. For example, regulations of MMFs limit their holdings of commercial paper rated less than A1-P1. Other market participants, including dealers and clearing agencies, also generally require issuers to maintain a certain quality. Again, credit quality is measured by the rating.

Backup Liquidity

Commercial paper issuers maintain access to funds that can be used to pay off all or some of their maturing commercial paper and other short-term debt. These funds are either in the form of their own cash reserves or bank lines of credit. Rating agencies require evidence of short-term liquidity and will not issue a commercial paper rating without it. The highest-rated issuers can maintain liquidity backup of as little as 50 percent of commercial paper outstanding, but firms with less than a high A1-P1 rating generally have to maintain 100 percent backup.

Most commercial paper issuers maintain backup liquidity through bank lines of credit available in a variety of forms. Standard credit lines allow borrowing under a 90-day note. Swing lines provide funds on a day-to-day basis, allowing issuers to cover a shortfall in proceeds from paper issuance on a particular day. Increasingly, backup lines of credit are being structured as more secure multi-year revolver agreements in which a bank or syndicate of banks commit to loan funds to a firm on demand at a floating base rate that is tied to the prime rate, LIBOR rate, or Certificate of Deposit rate. The spread over the base rate is negotiated at the time the agreement is made and can either be fixed or dependent on the bond rating of the borrower at the time the loan is drawn down. The length of the revolver commitment varies, but the trend in revolvers has been towards shorter terms, typically around three years. As compensation for the revolver commitment, the firm pays various fees to the bank. The facility fee is a percentage of the credit line and is paid whether or

[8] Because ratings depend on historical operating results, researchers have had some success in predicting ratings based on accounting data. See, for example, Peavy and Edgar (1983).

not the line is activated. The commitment fee is a percentage of the unused credit line. This type of fee has become less common in recent years. A usage fee is sometimes charged if the credit line is heavily used.

Backup lines of credit are intended to provide funds to retire maturing commercial paper when an event prevents an issuer from rolling over the paper. Such an event may be specific to an issuer: an industrial accident, sudden liability exposure, or other adverse business conditions that investors perceive as significantly weakening the credit strength of the issuer. Or the event may be a general development affecting the commercial paper market. For instance, a major issuer might default, as Penn Central did in 1970, and make it prohibitively expensive for some issuers to roll over new paper, or a natural disaster such as a hurricane may interrupt the normal function of the market.

Backup lines of credit will generally not be useful for a firm whose operating and financial condition has deteriorated to the point where it is about to default on its short-term liabilities. Credit agreements frequently contain "material adverse change" clauses which allow banks to cancel credit lines if the financial condition of a firm significantly changes. Indeed, the recent history of commercial paper defaults has shown that as an issuer's financial condition deteriorates and its commercial paper cannot be rolled over, backup lines of credit are usually canceled before they can be used to pay off maturing commercial paper.

General factors affecting the commercial paper market may also result in the disruption of backup lines of credit. Standard & Poor's has emphasized this point in an evaluation of the benefits to investors of backup credit lines: "A general disruption of commercial paper markets would be a highly volatile scenario, under which most bank lines would represent unreliable claims on whatever cash would be made available through the banking system to support the market" (Samson and Bachmann 1990, p. 23). Part of the risk assumed by commercial paper investors is the possibility of this highly volatile scenario.

Credit Enhancements

While backup lines of credit are needed to obtain a commercial paper rating, they will not raise the rating above the underlying creditworthiness of the issuer. Issuers can significantly increase the rating of their paper, however, by using one of a variety of credit enhancements which lower default risk by arranging for an alternative party to retire the commercial paper if the issuer cannot. These credit enhancements differ from backup lines of credit in that they provide a guarantee of support which cannot be withdrawn. Some smaller and riskier firms, which normally would find the commercial paper market unreceptive, access the commercial paper market using these enhancements.

Some large firms with strong credit ratings raise the ratings of smaller and less creditworthy subsidiaries by supporting their commercial paper with outright guarantees or with less secure "keepwell" agreements which describe the commitment the parent makes to assist the subsidiary to maintain a certain creditworthiness (Moody's, July 1992). Since parent companies may have incentives to prevent default by their subsidiaries, the affiliation of a subsidiary with a strong parent can raise the credit rating of the subsidiary issuer.

Firms also raise their credit ratings by purchasing indemnity bonds from insurance companies or standby letters of credit sold by commercial banks. Both of these enhancements provide assurance that the supporting entity will retire maturing commercial paper if the issuer cannot. With a letter of credit, for example, the issuer pays a fee to the bank, attaches the letter of credit to the commercial paper and effectively rents the bank's rating. The attention of the rating agency and investors shift from the issuer to the supporting bank. The issue will generally receive the same rating as the bank's own commercial paper and offer an interest rate close to the bank's paper. Since relatively few U.S. banks have A1-P1 ratings, highly rated foreign banks are the primary sellers of commercial paper letters of credit. At the end of the first quarter of 1992, approximately 6 percent of commercial paper was fully backed by a credit enhancement, primarily bank letters of credit, issued by a third party unaffiliated with the issuer (Federal Reserve Bank of New York 1992).

Slovin et al. (1988) show that the announcement of a commercial paper program with a credit enhancement[9] has been associated with a significant increase in the value of the issuer's equity, but the announcement of a commercial paper program with no credit enhancement has no impact on firm value. This evidence suggests that by issuing a letter of credit and certifying the creditworthiness of the issuer, the commercial bank provides new information to the capital markets. These results provide support for the hypothesis that banks generate information relevant for assessing credit risk that the securities markets do not have. Banks supply this information to the capital market through commercial paper programs supported by letters of credit.

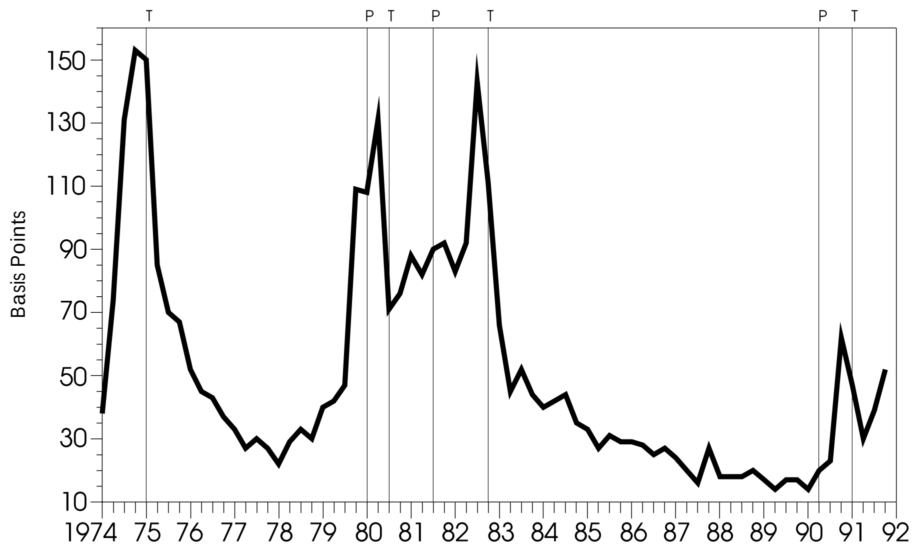
Default History and Yields

Commercial paper pays a market-determined interest rate that is closely related to other market interest rates like the rate on large certificates of deposit. Because commercial paper has default risk, its yield is higher than the yield on Treasury bills. From 1967 through 1991, the spread of the one-month commercial paper rate over the one-month Treasury bill rate averaged 117 basis points.

Default risk also creates a differential between the rates on different quality grades of commercial paper. Figure 4 shows the spread between the yield on commercial paper rated A1-P1 and the yield on paper rated A2-P2. This spread

[9] The credit enhancements examined were standby letters of credit and, for programs outside the United States, note issuance facilities.

Figure 4 Spread Between the Rates on Prime- and Medium-Grade Commercial Paper



Source: Board of Governors of the Federal Reserve System.

averaged 52 basis points from 1974 through 1991. Default risk as measured by the quality spread shows some variation over time, rising during recessions and falling during expansions.

Historically, the commercial paper market has been remarkably free of default. As shown in Table 5, in the 20-year period from 1969 through 1988 there were only two major defaults. The low default rates in the commercial paper market largely reflect the tastes of commercial paper investors. As shown in Table 4, investors typically prefer commercial paper issued by large firms with long track records, conservative financing strategies, and stable profitability. Most investors will not buy paper from small, unknown, highly leveraged issuers unless the paper has credit enhancements attached. Moreover, rating services will not assign a prime rating to these issues and most dealers will not distribute the paper.

Even a major issuer can find the commercial paper market unreceptive if its financial condition is perceived by the market to have weakened. Fons and Kimball (1992) estimate that issuers who defaulted on long-term debt withdrew from the commercial paper market an average of almost three years prior to default. As ratings declined, these issuers significantly decreased their commercial paper borrowings. Fons and Kimball (1992) take this "orderly exit"

Table 5 Major Defaults in the U.S. Commercial Paper Market

Issuer	Date of Default	Amount Outstanding at Default ($ millions)	Original Rating of Longest Outstanding Defaulting CP Moody's	S&P
Penn Central	6/21/70	82.0	NR	NR
Manville Corp.	8/26/82	15.2	P-2	A-2
Integrated Resources	6/15/89	213.0	NR	A-2
Colorado Ute Electric	8/17/89	19.0	P-1	A-1
Equitable Lomas Leasing	9/12/89	53.0	P-3	A-3
Mortgage & Realty Trust	3/15/90	166.9	NR	A-2
Washington Bancorp	5/11/90	36.7	NR	NR
Stotler Group	7/25/90	0.75	NR	NR
Columbia Gas	6/12/91	268.0	P-2	A-2

Source: Fons and Kimball (1992), *Wall Street Journal, Dow Jones News Wire, Business Week,* Standard & Poor's.

mechanism as evidence that investors in the commercial paper market are "unreceptive to lower-quality paper." Crabbe and Post (January 1992) document the orderly exit mechanism using a sample of bank holding company issuers during 1986 to 1990. For issuers which experienced Moody's commercial paper rating downgrades, commercial paper outstanding declined on average by 12.2 percent in the ten weeks prior to the rating change and 15.7 percent in the first four weeks after the change.

The number of commercial paper defaults rose to seven in 1989 to 1991, but even in this period the default rate was low. Fons and Kimball (1992) estimate the dollar amount of defaults over this period as a percentage of the total volume issued.[10] They find that the default rate for the United States was only 0.0040 percent in 1989–91, which means that "an investor purchasing U.S.-issued commercial paper:. throughout the 1989–1991 period experienced, on average, interruption in promised payments of roughly [40/100] of a penny for every $100 invested" (p. 13).

The rise in defaults in the 1989 to 1990 period may have partially reflected an increased tolerance for riskier paper in the later part of the 1980s. Unrated commercial paper grew significantly in the late 1980s to $5 billion in January 1990. Over the same period, the spread between the yields on A1-P1 paper and A2-P2 paper was unusually low (averaging less than 30 basis points). These

[10] Fons and Kimball (1992) estimate the total volume of commercial paper issuance as average outstanding commercial paper times (365/average maturity). Average maturity is estimated at 30 days.

developments were reversed in the early 1990s following the rise in commercial paper defaults, the deterioration in economic conditions, and the bankruptcy of Drexel Burnham, a major dealer and promoter of unrated commercial paper. By early 1991, unrated paper outstanding had fallen to below $1 billion and the A1-A2 spread had risen to almost 50 basis points, its highest level since 1982.

The commercial paper defaults in 1989 and 1990 had a significant impact on the demand for lower-rated paper by money market mutual funds. Several MMFs were major holders of defaulted paper of Integrated Resources and Mortgage & Realty Trust.[11] Following these defaults, some MMFs began to voluntarily restrict their commercial paper holdings to A1-P1 issues. Then in June 1991, SEC regulations became effective that limited MMFs to investing no more than one percent of their assets in any single A2-P2 issuer and no more than 5 percent of assets in A2-P2 paper. Previously, there had been no restriction on MMF total holding of A2-P2 paper, and MMFs had held approximately 10 percent of their assets in A2-P2 paper at the end of 1990. Crabbe and Post (May 1992) find that by the end of 1991, MMFs had reduced their holdings of A2-P2 commercial paper to almost zero. Along with the 1989 and 1990 defaults, they point to the June 1991 regulations as an important factor influencing MMF investment choices.

4. INNOVATIONS

Asset-Backed Commercial Paper

A relatively new innovation in the commercial paper market is the backing of commercial paper with assets. The risk of most commercial paper depends on the entire firm's operating and financial risk. With asset-backed paper, the paper's risk is instead tied directly to the creditworthiness of specific financial assets, usually some form of receivable. Asset-backed paper is one way smaller, riskier firms can access the commercial paper market. The advantages of asset-backed securities have led large, lower-risk commercial paper issuers to also participate in asset-backed commercial paper programs. Asset-backed programs have grown rapidly since the first program in 1983. Standard & Poor's has rated more than 60 asset-backed issues (Kavanagh et al. 1992, p. 109) with an estimated $40 billion outstanding.

Asset-backed commercial paper is issued by a company, called a special purpose entity, which purchases receivables from one firm or a group of firms and finances the purchase with funds raised in the commercial paper market. The sole business activity of the special company is the purchase and finance of

[11] Value Line's MMF, for example, held 3.5 percent of its portfolio in $22.6 million of Integrated's paper. Value Line protected the fund's investors, absorbing the loss at an after-tax cost of $7.5 million.

the receivables so the risk of the company and the commercial paper it issues is isolated from the risk of the firm or firms which originated the receivables.

The trade receivables and credit card receivables that are typically used in asset-backed programs have a predictable cash flow and default rate so the risk of the assets can be estimated. Asset-backed paper programs are structured so that the amount of receivables exceeds the outstanding paper. In addition to this over-collaterization, credit enhancements are used, including guarantees by the firm selling the receivables, bank letters of credit, or surety bonds. As with all commercial paper issues, rating agencies require backup liquidity.

The combining of similar receivables from a group of companies into a pool large enough to justify a commercial paper program allows small firms to participate in asset-backed programs and serves to diversify some of the receivables' default risk. Typically, the financing firm which pools the receivables is managed by a commercial bank which purchases assets from its corporate clients.

Swaps

A factor in the growth of the commercial paper market during the 1980s has been the rapid growth in the market for interest rate swaps. Interest rate swaps are one of a variety of relatively new instruments that have significantly increased the financing options of commercial paper issuers. Swaps provide issuers with flexibility to rapidly restructure their liabilities, to raise funds at reduced costs, and to hedge risks arising from short-term financing programs.

Interest rate swaps are agreements between two parties to exchange interest rate payments over some specified time period on a certain amount of unexchanged principle. To appreciate the role of swaps it is necessary to understand that there are two interest rate risks associated with commercial paper borrowing. First, the firm faces market interest rate risk: the risk that the rate it pays on commercial paper will rise because the level of market interest rates increases. A change in the risk-free rate, such as the Treasury bill rate, will cause a corresponding change in all commercial paper and borrowing rates. Second, the firm faces idiosyncratic interest rate risk: the risk that commercial paper investors will demand a higher rate because they perceive the firm's credit risk to have increased. With idiosyncratic risk, the rate on its commercial paper can rise without an increase in the risk-free rate or in other commercial paper rates.

A commercial paper issuer can eliminate market interest rate risk by entering into a swap and agreeing to exchange a fixed interest rate payment for a variable interest rate. For example, in the swap the firm may pay a fixed interest rate that is some spread over the multi-year Treasury bond rate and receive the floating six-month LIBOR rate. If the commercial paper rate rises because of a general rise in the market interest rate, the firm's increased interest payment on its commercial paper is offset by the increased payment it receives from the swap. This swap allows the firm to transform its short-term, variable-rate commercial paper financing into a fixed-rate liability that hedges market interest rate risks in the same manner as long-term fixed-rate, noncallable debt. Note that the firm still bears the risk of idiosyncratic changes in its commercial paper rate. If its own commercial paper rate rises while other rates, including the LIBOR rate, do not rise, the cost of borrowing in the commercial paper market will rise without a corresponding increase in the payment from the swap.

Alternatively, the firm can fix the cost of its idiosyncratic risk by borrowing in the long-term market at a fixed rate and entering into a swap in which it pays a floating rate and receives a fixed rate. The swap effectively converts the long-term fixed-rate liability into a floating-rate liability that is similar to commercial paper. The firm now faces the risk of a general change in the level on interest rates, just like a financing strategy of issuing commercial paper, but has fixed the cost of its idiosyncratic risk by borrowing long-term in the bond market at a fixed-rate.

One important and unresolved issue is what the advantage of swaps are relative to alternative financing strategies. For example, why would a firm issue short-term debt and swap the flexible rate into a long-term rate instead of issuing long-term debt? Researchers have advanced a variety of hypotheses to explain the rapid growth of the interest rate swap market, but no real consensus has been reached. Many explanations view swaps as a way for firms to exploit differences in the premium for credit risk at different maturities and in different markets. For example, one firm may find it can issue commercial paper at a rate close to the average for similarly rated issuers but pays a significantly higher spread in the long-term fixed-rate market. If the firm prefers fixed-rate financing, a commercial paper program combined with a swap may provide cheaper financing than issuing fixed-rate debt. But it is uncertain what causes these borrowing differentials.[12]

The two interest rate swaps discussed above are the most basic examples of a wide variety of available swaps. The examples are constructed to highlight some important aspects of interest rate swaps, but it is not known how many of these swaps are currently being used in conjunction with commercial paper programs.[13] Some commercial paper programs involve international debt issues in conjunction with both interest rate and currency swaps.

Foreign Commercial Paper Markets

While the U.S. market is by far the largest, a variety of foreign commercial paper markets began operating in the 1980s and early 1990s. Table 6 lists the international markets and shows estimates of paper outstanding at the end of 1990. Even though the U.S. commercial paper market continued to grow in

---

[12] Some suggested reasons include market inefficiencies and differences in agency costs and bankruptcy costs across various forms of debt. Wall and Pringle (1988) provide a review of the uses and motivations for interest rate swaps.

[13] Einzig & Lange (1990) discuss some examples of interest rates swaps used in practice.

Table 6 International Commercial Paper Markets Amounts Outstanding, End of 1990
Billions of U.S. dollars

United States	557.8
Japan	117.3
France	31.0
Canada	26.8
Sweden	22.3
Spain	20.0*
Australia	10.9
United Kingdom	9.1
Finland	8.3
Norway	2.6
Netherlands	2.0
Euro-CP	70.4
Total	878.5

*Estimate
Source: Bank for International Settlements.

the later 1980s, its share of the worldwide commercial paper market fell from almost 90 percent in 1986 to less than 65 percent in 1990. The Japanese market, which began in 1987, is the largest commercial paper market outside the United States. In Europe, the French, Spanish, and Swedish commercial paper markets are well established and the German market has shown rapid growth since it began in 1991.[14]

Some U.S. firms simultaneously maintain a commercial paper program in the United States and issue dollar-denominated commercial paper abroad in the Euro commercial paper market. The Euro commercial paper market developed from note issuance and revolving underwriting facilities of the late 1970s in which firms issued tradable notes with the characteristics of commercial paper in conjunction with a loan agreement in which a bank or bank syndicate agreed to purchase the notes if the issuer was unable to place them with investors. In the early 1980s, higher-quality issuers began issuing notes without the backup facilities. The Euro commercial paper market grew rapidly from 1985 to 1990. By the middle of 1992, outstanding Euro commercial paper totaled $87 billion. U.S. financial and industrial firms are important issuers, either directly or through their foreign subsidiaries. Approximately 75 percent of Euro commercial paper is denominated in U.S. dollars while the remainder is denominated in European currency units, Italian liras, and Japanese yen. Issuers commonly issue Euro commercial paper in dollars and use swaps or foreign exchange transactions to convert their borrowings to another currency. The foreign mar-

[14] Bank of International Settlements (1991) reviews the international commercial paper markets. Also see Euromoney (1992) for a review of the European money markets.

kets, including the Euro commercial paper market, provide issuers flexibility in raising short-term funds, allowing them to diversify their investor base, to establish presence in the international credit markets, and to obtain the lowest cost of funds.

While the Euro commercial paper market has similarities to the U.S. market, there are some important differences. The maturity of Euro commercial paper has been longer than in the United States, typically between 60 to 180 days, and, partly reflecting the longer maturities, there is an active secondary market. There is some evidence that the credit quality of the typical issuer in the Euro commercial paper market is not as high as in the U.S. market. Both Standard & Poor's and Moody's rate Euro commercial paper programs, but ratings have not been as crucial in the Euro market as they have been in the U.S. market. U.S. firms with less than A1-P1 ratings have found that the Euro market has been more receptive than the domestic market to commercial paper issues with no credit enhancements attached. Higher default rates abroad reflect the less stringent credit standards. Fons and Kimball (1992) estimate that the amount of defaults as a percent of the total volume of commercial paper issued in the non-U.S. markets (including the Euro commercial paper market) in 1989 to 1991 was 0.0242 percent, which was significantly greater than the 0.0040 percent in the U.S. market. In 1989, the four Euro commercial paper defaults affected almost 1 percent of the market.

The Growing Importance of Commercial Paper

The rapid growth of commercial paper shown in Figure 1 reflects the advantages of financing and investing using the capital markets rather than the banking system. To a significant extent, the advantage of commercial paper issuance is cost: high-quality issuers have generally found borrowing in the commercial paper to be cheaper than bank loans. The cost of commercial paper programs, including the cost of distribution, agent fees, rating fees, and fees for backup credit lines, are small, amounting to perhaps 15 basis points in a large program. A highly rated bank borrows at a cost of funds comparable to other commercial paper issuers, and it must add a spread when lending to cover the expenses and capital cost of its operations and to cover any reserve requirements. Riskier firms are willing to pay this spread because the bank adds value by generating information about the creditworthiness of the borrower which enables it to lend at less cost than the commercial paper market. A large creditworthy issuer will generally find it cheaper to bypass the bank and raise funds directly in the credit market.

The growth of the commercial paper market can be viewed as part of a wider trend towards corporate financing using securities rather than bank loans. Other aspects of this trend, commonly referred to as asset securitization, include the rapid growth of the bond and junk bond markets and the market for asset-backed securities. The pace of asset securitization increased sharply in the

1980s. New security technology, including the development of risk management tools like swaps and interest rate caps, became widespread. At the same time, established markets expanded to include new issuers. Smaller, riskier firms increased their issuance of long-term bonds and entered the commercial paper market with asset-backed paper and letter of credit programs. Commercial paper is likely to remain a significant source of financing for domestic and foreign firms and a relatively safe short-term security for investors.

REFERENCES

Bank for International Settlements. *International Banking and Financial Market Developments.* Basle, Switzerland, August 1991.

Board of Governors of the Federal Reserve System. *Flow of Funds Accounts, Financial Assets and Liabilities, First Quarter 1992.* Washington: Board of Governors, 1992.

Crabbe, Leland, and Mitchell A. Post. "The Effect of a Rating Change on Commercial Paper Outstandings." Washington: Board of Governors of the Federal Reserve System, January 1992.

__________. "The Effect of SEC Amendments to Rule 2A-7 on the Commercial Paper Market." Washington: Board of Governors of the Federal Reserve System, May 1992.

The Depository Trust Company. Final Plan for a Commercial Paper Program. Memorandum, April 1990.

Einzig, Robert, and Bruce Lange. "Swaps at TransAmerica: Analysis and Applications," *Journal of Applied Corporate Finance,* vol. 2 (Winter 1990), pp. 48–58.

"1992 Guide to European Domestic Money Markets," published with Euromoney, September 1992.

Federal Reserve Bank of New York. Press Release No. 1932, Market Reports Division, May 13, 1992.

Felix, Richard, ed. *Commercial Paper.* London: Euromoney Publications, 1987.

Fons, Jerome S., and Andrew E. Kimball. "Defaults and Orderly Exits of Commercial Paper Issuers," *Moody's Special Report,* Moody's Investor Service, February 1992.

General Motors Acceptance Corporation. *1991 Annual Report.* Detroit, Michigan: 1992.

Kakutani, Masaru, M. Douglas Watson, Jr., and Donald E. Noe. “Analytical Framework Involving the Rating of Debt Instruments Supported by ‘Keepwell’ Agreements,” *Moody’s Special Comment*, Moody’s Investors Service, July 1992.

Kavanagh, Barbara, Thomas R. Boemio, and Gerald A. Edwards, Jr. “Asset-Backed Commercial Paper Programs,” *Federal Reserve Bulletin*, vol. 78 (February 1992), pp. 107–16.

Peavy, John W., and S. Michael Edgar. “A Multiple Discriminant Analysis of BHC Commercial Paper Ratings,” *Journal of Banking and Finance*, vol. 7 (1983), pp. 161–73.

Samson, Solomon B. and Mark Bachmann. “Paper Backup Policies Revisited,” *Standard & Poor’s Creditweek*, September 10, 1990, pp. 23–24.

Slovin, Myron B., Marie E. Sushka, and Carl D. Hudson. “Corporate Commercial Paper, Note Issuance Facilities, and Shareholder Wealth,” *Journal of International Money and Finance*, vol. 7 (1988), pp. 289–302.

Standard & Poor’s Corporation. *Commercial Paper Ratings Guide.* April 1991.

Stigum, Marcia. *The Money Market.* Homewood, Ill.: Dow Jones-Irwin, 1990.

Wakeman, L. Macdonald. “The Real Function of Bond Rating Agencies,” *Chase Financial Quarterly*, vol. 1 (Fall 1981), pp. 19–25.

Wall, Larry D., and John J. Pringle. “Interest Rate Swaps: A Review of the Issues,” Federal Reserve Bank of Atlanta *Economic Review*, November/December 1988, pp. 22–40.

Personal Saving Behavior and Real Economic Activity

Roy H. Webb

Many analysts view personal saving behavior, summarized by statistics such as the personal saving rate or household debt acquisition, as a key determinant of real economic activity. Some blame the recent sluggishness of output and employment growth on low personal saving in recent years.

> The low rate of personal saving leaves consumers unprepared for their customary role of pulling the economy out of recession, according to Lacy Hunt, chief economist for the Hong Kong Bank Group in the United States.[1]

> The biggest problem in America's economy:. is debt. It is not so much corporate debt,:. but consumer debt:.. [I]ndividuals are in no position to spend the economy out of recession:. no room to raise borrowing, no savings to run down.[2]

> The past three years have not been a normal postwar recession, but a depression:.. [T]he current episode of strength will soon peter out in a triple dip, followed by a deeper stage of depression:.. Debt has grown too large to be sustained out of cash flow. As soon as the balance sheet is depleted, a deeper crisis of asset liquidation will catch the world by surprise.[3]

The author gratefully acknowledges helpful comments from Dan Bechter, Gary Burtless, Mary Finn, Marvin Goodfriend, and Thomas M. Humphrey. Max Reid provided valuable research assistance. The views and opinions expressed in this article are solely those of the author and should not be attributed to the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] *American Banker* 1992, p. 11.

[2] *The Economist* 1992, pp. 13–14.

[3] Davidson 1993, p. A15. While the author is also looking at saving by other sectors in the United States as well as saving in other leading economies, the saving behavior of U.S. households plays an important role in his analysis.

Others take a longer-term view and see low personal saving lowering capital formation, thereby leading to lower growth in real output, productivity, and future standards of living.

> Mr. Hunt said savings as a percentage of disposable income are lower than at the end of any of the five previous recessions. This will have adverse long-term ramifications for investment in plant and equipment and for U.S. competitiveness in world markets:.. With the U.S. saving rate consistently less than one-half those in Japan and Germany, "it would appear that the nation is ill-prepared:. to compete effectively." [4]

> The United States has long had one of the lowest saving rates in the world:.. The low rate of saving means that the United States has a lower level of income and possibly a substantially lower rate of income growth than would otherwise be possible.[5]

> [L]ow national saving is the most serious problem facing the U.S. economy. Low saving accounted for:. the slow growth in standards of living that continued throughout the 1980s:.. [T]he low saving rate is increasingly the result of insufficient personal saving by U.S. households.[6]

In contrast to these views, this paper argues that personal saving data alone reveal little about the current or future state of the economy. Consider first the assertion that low saving, as it is conventionally measured, is to blame for the recent sluggishness of real economic activity. Most economists would agree that a proposition is valid if (1) it accords with a generally accepted, internally consistent theoretical framework, (2) measurements implied by the theory are consistent with its predictions, and (3) alternative theories are not consistent with some of the measurements. The assertion that recent economic sluggishness is tied to low saving is questionable on several grounds. Proponents of that view have failed to present a well-articulated theory; the view is not consistent with data on household wealth; and the basic data on personal saving rates can be explained in a way that does not imply a linkage of low saving rates and recent economic sluggishness.

Next, the assertion that current low saving will result in lower long-term growth does follow from an influential theoretical framework, unlike the asserted linkage between personal saving and current economic activity. That long-run relation, however, is more complex than suggested by simple theory. Determining the adequacy of the nation's prospects for real growth will require much more data than simple saving rates.

A by-product of this inquiry is an exposition and explanation of several statistics that help describe the saving behavior of households. The most commonly cited statistic is not consistent with standard consumer theory, and may

[4] *American Banker* 1992, p. 11.

[5] Feldstein 1989. While the author is referring to national saving, household saving is an important part of his analysis.

[6] Summers 1990, p. 153.

also be subject to substantial measurement error. All told, one should not view household saving or debt rates as conclusive indicators of real economic vitality; at best they may suggest that a look at more relevant data is in order.

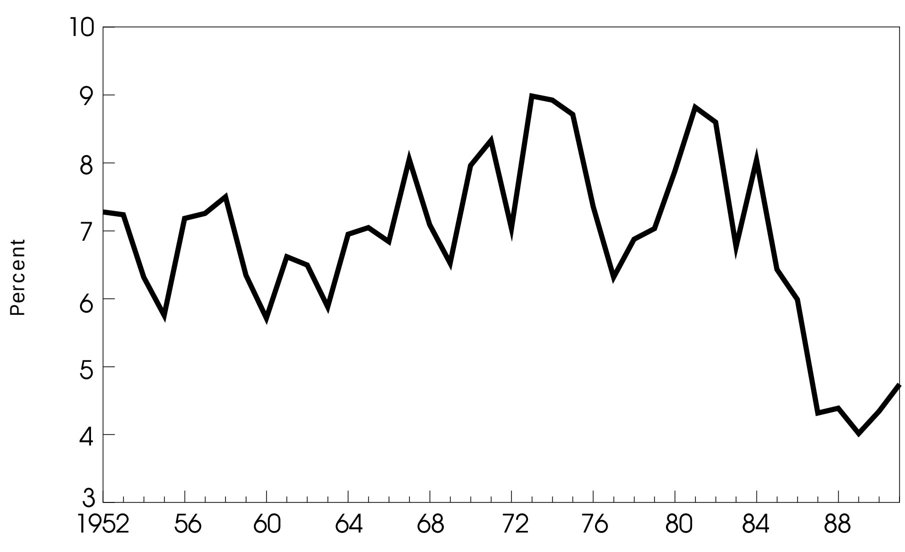
1. WEALTH AND SAVING

This section tackles the assertion that recent economic sluggishness is due to recent rates of household saving and debt acquisition. The first step is to question the relevance of the most widely cited statistic, which was not derived from the most widely used consumer theory. A measure that is more relevant to consumer spending is then discussed.

Measuring Saving

Analysts who view recent personal saving with alarm often focus on a particular statistic published in the National Income and Product Accounts (NIPAs). That statistic, which is often referred to as the *saving rate*, is simply the ratio of saving to disposable personal income. As shown in Figure 1, it declined from 9 percent to 4 percent in the 1980s and remains well below levels of the 1950s,1960s, and 1970s. To understand the significance of that decline, note first that personal saving is defined as unspent income. The definition suggests the indirect approach actually used to estimate saving, which is to subtract estimated

Figure 1 Personal Saving Rate



Note: Ratio of personal savings to disposable personal income.
Source: National Income and Product Accounts

outlays (mostly consumer spending for goods and services) from estimated income. Saving, however, is much smaller than either income or spending; therefore any error in estimating either item will cause a much larger error in estimating saving. For example, in 1991 personal income was $4.8 trillion and personal saving was $0.2 trillion; thus a 1 percent error in estimating personal income would result in a 24 percent error in estimating personal saving. Since neither income nor spending is measured precisely, personal saving is probably estimated with a large error. That fact alone should make any user of saving data especially cautious.[7]

Even if NIPA income and consumption were both measured precisely, one might question the relevance of the particular definitions employed. Two examples illustrate this point.

(1) NIPA income is defined as income from current production. By definition asset revaluations are not part of NIPA income. A country could therefore boost its NIPA income by depleting its exhaustible mineral reserves without having to account for the reduced land values that would result. Similarly, NIPA personal income is not affected when the market value of assets owned by individuals changes. Accordingly, the bull market in residential real estate in the 1970s and bull markets in stocks and bonds in the 1980s did not directly affect NIPA measures of income and saving. As will be discussed below, asset appreciation, whether or not officially measured, can substitute for saving in that both provide the means for future consumption.

(2) Another definitional problem is dividing private spending between consumption and investment. NIPA investment is defined as the purchase of physical assets. If a person acquires productive capabilities through additional schooling, any payments for tuition, textbooks, and related items are defined as consumption. Many economists, however, see a strong analogy between *physical capital,* the tangible assets that can be used for future production, and *human capital,* the skills and abilities that people can use for future production. Since future production can be boosted by either physical or human capital formation, and since the purchase of either human or physical capital involves a trade-off of consumption today for future productive capacity, it is somewhat arbitrary to label spending for one as investment while labeling spending for the other as consumption.[8]

[7] There is even more reason to be suspicious of early estimates of saving rates, which are based on incomplete data. Months or even years after the first estimates, revised values based on more complete information can substantially change the reported saving rates. At times the first reports have had significant bias. For example, from 1980 to 1987 initial releases underestimated saving rates by an average of 2 full percentage points (200 basis points). It is certainly conceivable that current reports of low saving will be revised upward at some future date.

[8] Estimates of the size of the stock of human capital suggest it is no minor matter. Jorgenson and Fraumeni (1989), for example, estimated the value of the stock of human capital to be $194 trillion in 1984, versus $16 trillion for tangible physical capital.

Two Nobel Laureates, Sir John Hicks and Milton Friedman, have separately noted that the NIPA definition of income (income from current production) was not derived from mainstream economic theory. "[A]ny one who seeks to make a statistical calculation of social income is confronted with a dilemma. The income he can calculate is not the true income he seeks; the income he seeks cannot be calculated."[9] "I do not believe that [terms such as Income and Saving] are suitable tools for any analysis which aims at logical precision."[10] "The designation of current receipts as 'income' in statistical studies is an expedient enforced by limitations of data."[11]

A concern of both Hicks and Friedman was that while national income accountants were developing a system that could display an abundance of consistent information *at any particular point in time*, the information would not be consistent over time.[12] Because the relative price of capital changes over time, the evolution of the market value of the capital stock cannot be measured by NIPA investment and depreciation. Accordingly, the time profile of personal wealth cannot be constructed from initial wealth holdings and NIPA saving data.

Wealth

One could approach the NIPAs with a fresh eye and reconstruct aggregates such as income and saving that are based more firmly on economic theory. Since that task is beyond the scope of a single paper, an interested reader is referred to a good book on the subject such as Eisner (1989). The more modest aim of this section is to suggest that aggregate wealth figures are relevant for analysis of consumer spending and real economic activity.[13]

In order to appreciate the linkage of consumption and wealth, consider the following problem. An individual wishes to consume at a constant rate over his lifetime. His salary, however, will rise over time but then cease during retirement. What constant level of consumption can be maintained?

The answer is illustrated in Figure 2, which illustrates a simple version of the *life-cycle* theory of consumption. Early in life when earnings are low he

[9] Hicks 1939, p. 179.

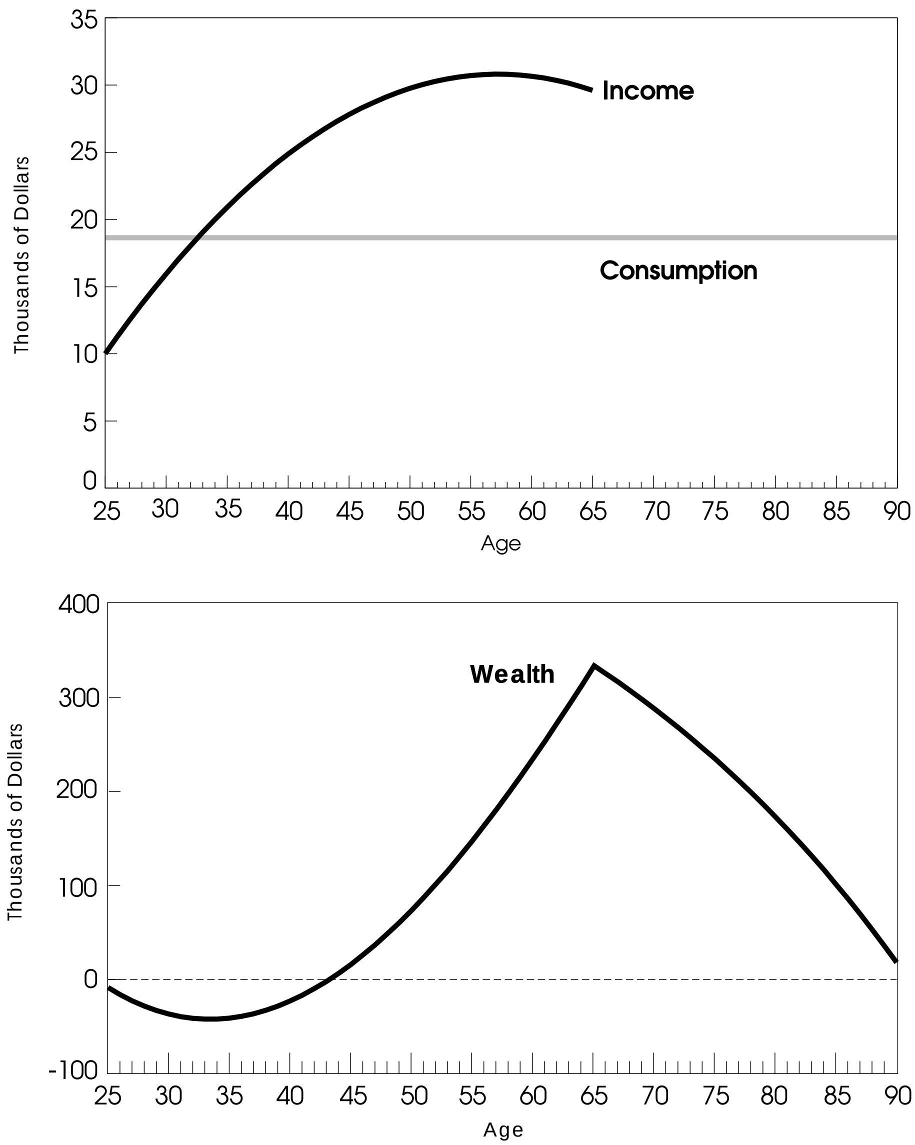
[10] Ibid., p. 171.

[11] Friedman 1957, p. 10.

[12] The choice of words is deliberate; consider "Since we have shown in the preceding chapters what determines the volume of employment *at any time*, it follows, if we are right, that our theory must be capable of explaining the phenomena of the Trade Cycle" (Keynes 1936). The textbook IS-LM presentation of Keynesian theory continues the point-in-time focus, thereby leading to shortcomings such as (1) investment not affecting the capital stock and (2) expectations taken as given rather than explained. In contrast, more recent dynamic equilibrium models such as those surveyed in Sargent (1987) or Barro (1989) explicitly model the evolution of the capital stock, expectations, and other variables over time. These newer models highlight the shortcomings of the NIPA definition of income, whereas models of the IS-LM type fit well with the NIPA definition.

[13] A more detailed look at saving, wealth, and economic activity is taken by Bradford (1990).

Figure 2 An Individual's Optimal Pattern of Wealth



Note: The horizontal axis represents time, divided between a working life of 40 years and retirement of 25 years. The vertical axis represents dollars of constant purchasing power. Given the path for income, consumption is the largest constant value consistent with a real interest rate of 3 percent and zero initial and final wealth, and the path for wealth is then calculated. The general shape of the income line, including the ratio of peak to initial income and the age of peak income, was taken from Graham and Webb (1979), and was calculated from cross-sectional estimates of lifetime earnings of men with college degrees.

would like to borrow to raise consumption; as earnings rise he would repay the accumulated debt and then build wealth that could be consumed during retirement.[14] While this simple example abstracts from uncertainty and other complexities of the real world, it serves to present the intuition of the basic economic theory of consumption.[15] A few key points should be noted. (1) Optimal saving varies substantially over an individual's life, swinging from negative to positive to negative. (2) A single observation of income and consumption describes just what the individual is doing at a single point in time. (3) An observation of wealth adds the additional information on the results of all past saving. (4) The ability to borrow and save allows the individual to enjoy a stable consumption stream despite a variable income stream. (5) The ability to borrow presupposes that someone else has already accumulated wealth and is willing to lend; in this example, an older individual with positive wealth might wish to lend to a younger one wishing to borrow.

To measure wealth one must track over time the prices and quantities of commodities that are not immediately consumed. For the whole economy, land, residential structures, and business plant and equipment are important items that can be productively employed for substantial lengths of time. Individuals, however, often do not own such assets directly but instead own financial assets—the paper claims to the physical assets or the income streams resulting from their use.

When people acquire physical and financial assets in order to smooth consumption over time, one can track their ability to pay for future consumption. A *balance sheet* for the household sector presents the assets and liabilities held by persons (rather than firms or government agencies).[16] One can look at the detailed information or use a simple summary statistic, such as households' net financial assets (or financial net worth), which is defined as financial assets such as cash, bank accounts, stocks, bonds, mutual fund shares and pension fund reserves, minus financial liabilities such as mortgages, revolving credit, and installment loans. This magnitude does measure the capacity of households to spend, whereas any period's saving rate does not.

Although the NIPAs do not contain comprehensive statements of wealth, estimates are contained in the Flow of Funds Accounts (FFAs) published by the

[14] An implication of this model is that saving should anticipate changes in labor income to the extent that such changes can be predicted. Campbell (1987) has found some evidence for this even with a NIPA saving measure.

[15] The extent to which a model of this type has been consistent with actual consumption behavior was examined by Fuhrer (1992), who found that while the model predicts long-run behavior well, it did not predict the downturn in auto sales in the 1990–91 recession.

[16] The NIPA definition of personal saving includes saving from families and single persons plus saving by unincorporated businesses, nonprofit institutions serving persons, and private welfare funds and private trust funds. Included under this definition are certain investment returns from private, but not public, pension funds. The reasoning underlying this definition of the household sector is discussed by Holloway (1989).

Table 1 Assets and Liabilities of Households, 1991
Billions of dollars

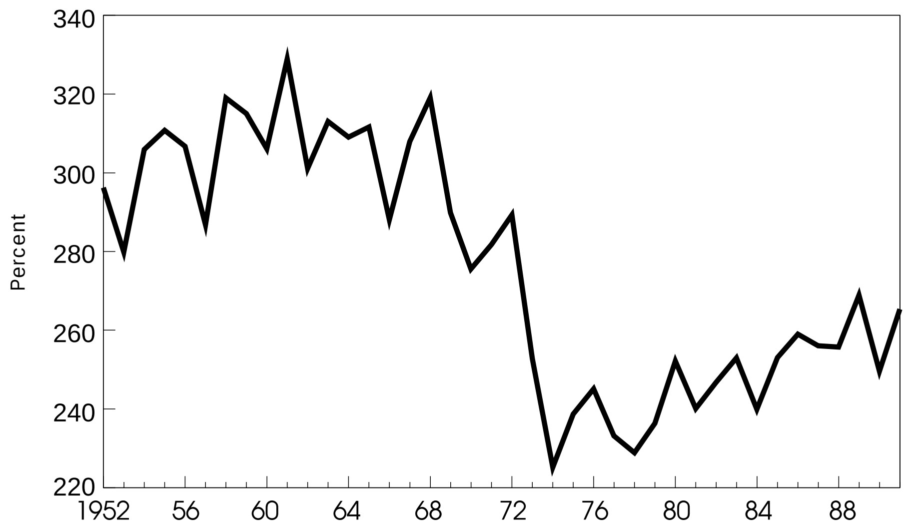
Assets	24,292		
Tangible		9,102	
Owner-occupied housing			3,712
Land			2,624
Consumer durables			2,099
Tangible assets of nonprofit institutions			677
Financial		15,190	
Deposits			3,376
Government securities			838
Bonds and other credit market instruments			1,156
Mutual fund shares			734
Corporate equity			2,334
Equity in noncorporate business			2,568
Pension fund reserves			3,473
Other			711
Liabilities	4,190		
Home mortgages			2,854
Installment consumer credit			744
Other			592
Net Worth	20,102		
Financial Net Worth		11,000	

Note: Data represent year-end values and are taken from the Federal Reserve Board's "Balance Sheets for the U.S. Economy 1960–91," (Release C.9) March 1992.

Board of Governors of the Federal Reserve System.[17] The FFAs contain detailed balance sheets for financial intermediaries, other businesses, households, and the federal government. A simplified balance sheet for the household sector is presented in Table 1 that includes major categories of assets and liabilities. A word of warning: many items on the household balance sheet are estimated as residuals, just as NIPA saving is a residual. Accordingly, measurement errors for many items can affect estimates of household net worth. Also, the FFAs take NIPA values as starting points for many estimates, and thus measurement errors in items such as income will be present in both sets of accounts. Moreover, some items in the FFAs are inherently difficult to estimate with precision. Also, while in principle each item should be measured at market value, in practice market values are not calculated for debt instruments such as mortgages and corporate, federal government, and municipal bonds. Corporate equity holdings, however, are presented at market value. For these reasons net worth is a statistic that is best used with caution.

[17] An introduction to the FFAs is given by Ritter (1974).

Figure 3 Financial Net Worth



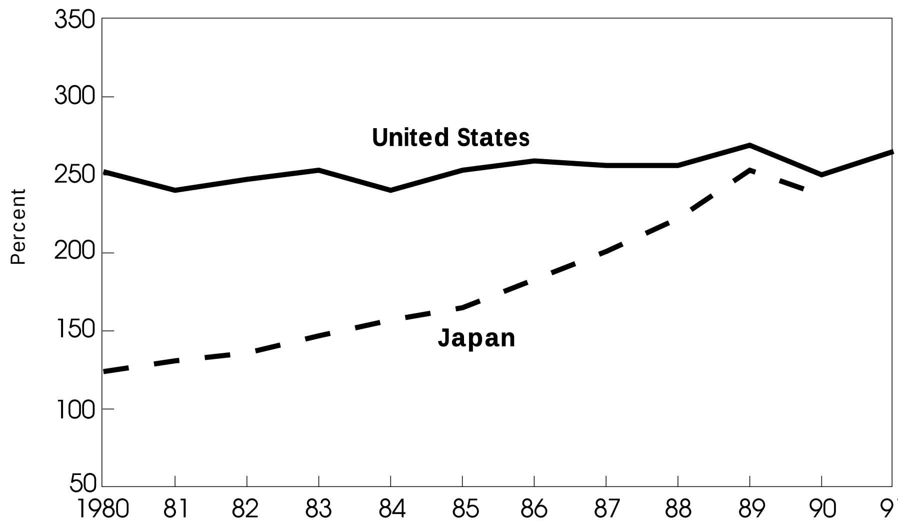
Note: Ratio of net financial assets, household sector (end-of-year) to disposable personal income (annual average).
Source: Flow of Funds Accounts and National Income and Product Accounts.

Since asset holdings grow over time due to real economic growth and inflation, it is useful to scale them by considering the ratio of net financial assets to disposable personal income.[18] As Figure 3 illustrates, the asset-income ratio has fluctuated between 2.2 and 3.3 times income over the past 40 years. The most dramatic change was the decline from 3.2 in 1968 to 2.25 in 1974. During the 1980s the ratio was reasonably stable, rising slightly over the decade. There is nothing in this figure to suggest that consumers have saved so little that

[18] Why not net assets, rather than the less comprehensive net financial assets? The two differ by the amount of real assets owned by households, that is, durable goods, land, and housing. Adequate treatment of the reliability of estimated market values of land and housing would require a good bit of additional discussion that would be tangential to this paper's topic. At a later date an article is planned for this *Quarterly* that will address land and housing values. One piece of evidence is that survey estimates of residential housing values often report much higher values than are given in the FFAs, with analysts such as Curtin, Juster, and Morgan (1989) judging the surveys to be more accurate.

For what they are worth, the FFA housing figures show a decline in household housing and land values of $160 billion in 1990, but with gains of $402 billion in 1989 and $397 billion in 1991. Other figures, such as constant quality price indexes for existing homes do not show a downward movement in 1990 or other years. Thus despite anecdotes of house prices falling these figures do not reveal an aggregate sustained fall in housing values that would affect the conclusion of this section that household wealth rose more rapidly than income over the 1980s and early 1990s.

Figure 4 Net Financial Wealth of Household Sector Ratio to Disposable Personal Income



they cannot now afford to purchase goods and services. If current saving looks low to some observers, that may simply reflect households having accumulated a level of wealth they consider satisfactory. That interpretation is consistent with the decline in the personal saving rate in the 1980s being accompanied by rising household wealth.

How does household wealth in the United States compare with similar statistics in foreign countries? Although the saving rate of U.S. households is below that rate in other countries, Figure 4 illustrates that U.S. households are wealthier. Households in countries holding less wealth often save more in order to accumulate wealth, as exemplified by Japan.

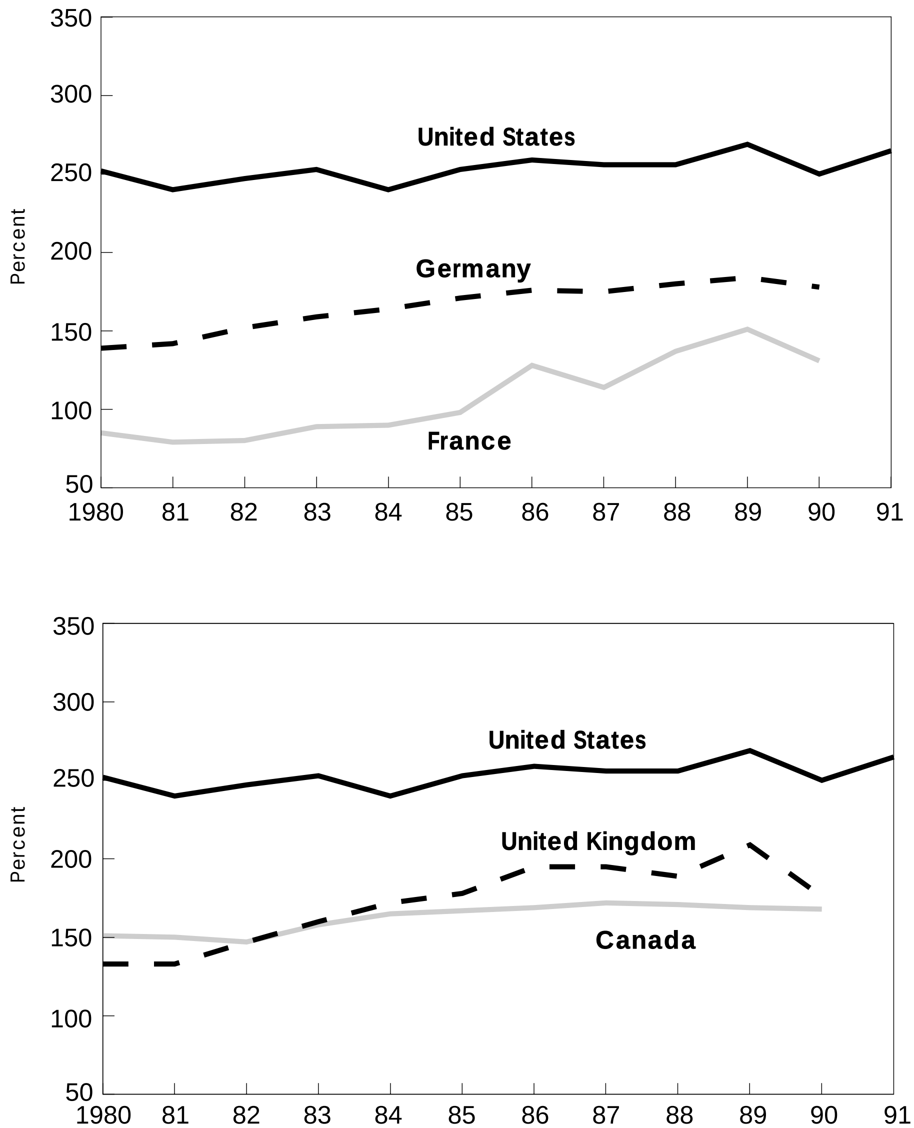
Possible Objections

One might acknowledge the potential usefulness of data on household wealth while raising objections to its current validity. This portion of the paper attempts to address some of the most important concerns.

Concentration of Wealth

Aggregate net wealth would not be a good measure for the majority of households if most wealth were held by relatively few households. While wealth

Figure 4 (continued)



Note: Ratio of net financial assets, household sector (end-of-year) to disposable personal income (annual average).

Source: United States, Flow of Funds Accounts and National Income and Product Accounts; other countries, *OECD Economic Outlook*, December 1991: Organisation for Economic Co-operation and Development, Paris, p. 21.

in the United States is more concentrated than income, it does not appear so concentrated as to render data on wealth irrelevant. The 1989 Survey of Consumer Finances[19] reported that families with incomes below $10,000 had a median net worth of $2,300. It is certainly likely that these families in the lowest quintile of the income distribution would be unable to draw on savings to finance additional consumption. Families in the next quintile, earning between $10,000 and $20,000, had a median net worth of over $27,000, and higher-income groups also had median net worth well over annual income. Another way of describing the survey data is that the median family of a subset of the population that accounts for 80 percent of household income and a greater percentage of aggregate consumption had accumulated a significant amount of wealth. The data therefore do not appear to support the view that wealth is too narrowly distributed to be a useful indicator of potential aggregate consumption.

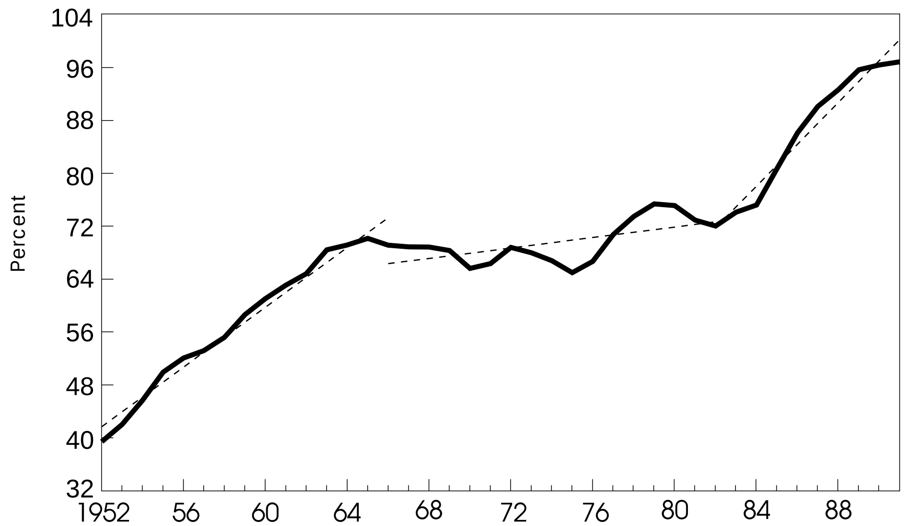
Debt

Another objection addresses the role of debt. If people are highly indebted and many of their assets are illiquid, then the burden of debt repayment might restrict their consumption even if the value of their assets is relatively large. As Figure 5a indicates, household debt is indeed high, relative to the recent past; in 1991 it was almost equal to a full year's disposable personal income, whereas in the mid-1950s debt was less than half a year's income.[20] As the figure indicates, the debt-income ratio grew about 2 percentage points per year from 1952 to the mid-1960s, grew fairly slowly until the early 1980s, and has since grown by 3 percentage points per year. Interestingly, corporate debt shows somewhat similar behavior in Figure 5b, namely, an initial period of growth that was interrupted in the 1970s and resumed in the 1980s.[21] Is that simply a coincidence?

What follows is one possible explanation of the data. The behavior of debt, income, and wealth can be reconciled by noting that a revolution in financial intermediation has occurred over the past 40 years, as a few examples indicate. Credit cards serve two functions: allowing routine transactions to be made without currency and supplying widespread unsecured lines of credit. Mutual funds allow individuals, even those who have fairly small amounts to invest, to benefit from broadly diversified and professionally managed equity and bond portfolios. Home equity lines of credit allow easy, tax-advantaged access to equity in owner-occupied housing. Corporate lending has also been transformed, as many firms that would have borrowed from banks in the 1950s now have access to security markets. In short, the efficiency of financial intermediation

[19] Data are taken from Kennickell and Shack-Marquez (1992).
[20] This figure is similar to Figure 2 in Altig, Byrne, and Samolyk (1992).
[21] This figure is similar to Figure 1 in Paulus (1991).

Figure 5a Household Debt to Income Ratio

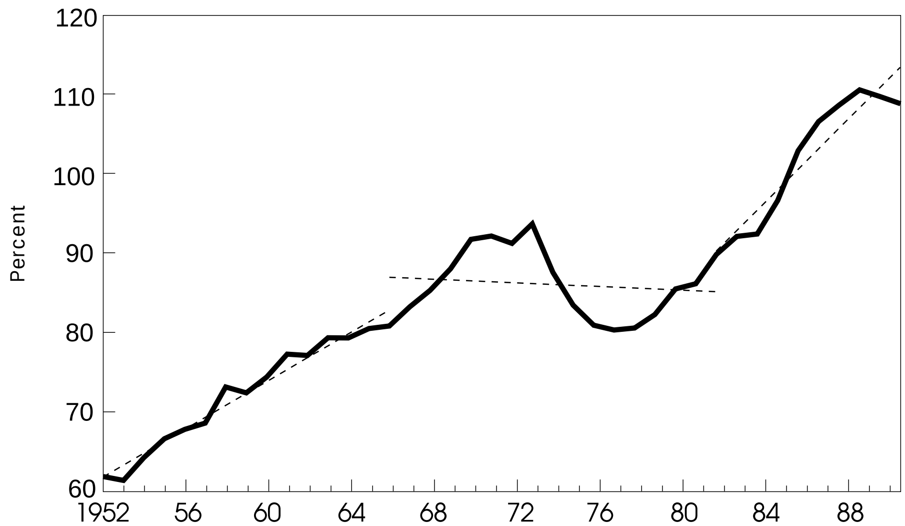


Notes: Ratio of household liabilities (end-of-year) to disposable personal income (annual average). The trend line from 1952 to 1966 represents annual growth of 2.5 percentage points per year, the trend line from 1966 to 1982 represents growth of 0.40 percentage points per year, and the trend line from 1982 to 1991 represents growth of 3.1 percentage points per year.
Source: Flow of Funds Accounts and National Income and Product Accounts.

has improved, in the sense that individuals can better smooth consumption over time and many producers can more readily finance investments yielding high returns.

The trend toward more efficient financial intermediation was interrupted from 1967 to 1981. In a different context Webb (1992) argued that this period had an inflation-tolerant monetary policy. Inflation averaged 1.5 percent from 1952 to 1966, 7.3 percent from 1966 to 1982, 3.9 percent from 1982 to 1991, and 3.0 percent in 1992. Rising inflation in the late 1960s and early 1970s and high, variable inflation in the remainder of the decade played havoc with investing in financial instruments that had traditionally been denominated in nominal terms. During this period Regulation Q restricted the nominal interest rates payable on many deposits, and taxes were levied on nominal rather than real returns. The always difficult process of channeling savings to their most productive uses became even more difficult as financial intermediation was thereby strained and distorted. Debt-income ratios stagnated during this period despite the benefit debtors received from unanticipated inflation and the bias in the tax laws at the time that favored financing by debt rather than by equity.

Figure 5b Corporate Debt to Income Ratio



Notes: Ratio of nonfinancial corporate liabilities (end-of-year) to national income originating in private nonfinancial business. The trend line from 1952 to 1966 represents annual growth of 1.5 percentage points, the trend line from 1966 to 1982 represents a 0.1 percentage point rate of decline, and the trend line from 1982 to 1991 represents growth of 2.6 percentage points per year. Source: Flow of Funds Accounts and National Income and Product Accounts.

The idea that the rise in private debt in the 1980s reflected increasing efficiency of financial intermediation is not consistent with the quotations at the beginning of this paper. The alternative view of the authors quoted seems to be that the rise in debt backed an unsustainable consumption binge. Proponents of that alternative have not, to the author's knowledge, recognized that wealth rose relative to income in the 1980s even though saving rates declined. Since higher levels of wealth allow higher levels of future consumption, consumption levels of the 1980s appear to be sustainable. Moreover, it appears that the rate of return on invested assets was relatively high in the 1980s.[22] A high rate of return is symptomatic of savings being put to highly productive uses, which in turn is symptomatic of efficient financial intermediation.

[22] The change in financial wealth can be stated as the saving from labor income and transfer payments plus the return on assets; in symbols, $dW = Y - C + rW \equiv S + R$. The change in the wealth-income ratio is by definition $d\left(\frac{W}{Y}\right) = \frac{Y\,dW - W\,dY}{Y^2}$; substituting from the previous expression and rearranging terms, an increasing wealth-income ratio means that $\frac{R}{Y} > \frac{W}{Y}g - \frac{S}{Y}$, where g is the growth rate of real income. Thus with a wealth-income ratio of 2.5, a real growth rate of 3.5 percent, and a saving rate of 4.5 percent, the real rate of return on net financial wealth must exceed 4.25 percent.

If one still wished to argue that (1) financial asset holdings are irrelevant due to imperfect markets and (2) household debt levels are nonoptimal in the sense of being higher than fully informed borrowers and lenders would choose, then there are further problems. Why would large numbers of borrowers and lenders make the same mistake in the 1980s? Did they believe there would never be a recession? Unless some such widespread error occurred, what is the basis of the assertion that debt levels are too high? The author is unaware of such questions being seriously addressed; as a result, assertions of debt being too high do not appear to be based on an economic theory involving rational people with stable preferences. When economic theory is used to study imperfect markets, the usual result is that some people are able to borrow *too little,* not too much.[23]

Measurement of Wealth

Another objection concerns the relevance of the FFAs. Questions of definition can arise over items included in household financial assets such as (1) nonprofit institutions as part of the household sector, (2) substantial assets and liabilities recorded at historical values rather than market values, (3) government bonds recorded without excluding a liability for the future taxes that will be levied to pay interest on the bonds, and (4) pension fund reserves that are far removed from household control. These are valid concerns which demonstrate that these statistics from the FFAs, like every other macroeconomic statistic, are not estimated in the exact form that many users would prefer.

Addressing the objections in order, (1) nonprofit institutions account for a small fraction of the household sector's economic activity; note in Table 1 that tangible assets of nonprofit institutions are less than 8 percent of the household sector total. (2) Marking debt instruments to market would strengthen the argument that household wealth is not unusually low, because the increase in bond prices as interest rates fell over the last ten years is excluded from the figures presented in this paper. (3) Excluding government bonds from these figures would not alter any conclusions, since they account for less that 6 percent of household financial assets. Also, economists are divided on the extent to which one should offset government bonds with anticipated future tax liabilities. (4) Wealth held in pension funds can affect household behavior. Households with a large amount of pension wealth can consume more today precisely because they do not have to save as much from current cash flows in order to provide for retirement.

The conceptual and measurement problems with the FFAs suggest that the accounts should be used with caution. Analysts who keep the accounts'

[23] For example, Bernanke and Gertler (1989) present a model in which potential borrowers with low net worth are unable to finance productive investment projects. Whited (1992) presents empirical evidence consistent with the view that financial constraints can reduce investment.

weaknesses in mind will find the data useful. The alternative is to ignore relevant balance sheet data.

Evaluation

Figure 1 illustrates that saving from current income is relatively low, and Figure 5a illustrates that the household debt to income ratio is relatively high. These phenomena can be explained without asserting that something has been so seriously wrong with consumer finances as to explain the past recession and the subpar expansion that followed. The explanation instead notes that the relatively large net worth of households contradicts the notion that consumers are currently unable to finance optimal spending plans.

Authors who have linked saving behavior with recent economic weakness have several obstacles to overcome to establish their point. First, they need to detail the theoretical model of consumer spending that they use to define recent saving rates as too low or recent debt levels as too high. The widely used life-cycle model discussed above is apparently not the basis for such assertions, due to rising wealth levels in recent years.[24] A possible alternative could be models with imperfect loan markets, although these usually imply that debt levels are too low. And when a theoretical model is used to show that savings are too low, or debt too high, the authors then need to explain why consumers saved too little or borrowed too much, and why lenders willingly lent too much. Finally, it would help if the authors explained why they believe the conventional saving rate is measured with sufficient accuracy to allow confident assertions to be made.

The case has not been made that personal saving behavior has much to do with the recent subpar economic performance. While an unproven case might still be valid, there are plausible alternative explanations of the basic data. For example, if weak consumer spending was an important factor in explaining the slow recovery in 1991, that weakness could reflect the uncertainty caused by the large number of permanent job losses in the last few years and the prospect of more losses ahead due to job reductions announced but not implemented by many large organizations.[25]

2. LONG-RUN GROWTH

The previous section found only a questionable theoretical link between saving measures from the recent past and current spending. In contrast, standard economic theory posits a firm link between saving and the long-run level of

---

[24] An exception is the analysis of Bernheim and Scholz (1992), who present evidence that they interpret as showing that people without college educations do not save enough to maintain their standard of living in retirement.

[25] See, for example, Carroll (1992).

real output; in addition, there are theoretical frameworks in which saving can also affect the rate of growth. The linkage is that saving allows investment to raise the stock of productive capital. Many analysts are concerned about growth because recent growth rates in the United States appear low, relative to either growth in the United States in the 1950s and 1960s, or to growth rates in many other countries.

This section first reviews some data on economic growth in the United States, and possible interpretations of that experience. Next, the potential role of saving in a widely used theoretical model is examined. Some recent advances in growth theory that affect the interpretation of saving are next discussed. Unlike the first section of this paper, this section finds that a *properly measured* saving rate would be a useful statistic to the extent that it is a valid indicator of capital formation and thereby also an indicator of future growth prospects. Some empirical evidence on the correlation of saving and investment rates concludes this section.

Recent Experience

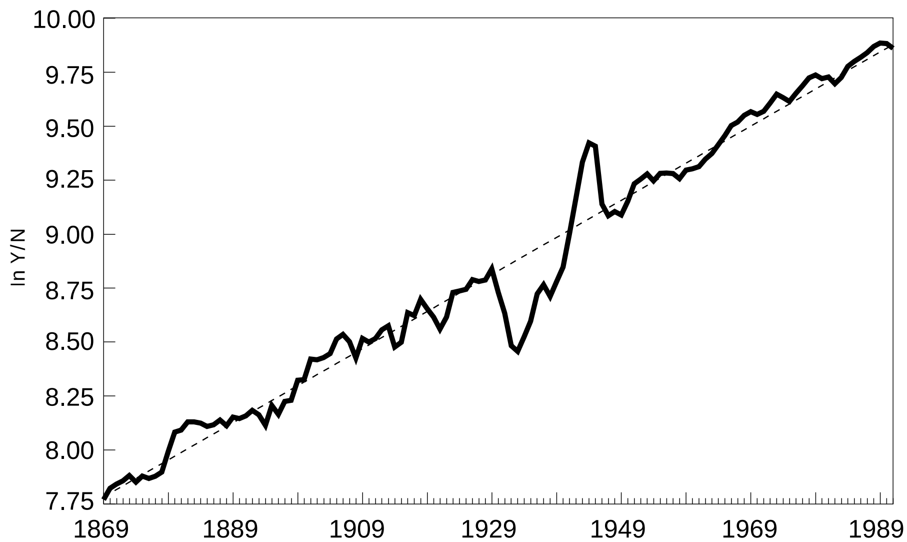
Figure 6 and Table 2 contain some basic data on Gross Domestic Product (GDP) per capita for over a century.[26] Figure 6 illustrates that the growth rate of per capita GDP has fluctuated around a trend of 1.7 percent, which means that it has doubled approximately every 40 years. The two largest departures from trend are the Great Depression and World War II. Table 2 allows one to calculate growth rates for shorter periods. Of particular interest is the most recent experience, in which growth declined from 2.1 percent between 1950 and 1973 to 1.6 percent between 1973 and 1989. Despite the fact that growth in the latter period is close to its long-run trend, some observers believe that the decline in growth indicates that the United States is failing to realize its economic potential.

Table 2 also indicates that while the level of output per capita is higher in the United States than in other major countries, several other countries have grown more rapidly in recent years. Most spectacular is Japan, where output per capita grew by 7.7 percent from 1950 to 1973, and by 3.1 percent from 1973 to 1989. Simply extrapolating the latest growth rates puts several countries ahead of the United States early in the next century. That too leads some to

[26] GDP is used to facilitate comparisons over time and across countries. It is not and was not designed to be a measure of economic welfare. There are also better measures of product one could devise; many investigators, however, believe that the correlation between GDP and a better measure is sufficiently high to warrant the use of GDP statistics.

It should also be noted that the accuracy of almost every economic statistic declines as one goes farther back in time. Analysts who produce the NIPAs today have much more raw data to use to construct aggregate statistics than did their counterparts 40 years ago, who in turn had much more raw data than did the individuals who have constructed estimates for GDP before 1929.

Figure 6 Output per Capita with Trend



Notes: Gross domestic product divided by population, annual data, logarithmic scale. Trend line represents annual growth at a 1.7 percent rate and is based on estimates from 1869 to 1929 and extrapolated for 1930 to 1991.
Source: GDP, National Income and Product Accounts, 1929–91, and Balke and Gordon (1989), 1869–1928; population, United States Census, 1950 to 1991, and *Historical Statistics of the United States: Colonial Times to 1970,* U.S. Government Printing Office, Washington, 1975: Series A7.

Table 2 Gross Domestic Product per Capita, 1985 Dollars, United States Prices

	1870	**1913**	**1950**	**1973**	**1989**
United States	2,247	4,854	8,611	14,103	18,317
Canada	1,347	3,560	6,113	11,866	17,576
France	1,571	2,734	4,149	10,323	13,837
Germany	1,300	2,606	3,339	10,110	13,989
Japan	618	1,114	1,563	9,237	15,101
United Kingdom	2,610	4,024	5,651	10,063	13,468

Note: These figures are taken from Maddison (1991), Table 1.1. They represent per capita GDP, expressed in constant dollars to remove the effects of inflation, and adjusted for differing purchasing power of currencies.

believe that the United States is growing too slowly, and to view low saving as a possible cause.

The Solow Growth Model

The name of Nobel Laureate Robert Solow is linked with a straightforward and influential theoretical model of economic growth.[27] Consider a specific production function, which states with symbols that national product depends on the amounts of capital and labor employed, as well as the state of knowledge:

$$Y_t = K_t^{\alpha}(A_t L_t)^{1-\alpha}, \tag{1}$$

where Y is the quantity of output, K is the stock of capital, L is the labor force, A can be interpreted as the state of knowledge about producing output, t indexes time, and α is a parameter between zero and one, the value of which can be statistically estimated. If one assumes (1) that the labor force and knowledge grow at given exponential rates of n and g, respectively, (2) that a constant fraction s of output is saved and invested,[28] and (3) that capital depreciates at an exponential rate d, it then follows that

$$\ln \frac{Y_t}{L_t} = gt + \left[\frac{\alpha}{1-\alpha} \ln(s) - \frac{\alpha}{1-\alpha} \ln(n+g+d) + \ln A_0 \right] \tag{2}$$

for a country experiencing steady-state growth, that is, a country for which the capital stock is consistent with the model's parameters and initial conditions. Note in equation 2 that the growth rate of output per capita is determined solely by the exogenous parameter g, the growth rate of knowledge. Other parameters in the bracketed term, *including the saving rate,* only affect the *level* of output per capita.

Differences in Growth Rates Across Countries

For a country like the United States in which output per capita does not depart too much from a constant trend over a long interval of time, the assumption of steady-state growth appears reasonable. An opposite case would be a country like Japan immediately after World War II where much of the capital stock had been destroyed. The Solow framework can be used to determine how fast a country off its steady-state growth path would converge to that path. Assuming that the speed of convergence is proportional to the difference (in logarithms) between the steady-state and the actual levels of output per capita, then

$$\theta = (n+g+d)(1-\alpha), \tag{3}$$

[27] A good exposition is Solow (1969).

[28] In the growth literature, the saving rate almost always refers to the national saving rate, which is the personal saving rate plus saving by firms and by the government.

where the parameter θ denotes the speed of convergence to the steady-state path. Note that the speed of convergence does *not* depend on the saving rate.[29] For example, if population growth n is 1 percent per year, the steady-state growth rate g is 2 percent, the depreciation rate d is 4 percent, and α is 0.3, then the speed of convergence would be about 5 percent. In other words, about 5 percent of the percentage gap between actual and steady-state output per capita would be eliminated each year, or half the gap would be closed in about eight years.

The idea of convergence has been used to interpret differential growth rates among different areas or countries. Mankiw, Romer, and Weil (1992), for example, augment the basic Solow model by adding a third factor of production, human capital, to physical capital and labor. Looking at a group of 98 countries and two smaller groups, they found that poorer countries in 1960 tended to grow faster from 1960 to 1985 than did richer countries; the estimated speed of convergence was about 2 percent. Barro and Sala-i-Martin (1992) also found evidence for convergence, both among states in the United States from 1880 to 1988 and in the set of 98 countries over a shorter interval; interestingly, they also estimate speeds of convergence of about 2 percent.

If convergence in the level of per capita output accounted for all the differences in growth, then one would not be concerned that countries with lower output were growing more rapidly than the United States. That faster growth would be a temporary phenomenon and would slow as a country's level of output per capita approached that of the United States. Evidently, however, more than just convergence is needed to account for all the variation in output growth. In Table 2, note that output per capita was higher in the United Kingdom than in the United States in 1870; by 1913 the countries' standings reversed. What accounts for the reversal? Between 1870 and 1913 the United States, Canada, and Germany grew faster than Japan, the poorest country. What accounts for this divergence? What accounts for the experience in the United States from 1950 to 1973 when growth was above the previous trend? And why have many countries remained poor over the last 40 years without showing any tendency toward rapid growth?

Endogenous Growth

These questions illustrate why some economists believe that while convergence is probably an important factor in many cases, other explanations of differential growth rates should also be examined. They have accordingly constructed models that depart in an important way from the basic Solow model. Instead of assuming that the economy's steady-state growth rate is a given value g based

[29] A change in the saving rate can change the steady-state capital stock, however, and thus influence the growth rate off the steady-state path.

on the automatic growth of knowledge, they emphasize the individual decisions that result in growth. This is now an especially rich area of macroeconomic research, and there will be no attempt to mention all the important models. Two examples of such endogenous growth models that are relevant for this paper are Lucas (1988) and Greenwood and Jovanovic (1990).

The Lucas model is of interest in that it provides a reason why growth might be too low, and points to the types of public policies that would raise the rate of growth. The model contains human capital, as do many in the endogenous growth literature, but notably makes an individual's productivity depend on both the individual's level of human capital and the community's average level of human capital. In other words, there is a positive externality to human capital accumulation: an individual's decision to acquire additional human capital would balance his own costs and benefits without taking into account that raising one's own stock of human capital also raises the community's stock and thereby raises the productivity of other members of the community. Human capital accumulation is the basic engine of growth in this model, analogous to the exogenous value g in the Solow model. Due to the positive externality, public policies such as subsidies to education can raise the growth rate and aggregate economic welfare.

An implication of the Lucas model is that saving is relevant, in that it coincides with the capital formation that affects the level and rate of growth of output. The measure of saving implied by his model includes both saving as conventionally measured plus investment in human capital. A generally acceptable measure of the latter would require an ambitious research undertaking. Individual researchers have proposed strategies for estimating investment in human capital, but different strategies have led to vastly different results. Measurement of some of the resources that are used for investment in human capital, such as teachers' salaries, buildings, and textbooks is straightforward. A more difficult question is valuing a student's time in school. How are differences in the quality of education to be estimated? When a person develops skills through experience, how is that measured? A professional consensus has not emerged on these and other difficult questions. But any saving statistic that fails to confront human capital measurement is omitting a very important part.

Financial institutions play a key role in the model of Greenwood and Jovanovic, in which the extent of financial intermediation and the degree of development are linked. Financial intermediation allows a given amount of saving to finance a greater amount of investment than could occur without intermediation. And mature economies can have relatively low saving rates with high growth due to well-developed financial intermediation. Therefore, simply comparing saving rates in different countries would not provide useful information on the adequacy of investment or on future growth prospects.

The theoretical linkage of financial intermediation and growth is supported by empirical evidence. King and Levine (1993) studied real growth and several measures related to financial intermediation in 80 countries from 1960 to 1989. They found a robust correlation between the extent of financial development and contemporaneous growth, and also that financial development predicts future growth.

The Correlation of National Saving and Investment

The intuition linking saving and growth is highlighted by the formal models examined. In the basic Solow model the long-run growth rate is exogenous and is therefore unaffected by saving. Over shorter time spans, however, growth can be affected by saving. Once the growth rate is made endogenous, interpreting saving data can become even more difficult. If the Lucas assumption of externalities in human capital accumulation is important, then we should be focusing on a better understanding and measurement of human capital. And to the extent that more highly developed financial intermediation raises the return to saving, the meaning of a given rate of saving changes as an economy matures.

Despite these difficulties, researchers have presented empirical evidence that suggests a strong linkage between a country's saving and investment. One of the most influential studies, by Feldstein and Horioka (1980), found a strong correlation between rates of saving and investment for 21 countries. Figure 7 presents national saving and gross investment data, relative to GDP, for the postwar United States. The two series clearly move together over the 1960–74 interval studied by Feldstein and Horioka, although for much of the 1980s investment outpaced saving as foreign investment in the United States was relatively large. In 1991 both saving and investment hit postwar lows, reinforcing the concerns of many over inadequate investment due to inadequate saving.

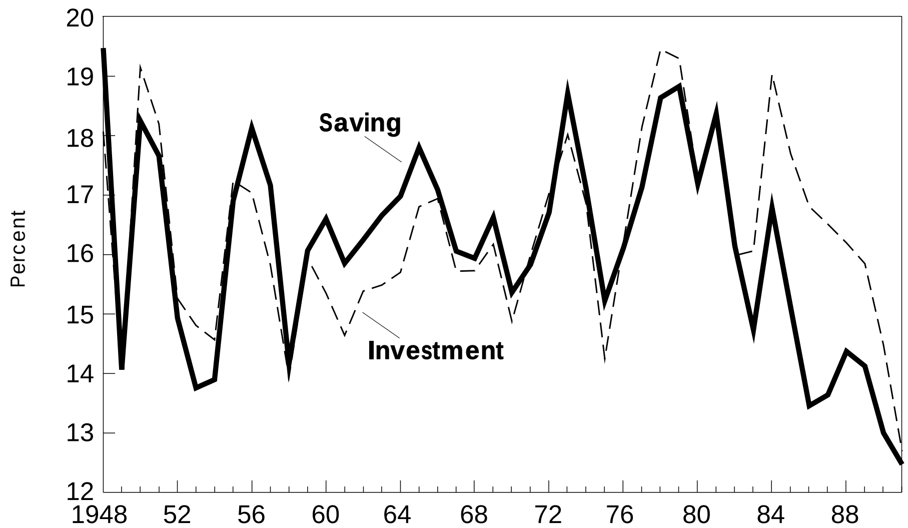
Simple correlations such as this are always difficult to interpret. Two variables can be correlated, even if movements in one do not cause movements in the other, if both are responding to movements of a third variable. There are many possible factors that might explain movements in both saving and investment. For example, both are low at business cycle troughs and rise during cyclical expansions. Interest rates are another factor affecting saving and investment.

A quick look confirms the possibility that the correlation might vanish after allowing for other factors. Table 3 contains empirical results based on:

$$V_t = c + \sum_{i=1}^{4} \alpha_{t-i} \frac{I_{t-i}}{Y_{t-i}} + \sum_{i=1}^{4} \beta_{t-i} \frac{S_{t-i}}{Y_{t-i}} + \sum_{i=1}^{4} \gamma_{t-i} R_{t-i} + \sum_{i=1}^{4} \delta_{t-1} U_{t-i} + e_t, \quad (4)$$

where V is the dependent variable, either the gross investment to GDP ratio or the national saving to GDP ratio, I is investment, Y is GDP, S is saving, R is the interest rate on 90-day Treasury bills, U is the capacity utilization rate in manufacturing, e is an error term, t indexes time, and the remaining symbols are coefficients that can be estimated by ordinary least squares. The

Figure 7 Gross Saving and Investment Relative to GDP



Note: Gross private domestic investment, divided by gross domestic product, and gross national saving, divided by gross domestic product.
Source: National Income and Product Accounts.

two equations can be used to examine the extent to which either investment or saving is correlated with previous values of those two series and also with previous values of an interest rate and the capacity utilization rate (which can be interpreted as an indication of the stage of the business cycle). Especially notable results from the investment equation are (1) the lagged variables are associated with a large portion of the movement of the investment-GDP ratio, and (2) the coefficients on lagged saving are not significantly different from zero, unlike coefficients on all the other variables.

If taken at face value, these results suggest that savings in the recent past do not directly affect investment; however, the results in Table 3 are suggestive rather than definitive. Most importantly, there was no experimentation with other measures of saving, investment, and output,[30] and there was no analysis of contemporaneous correlations of saving, investment, output, interest rates, and possibly other variables. The results do show, however, that the empirical correlation of investment and saving is not easy to interpret since it could well reflect the business cycle and possibly other influences.

[30] Cullison (1991) studied the relation of several measures of saving to quarterly GDP growth.

Table 3 Regression Results for Investment and Saving Rates

(1) $$\frac{I_t}{Y_t} = c + \sum_{i=1}^{4} \alpha_{t-i} \frac{I_{t-i}}{Y_{t-i}} + \sum_{i=1}^{4} \beta_{t-i} \frac{S_{t-i}}{Y_{t-i}} + \sum_{i=1}^{4} \gamma_{t-i} R_{t-i} + \sum_{i=1}^{4} \delta_{t-1} U_{t-i}$$

Time bounds: 1952 Q2 to 1992 Q2 $\overline{R}^2 = .84$

Variable	F-Statistic	Significance Level
I/Y	54.06	.00
S/Y	0.80	.53
R	5.09	.00
U	3.61	.01

(2) $$\frac{S_t}{Y_t} = c + \sum_{i=1}^{4} \alpha_{t-i} \frac{I_{t-1}}{Y_{t-i}} + \sum_{i=1}^{4} \beta_{t-i} \frac{S_{t-i}}{Y_{t-i}} + \sum_{i=1}^{4} \gamma_{t-i} R_{t-i} + \sum_{i=1}^{4} \delta_{t-1} U_{t-i}$$

Time bounds: 1952 Q2 to 1992 Q2 $\overline{R}^2 = .87$

Variable	F-Statistic	Significance Level
S/Y	101.45	.00
I/Y	1.64	.17
R	6.07	.00
U	2.26	.07

Note: *I* is gross private domestic investment, *Y* is GDP, *S* is gross national saving, *R* is the 90-day Treasury bill rate, and *U* is the capacity utilization rate in manufacturing.

3. CONCLUSION

Although many analysts cite the personal saving rate as a key indicator of the current and prospective strength of the economy, the saving rate alone actually reveals little about current and future conditions. Difficulties in defining, measuring, and interpreting saving should be kept in mind by prospective users.

Current saving data reveal little about prospective consumer spending. Basic economic theory instead indicates that household wealth measures resources accumulated for future spending. In addition, it would be a mistake to focus simply on one part of the household balance sheet, debt, without first determining its optimal level. Since debt has the positive roles of allowing individuals to smooth consumption over time as income varies and of financing productive investment, it should not be simply assumed that current debt levels are too high.

In contrast to the weak link between recent saving and current consumer spending, there is a well-established theoretical link between saving and investment, and therefore between saving and economic growth. Even here the message given by saving data can be difficult to interpret, since low national saving can occur while investment is buoyed by inflows of foreign funds; in addition, human capital formation is omitted from the usual saving measure. To determine whether national investment is adequate it could be more productive to look directly at detailed investment data. If profitable investments were not being made, one might wish to search for underlying causes such as taxes, regulation, externalities, or inadequate financing. A focus on conventionally measured saving may well divert attention from these important fundamentals.

REFERENCES

Altig, David, Susan M. Byrne, and Katherine A. Samolyk. "Is Household Debt Inhibiting the Recovery?" Federal Reserve Bank of Cleveland *Economic Commentary,* February 1, 1992.

American Banker. "Low Consumer Savings Rate Seen as a Hurdle to Recovery," February 10, 1992, p. 11.

Balke, Nathan S., and Robert J. Gordon. "The Estimation of Prewar Gross National Product: Methodology and New Evidence," *Journal of Political Economy,* vol. 97 (February 1989), pp. 38–92.

Barro, Robert J. *Modern Business Cycle Theory.* Cambridge, Mass.: Harvard University Press, 1989.

__________, and Xavier Sala-i-Martin. "Convergence," *Journal of Political Economy,* vol. 100 (April 1992), pp. 223–51.

Bernheim, B. Douglas, and John Karl Scholz. "Private Saving and Public Policy," in James M. Poterba, ed., *Tax Policy and the Economy.* Cambridge, Mass.: National Bureau of Economic Research, 1992.

Bernanke, Ben, and Mark Gertler. "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review,* vol. 79 (March 1989), pp. 14–31.

Bradford, David F. "What is National Saving? Alternative Measures in Historical and International Context," in Charls E. Walker, Mark A. Bloomfield, and Margo Thorning, eds., *The U.S. Savings Challenge.* Boulder: Westview Press, 1990.

Campbell, John Y. "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica,* vol. 55 (November 1987), pp. 1249–73.

Carroll, Christopher. "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity,* 2:1992, pp. 61–156.

Cullison, William E. "Measures of Saving as Indicators of Economic Growth." Paper presented to the Western Economic Association, Seattle, 1991; forthcoming in *Contemporary Policy Issues.*

Curtin, Richard T., F. Thomas Juster, and James N. Morgan. "Survey Estimates of Wealth: An Assessment of Quality," in Robert E. Lipsey and Helen Stone Tice, eds., *The Measurement of Saving, Investment, and Wealth.* Chicago: University of Chicago Press, 1989.

Davidson, James Dale. "The Depression of the 1990s—It's Already Here," *The Wall Street Journal,* February 11, 1993, p. A15.

The Economist. "The Confidence Factor," April 18, 1992, pp. 13–14.

Eisner, Robert. *The Total Incomes System of Accounts.* Chicago: University of Chicago Press, 1989.

Feldstein, Martin, and Charles Horioka. "Domestic Saving and International Capital Flows," *Economic Journal,* vol. 90 (June 1980), pp. 314–29.

__________. "Tax Policy for the 1990s: Personal Saving, Business Investment, and Corporate Debt," *American Economic Review,* vol. 79 (May 1989), pp. 108–12.

Friedman, Milton. *A Theory of the Consumption Function.* Princeton: Princeton University Press, 1957.

Fuhrer, Jeffrey C. "Do Consumers Behave as the Life-Cycle/Permanent-Income Theory of Consumption Predicts?" *New England Economic Review,* September-October 1992, pp. 3–14.

Graham, John W., and Roy H. Webb. "Stocks and Depreciation of Human Capital: New Evidence from a Present Value Perspective," *Review of Income and Wealth,* June 1979, pp. 209–24.

Greenwood, Jeremy, and Boyan Jovanovic. "Financial Development, Growth, and the Distribution of Income," *Journal of Political Economy,* vol. 98 (October 1990), pp. 1076–1107.

Hicks, J. R. *Value and Capital,* 1939, reprinted, Oxford: Oxford University Press, 1968.

Holloway, Thomas M. 1989. "Present NIPA Saving Measures: Their Characteristics and Limitations," in Robert E. Lipsey and Helen Stone Tice, eds., *The Measurement of Saving, Investment, and Wealth.* Chicago: University of Chicago Press, 1989.

Jorgenson, Dale W., and Barbara M. Fraumeni. "The Accumulation of Human and Nonhuman Capital," in Robert E. Lipsey and Helen Stone Tice, eds., *The Measurement of Saving, Investment, and Wealth.* Chicago: University of Chicago Press, 1989.

Kennickell, Arthur, and Janice Shack-Marquez. "Changes in Family Finances from 1983 to 1989: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, vol. 78 (January 1992), pp. 1–18.

Keynes, John Maynard. *The General Theory of Employment, Interest and Money,* 1936, reprinted, New York: Harbinger, 1964.

King, Robert G., and Ross Levine. "Finance and Growth: Schumpeter Might be Right," Working Paper WPS 1083, World Bank, 1993; forthcoming in *Quarterly Journal of Economics.*

Lucas, Robert E. "On the Mechanics of Economic Development," *Journal of Monetary Economics,* vol. 22 (July 1988), pp. 3–42.

Maddison, Angus. *Dynamic Forces in Capitalist Development: A Long-Run Comparative View.* Oxford: Oxford University Press, 1991.

Mankiw, N. Gregory, David Romer, and David N. Weil. "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* (May 1992), pp. 407–37.

Paulus, John D. "Bust and Boom in the Stock Market: Prospects for the 1990s," *Challenge,* January/February 1991, pp. 34–39.

Ritter, Lawrence S. "The Flow of Funds Accounts: A Framework for Financial Analysis," in William E. Mitchell, John H. Hand, and Ingo Walter, *Readings in Macroeconomics: Current Policy Issues.* New York: McGraw-Hill, 1974.

Sargent, Thomas J. *Dynamic Macroeconomic Theory.* Cambridge, Mass.: Harvard University Press, 1987.

Solow, R. M. *Growth Theory: An Exposition.* Cambridge, England: Oxford University Press, 1969.

Summers, Lawrence H. "Tax Policies for Increasing Personal Saving," in Charls E. Walker, Mark A. Bloomfield, and Margo Thorning, eds., *The U.S. Savings Challenge.* Boulder: Westview Press, 1990.

Webb, Roy H. "Forecasts of Inflation from VAR Models." Paper presented to the Western Economic Association International Conference, San Francisco, July 12, 1992.

Whited, Toni M. "Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data," *Journal of Finance,* vol. XLVII (September 1992), pp. 1425–60.