Money and the U.S. Economy in the 1980s: A Break From the Past?

Lawrence J. Christiano*
Economist
Research Department
Federal Reserve Bank of Minneapolis

There is a widespread perception that the relationship between money and the U.S. economy has changed in the 1980s. This view is based in part on the belief that the relationship must have changed because the Federal Reserve changed its operating procedure in 1979 and the financial industry began to be deregulated in 1980. It is also based on the perceived unusual recent behavior of the velocity of money (the rate that dollars are spent, as measured by the ratio of the dollar value of output in the economy to the quantity of money).

My purpose here is to investigate this view by quantitatively assessing the changes that have occurred since 1979 between money and four other macroeconomic variables: the industrial production index, the consumer price index, the three-month return on U.S. Treasury bills, and the trade-weighted value of the dollar. These variables represent the major categories of economic activity: output, inflation, financial markets, and foreign trade.

I find surprisingly weak evidence of a change in the relationship between money and the four macroeconomic variables so far in the 1980s. The results are sensitive to how money is measured and how the data are modeled.

Because how to measure money is a controversial question, I have done the entire analysis for six different definitions of money. When money is measured by the St. Louis Federal Reserve Bank's measure of the monetary base, MB, or by MQ, a definition proposed by Spindt (1985), I find no evidence of a change in the relationship between money and the four macroeconomic variables. I do find some statistically significant evidence of a shift when I test four other measures of money: the Federal Reserve Board's M1 and M2 and two definitions advocated by Barnett, MS1 and MS2 (in Barnett 1980; Barnett and Spindt 1979; and Barnett, Offenbacher, and Spindt 1984). The evidence is in the form of changes in the relation between money and interest rates and in the time response of inflation to a change in the money supply.

As often occurs in the analysis of data, my results depend on the particular model used. One model I use is a version of what is known as the *difference stationary* (DS) model. This model specifies that period-to-period changes in variables fluctuate in a similar way, about a constant mean throughout a data set.[1] Based on data from 1900 to 1970, Nelson and Plosser (1982) claim that the DS model is a good one for U.S. macroeconomic data. Since Nelson and Plosser wrote before the data of the 1980s were available, the fact that this model works well with these data vindicates their claim. However, when a *trend stationary* (TS) model is fit to the data, the widespread perception of a 1980s change in the relationship between money and the economy is confirmed, even if money is measured by

*I thank Lars Ljungqvist, now a visiting scholar at the Hoover Institution, Stanford University, for helpful research assistance.

[1] A *period* is the data sampling interval, which can be a month, a quarter, or a year. Changes, or differences, can be in terms of the logarithm of the data, in which case they represent growth rates.

MB or MQ. The TS model specifies that the data themselves—not their monthly changes—display similar fluctuations about a constant trend. I base my conclusions on the results of the DS model because there is evidence that it better represents U.S. data: its out-of-sample forecasts are superior.

My results do not definitively answer the question in the title of this paper. Perhaps changes in the relationship between money and the economy have occurred, but will not be detectable to the techniques I use until more data are available. Perhaps the TS model is right after all, and the change in the relationship it finds is really there, but not detectable to the DS model. What my results do indicate is that it is not a foregone conclusion that the relationship is different in the 1980s than it was in the 1970s.

Two Views of Velocity

As evidence that the relationship between money and the economy has broken down, analysts commonly point to the recent behavior of the velocity of money. Yet any attempt to draw inferences from data necessarily involves a model of that data, so those who claim that the recent behavior of velocity is unusual have a particular model in mind. There is another model, however, which fits the velocity data better, but which does not support the conclusion that the 1980s behavior of velocity is unusual. In other words, whether or not the behavior of velocity in the 1980s looks unusual depends on the model of velocity used.

The velocity of money in 1970–85 is plotted in Chart 1.[2] Those who see a break in velocity try to characterize the data as fluctuating about a smooth curve. Clearly, smooth curves drawn through the data of the 1970s and the first half of the 1980s look very different. This difference is what many cite as evidence of a shift in the behavior of velocity in the 1980s and as symptomatic of a more general breakdown in the relationship between money and the economy. An alternative possibility is that the sharp change simply reflects the effects of trying to fit the wrong model to the data.

Devising a simple structure that characterizes a set of data is what statisticians refer to as *fitting a model to data.* This can be done informally, by visually seeking a pattern in a graph of data, or formally, by constructing an explicit mathematical model and assigning values to its parameters (coefficients representing the relationships between the model's variables). The model alluded to in the last paragraph is a TS model. Evidently, such a model does not fit the 1970s and 1980s data on velocity well. Nelson and Plosser (1982) argue that the

Charts 1 and 2

Has Velocity Changed in the 1980s?

Chart 1 Yes

Levels and Trends of Velocity*

January 1970–November 1985

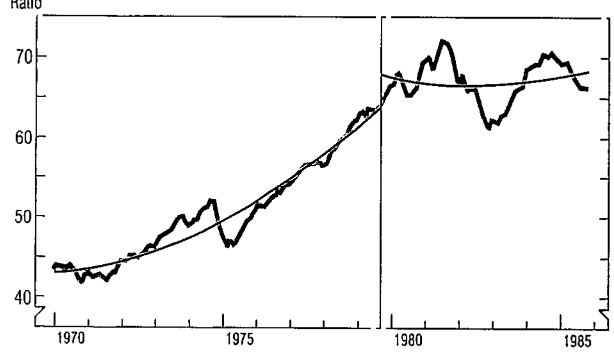
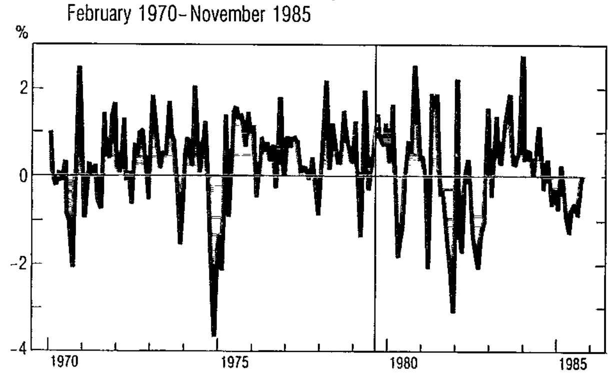


Chart 2 No

Changes in the Log of Velocity*

February 1970–November 1985



*Velocity = (industrial production index × consumer price index)/M1. The trend lines are the results of fitting the data to a quadratic function of time for January 1970–September 1979 and October 1979–November 1985. The data are split at a time that the Fed changed its operating procedure

Sources of basic data. Federal Reserve Board of Governors, U.S. Department of Labor

[2]In calculating velocity for Chart 1, I measure money using the Federal Reserve Board's M1. The nominal value of output is proxied by the product of the industrial production index (IP) and the consumer price index (CPI). Here, then, velocity = (IP×CPI)/M1. My measure of velocity is somewhat unconventional, since velocity is usually measured as the ratio of nominal gross national product (GNP) to M1. I use a proxy for nominal output instead because this study is based on monthly data and data on nominal GNP are not available monthly. The analysis would be unaffected if it were based on quarterly observations of the conventional measure of velocity. I choose not to use those in order to preserve comparability of the data. The two trend lines in Chart 1 are the results of fitting the velocity data to a quadratic function of time for the periods from January 1970 to September 1979 and from October 1979 to November 1985.

TS model is not well suited to most U.S. macroeconomic data. Their conclusions are based on annual U.S. data from 1900 to 1970—observations well before the supposed 1980s breakdown. They present evidence in favor of the DS model, which they regard as better. This model comes from a respected tradition in statistics and is also referred to as an *autoregressive integrated moving average* (ARIMA) model. The influential book by Box and Jenkins (1970) stimulated the widespread application of this model by statisticians in the early 1970s.

A DS model for a particular data series is one which says that the period-to-period changes of (possibly the logarithms of) the data display a tendency to revert to a constant mean, exhibit a roughly constant degree of persistence in deviations from the mean, and fluctuate with a roughly constant amplitude. Such a process is called *covariance stationary.* Thus, the DS model says that if the data are in changes, then they exhibit covariance stationarity.

When a DS model is fitted to velocity data, it does quite well. This can be seen informally in Chart 2, which shows the monthly change in the logarithm of velocity between 1970 and 1985. According to Chart 2, the fluctuations about the mean of velocity growth in the 1980s closely resemble those of the 1970s. In particular, the amplitude of the deviations of velocity growth from the mean as well as the persistence of those deviations appear similar in the two periods.

The apparent shift in trend in Chart 1 is not a puzzle from the perspective of the DS model. Data generated by a DS model are known to display nonrandom patterns and trends. However, these trends have no significance and are expected to undergo shifts of the kind seen in Chart 1. Gould and Nelson (1974) make this point in their analysis of the apparent break in the trend of velocity before and after World War II. They provide a particular DS model that is consistent with what appears to have been a switch from a downward to an upward trend.

My analysis reveals that although U.S. data on velocity appear to behave quite differently in the 1970s and 1980s if the data are interpreted using the TS model, the 1980s behavior of that data does not seem unusual if they are interpreted using the DS model. For this reason, I conclude that the velocity data alone do not provide persuasive evidence that any change has occurred in the relationship between money and the economy in the 1980s.

A Formal Analysis

Now I will extend the investigation by incorporating more data and using formal statistical techniques. I do this to be confident that the impressions gained from Charts 1 and 2 are not the result of including too little data in the analysis. I emphasize formal statistical techniques because simple, revealing graphical representations of the dynamic interactions among many data series are difficult to devise. Generally, the conclusions reached above survive greater scrutiny. Nevertheless, some evidence of a breakdown in the relationship between money and the economy does emerge for most definitions of money considered. The exceptions are, again, MB and MQ, which survive all the tests I devise.

The Data and the Models

The variables I use in the formal analysis are the industrial production index (IP), the consumer price index (CPI), the three-month Treasury bill rate (R), the trade-weighted value of the dollar ($), and six definitions of money: MB, M1, M2, MQ, MS1, and MS2 (described in the accompanying box). The data on these variables include the 191 months from January 1970 to November 1985. I do not use earlier data because some of the monetary aggregates are not available for months before January 1970. Later data were not available when most of this research was done. My sources are the U.S. Department of Labor for the CPI, the Federal Reserve Bank of St. Louis for MB, and the Federal Reserve Board for the rest (an unpublished data base at the Board for MQ, MS1, and MS2).[3]

As before, two types of model are used throughout this analysis: TS models and DS models. Each model includes six variables: the four nonmoney variables listed above and one of the monetary aggregates. Because I use six monetary aggregates, there are six TS models and six DS models. Each TS and DS model is distinguished by the particular monetary aggregate included. Thus, *TS(M1)* denotes the TS model with the M1 definition of money; *DS(M1)*, the DS model with M1 in it. The generic symbol for a monetary aggregate is *m*. In this way, I study a total of twelve models: TS(*m*) and DS(*m*) for *m* = MB, M1, M2, MQ, MS1, and MS2.

A TS(*m*) model is defined as a vector autoregression (VAR) in the logarithm of IP, CPI, $, and *m* and the level of R.[4] In addition, each equation of the VAR

---

[3]The data on IP, the CPI, and the monetary aggregates are seasonally adjusted. All data except those for the CPI are seasonally adjusted by the source. The CPI is adjusted as described in Amirizadeh 1985, pp. 9–13.

[4]For a description of VARs and how they are estimated and used, see Sargent 1979 and Sims 1980. I also did the analysis for VARs specified without a trend and with a quadratic trend. The conclusions are the same as those I report here.

Six Measures of Money

All six monetary aggregates tested in the accompanying study are attempts to define and measure the funds the public has readily available for spending. As the table here shows, the aggregates differ somewhat in the types of financial assets they count as money. But they differ more in the way they weight each asset:

- The *monetary base,* MB, and the *simple sum* aggregates, M1 and M2, simply sum the monthly values of their assets, which in effect gives each the same weight (one).*
- The *monetary transactions* aggregate, MQ, weights its assets by the share of total spending on goods and services they are involved in each month.
- The *monetary services* aggregates, MS1 and MS2, weight each of their assets by a measure of the cost of the service it provides: the difference between the market interest rate and the asset's own interest rate.

For further details about these aggregates, see Gilbert 1983, 1985; FR Board, various dates; Spindt 1985; Batten and Thornton 1985; Barnett 1980; Barnett and Spindt 1979; Barnett, Offenbacher, and Spindt 1984; and Lindsey and Spindt 1986.

The Assets in the M's

<table>
<tr><td>Depository Institution Reserves</td><td rowspan="2">MB</td><td></td><td></td><td></td></tr>
<tr><td>Currency</td><td rowspan="6">M1 and MS1</td><td rowspan="9">MQ</td><td rowspan="13">M2 and MS2</td></tr>
<tr><td>Demand Deposits</td><td rowspan="12"></td></tr>
<tr><td>NOW Accounts</td></tr>
<tr><td>ATS Accounts</td></tr>
<tr><td>Credit Union Share Drafts</td></tr>
<tr><td>Travelers' Checks</td></tr>
<tr><td>Money Market Deposit Accounts</td><td rowspan="7"></td></tr>
<tr><td>Money Market Mutual Fund Shares (Noninstitutional)</td></tr>
<tr><td>Savings Deposits Subject to Telephone Transfer</td></tr>
<tr><td>Savings Deposits Not Subject to Telephone Transfer</td><td rowspan="4"></td></tr>
<tr><td>Small Time Deposits</td></tr>
<tr><td>Overnight Repurchase Agreements</td></tr>
<tr><td>Overnight Eurodollars</td></tr>
</table>

*The base includes a factor added to take account of changes in reserve requirements. The St. Louis Fed's measure of the base (used in this study) accounts for such changes differently than the Federal Reserve Board's. The two measures also differ in the ways they adjust for seasonal influences and treat cash in the vaults of depository institutions. See Gilbert 1983.

includes a constant and linear trend term and six lags of each of the variables.[5] The reason for using VARs to capture the stationary part of the TS model is that they can approximate virtually all covariance stationary processes. Also, the parameters of a VAR are extremely easy to estimate, since estimation is done using ordinary least squares on each equation separately.

To construct a DS(m) model, I begin by differencing the variables in the TS model enough times so the resulting variables appear to be covariance stationary. This implies differencing the log of each of IP, \$, and m once and the log of CPI twice. The level of R is differenced once. Each DS(m) model is a six-lag VAR with constant term fit to the differenced data.[6]

The Trend Stationary Model

In several ways I test the proposition that the same TS

[5]I test the indicated lag length specification using the chi-square test described in Sims 1980, p. 17. In calculating the test statistic, I incorporate an adjustment to compensate for the tendency for the test statistic to reject the null hypothesis too often. (See note 7 for further details.) For each model, the lag length is six under the null hypothesis and ten under the alternative. The area under the chi-square distribution to the right of the test statistic is 0.41, 0.0009, 0.20, 0.13, 0.0002, and 0.090 for TS(m), m = MB, M1, M2, MQ, MS1, and MS2, respectively. Evidently, there is evidence against the null hypothesis in the cases of M1 and MS1. Examination of the individual VAR equations shows that the rejections are probably due to too short a lag length on the interest rate equation. I decided to work with the six-lag specification anyway because there is some evidence that the TS model is overparameterized when the lag length increases from six to seven. In particular, for most variables and forecast horizons, the root mean square error (RMSE) of out-of-sample forecasts at horizon 1–12 deteriorates with the addition of one lag. (The method of calculating the RMSEs is described later.) Exceptions are the long-run interest rate forecasts generated by TS(MQ), TS(M1), and TS(MS1). Among these, the greatest improvements are TS(M1)'s. For this model, when the lag length is increased from six to seven, the RMSEs fall from 5 to 10 percent for forecasts at horizons 7–12. At the short horizons, however, the model's RMSEs increase.

[6]Again, I test the indicated lag length specification using the procedure described in note 5. For each model, the lag length is six under the null hypothesis and ten under the alternative. The area under the chi-square

model, with unchanged parameter values, can explain both the 1970s and the 1980s. A test of this kind is called a *stability test* because it is a test of whether or not the parameter values—that is, the relationships in the data—are the same in the two periods. For practical purposes I must pick a particular date on which a possible shift occurred. I pick October 1979, a month in which the Fed changed the way it conducts monetary policy.

The null hypothesis in my first test is that all of the TS model parameters are stable. Results of a chi-square test for this null hypothesis are reported in Table 1.[7] I do not report the actual test statistics. Instead, Table 1 shows the area under the chi-square distribution to the right of the test statistic, the area called the *significance level* of the test. It is the probability of rejecting the null hypothesis if it is true. Clearly, when the significance level is very low, one has little reason not to reject the null hypothesis. For example, consider the number 0.00000115 for the TS(M1) model. It means that if the estimated TS(M1) model is exactly true, the probability of getting a test statistic as large as or larger than the one actually computed is only about 1 in 1 million. Obviously, rejecting the null hypothesis when the significance level is 0.00000115 is a pretty safe bet.

Of the TS test results in Table 1, the ones most favorable to the null hypothesis are those for TS(MB) and TS(MQ). Nevertheless, the significance levels for these tests are fairly small: between 0.01 and 0.02. These significance levels are sufficiently small to raise serious doubts about the validity of the null hypothesis as applied to TS(MB). Altogether, then, the TS model results in Table 1 constitute evidence against the null hypothesis that the relation between money and the economy did not change between the 1970s and the 1980s.[8]

The reliability of the chi-square test is suspect when the number of observations is small. (See Sims 1980, p. 17.) In such a case—which may include the present one—decisions should be based not only on the outcome of a chi-square test, but also on the results of other tests which one hopes do not share the chi-square's possible weakness. I have done several such tests, and they tend to confirm the strong suspicions raised by the chi-square results.

One particularly dramatic example is the one-step-ahead monthly inflation forecast errors plotted in Chart 3. Those errors are computed using the TS(MB) model estimated with data from January 1970 to September 1979. As a result, the post-September 1979 numbers are out-of-sample forecast errors. The chart indicates that the TS(MB) model repeatedly overpredicts inflation in the 1980s. At an annual rate, the average overprediction between October 1979 and November 1985 is a whopping 14.7 percentage points. (The standard error is 7.6 percentage points.)These findings, in addition to others not reported here, are consistent with Chart 1's impression that the TS model is not stable.

Table 1

Overall Model Tests of Stability

Model's Monetary Aggregate	Significance Level* for Stationary Model	
	Trend	Difference
MB	.01750000	.68
M1	.00000115	.43
M2	.00270000	.17
MQ	.01300000	.74
MS1	.00000016	.54
MS2	.00560000	.59

*Based on chi-square tests, this is the probability (÷100)—under the null hypothesis of stability—of getting a value of the chi-square statistic greater than the one computed.

distribution to the right of the test statistic is 0.59, 0.30, 0.27, 0.27, 0.029, and 0.38 for DS(m), m = MB, M1, M2, MQ, MS1, and MS2, respectively. Thus, except for DS(MS1), the null hypothesis that the lag length is six fails to be rejected at a very high significance level. The test statistic for DS(MS1) is somewhat quirky, since the significance level of the test of the six-lag versus the seven-lag model is 0.775, and for the six- versus eight-lag model it is 0.131. Because of these test results, I use the six-lag specification for DS(MS1).

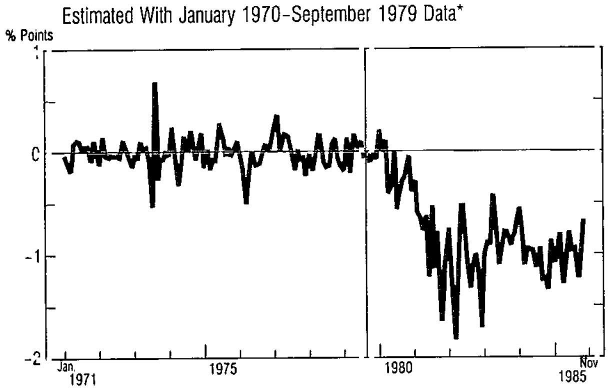
7My method for computing the chi-square test is the one described in Sims 1980, p. 17. In particular, for each TS model I specify an unrestricted TS with dummy variables which let all the coefficients in the model take on different values after September 1979. I then compute D_u, the matrix cross product of residuals from the estimated unrestricted TS. D_r is the same matrix for the restricted TS in which the pre- and post-October 1979 coefficients are restricted to be the same. The test statistic is then $(T-c)(\log|D_r| - \log|D_u|)$, where $|\cdot|$ denotes the determinant of the indicated 5×5 matrix, *log* denotes the natural logarithm, T is the number of observations (183), and c is the number of variables in each unrestricted regression (64). The parameter c is introduced to correct for a presumed small sample bias in favor of rejecting the null hypothesis. (See Whittle 1953 and Lissitz 1972.) Under the null hypothesis, the test statistic is approximately a realization from a chi-square random variable with 160 degrees of freedom. This is the difference in the number of estimated parameters between the restricted and unrestricted TSs.

8Stability tests on the individual equations indicate that the main source of instability is in the interest rate equation.

Chart 3

A Trend Stationary Model's Errors in Forecasting Inflation

One-Month-Ahead Forecast Errors of TS(MB) Model Estimated With January 1970-September 1979 Data*



*A forecast error is the actual value less the predicted value.

Table 2

Individual Equation Tests of Stability

	Significance Level* for Difference Stationary Model					
Equation	MB	M1	M2	MQ	MS1	MS2
Output	.68	.77	.68	.90	.80	.84
Prices	.68	.70	.44	.80	.65	.46
Money	.26	.69	.30	.43	.66	.53
Interest Rate	.29	.02	.02	.13	.01	.07
Exchange Rate	.51	.16	.25	.43	.24	.17

*This is the result of a Chow test of the null hypothesis of no structural change in October 1979.

The Difference Stationary Model

Results of chi-square tests for constancy of all the coefficients in the DS models are also reported in Table 1.[9] Like those for the TS models, these numbers are the significance levels of the test rather than the actual values of the test statistic.

Comparison of the TS and DS results reveals a striking difference. The very high significance levels reported for the DS models indicate that the values of the computed test statistics are quite plausible, under the assumption that the null hypothesis is true. Put differently, the results indicate that the data are consistent with the hypothesis of constancy of all coefficients in the VARs.

Since the DS results in Table 1 are so much at variance with the widespread perception of instability in the 1980s, it seems prudent to do further tests of parameter stability in the DS models. In doing so, I hope to show that the results are not an artifact of a possible deficiency in the chi-square tests.

I do two more classes of tests. First, I test each individual equation in the six VARs. The results basically corroborate the findings in Table 1. One difference is that some parameter instability is evident in the interest rate equation in all the DS models except those with MB and MQ. The second class of test investigates the stability of what some might regard as interesting functions of the VAR parameters. My choice of functions is motivated by the fact that a generally accepted important use for the monetary aggregates is to provide an indication about subsequent developments in inflation. These tests show evidence of instability in DS(M1), DS(MS1), and DS(MS2).

Table 2 presents the results of performing a Chow test for structural stability on each equation of each DS model.[10] Instead of the actual F-statistics for the test, Table 2 presents the corresponding significance levels. For example, the first value in the table is 0.68. That is the area under the F-distribution to the right of the computed F-statistic.

With one exception, the results are uniform across all models. The null hypothesis of no structural change fails to be rejected at the 15 percent significance level in all but the interest rate equation. That equation shows

[9]My method for computing the chi-square test for the DS models is, with some modifications, the same as that for the TS models. (See note 7.) The modifications are that in this case the value of c is 62 and the test statistic is chi-square with 155 degrees of freedom under the null hypothesis. The reason for these modifications is that the DS model has fewer parameters than the TS model because it has no trend.

[10]The Chow test for any given equation is done by first estimating the equation over three sample periods: January 1970–September 1979, October 1979–November 1985, and January 1970–November 1985. I denote the sum of squares of the fitted residuals for each of these by s_1, s_2, and s_3, respectively. The test statistic, s, is computed as $s = [(s_3 - s_1 - s_2)/s_3](d/n)$, where $d = 152$ and $n = 31$. Under the null hypothesis, s is asymptotically a realization from an F-distribution with n numerator degrees of freedom and d denominator degrees of freedom. For a further discussion of the Chow test, see Dhrymes 1978, p. 62.

evidence of instability for all the aggregates except MB and MQ. When either of those is in the equation, it shows no evidence of structural instability, even at the 10 percent significance level.

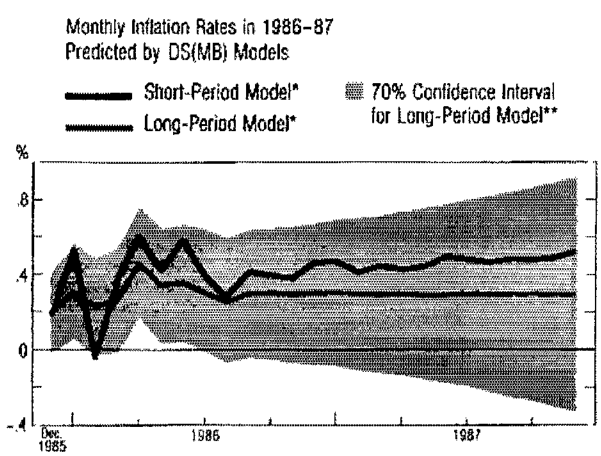
□ *Money and Inflation*

The following tests focus on several particular functions of the DS parameters which mediate the relationship between money and inflation. These tests turn up surprisingly little evidence of instability. In fact, DS(MB), DS(MQ), and DS(M2) show no evidence of instability.

My first test of this type compares the forecasts of inflation in 1986 and 1987 between DS models estimated using data up to September 1979 (what I'll call the *short period*) and up to November 1985 (the *long period*). To make the comparison easier, I also compute a 70 percent confidence interval for the forecast of the models estimated on the long period (an interval within which the model says the variable can be expected to fall with 70 percent probability).[11] The results for VAR(MB) are reported in Chart 4. (The results for the other models look very much like these.) A distinctive feature of this chart is the shape of the confidence interval: it widens as the forecast horizon increases, reflecting increased uncertainty.[12] Moreover, the confidence interval easily contains the inflation forecast based on the short-period model. In this sense, the short- and long-period models do not imply significantly different outlooks for inflation.

Chart 4

Difference Stationary Model
Forecasts of Inflation



*The short-period model is estimated with data from January 1970 to September 1979; the long-period model, with data from January 1970 to November 1985

**According to the long-period model, the odds are 7 in 10 that inflation will fall in this interval

The models' predicted levels of inflation nevertheless look quite different and in ways one might expect. In particular, the short-period models expect more inflation in 1986 and 1987 than do the long-period models. Presumably, this difference is due to the fact that the long-period models are estimated using data from the 1980s as well as the 1970s. Their parameter values therefore take into account the recent years' rapid money growth which has not been associated with a pickup in inflation. The problem with this argument is that it exaggerates the significance of the difference between the two models' forecasts. Compared to the uncertainty in the forecasts themselves, that difference is negligible.

A second money-inflation test for structural stability is to compute a set of inflation forecast errors for the period from January 1971 to November 1985. This is done using DS models estimated over the short period (that is, with data through September 1979). The post-1979 errors are thus out-of-sample forecast errors. If the relationship between money and inflation has changed in the 1980s, then one would expect the errors for the 1980s to look very different from those of the 1970s. This is precisely what the TS models found (recall the dramatic Chart 3).

Chart 5 shows the one-step-ahead inflation forecast errors for the DS model in which the monetary aggregate is MB. (Again, the results for the other models look very much like these.) Note that the amplitude of fluctuations of the errors increases in the early 1980s and then appears to return to its previous level. Yet the mean of these errors does not appear to shift.

It is logically possible for the short-period model to display no mean shift in its one-step-ahead forecast errors in the 1980s and yet for there to be a mean shift at longer horizons. This does not seem to be the case here. Table 3 reports the mean and standard errors for one-, two-, three-, and four-step-ahead forecasts of inflation for the DS(MB) model estimated over the short period.

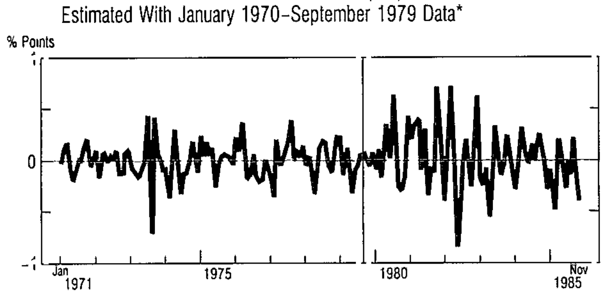
[11] The computed confidence intervals are too narrow for two reasons. One is that, in calculating them, I ignore parameter uncertainty. The other is that the formulas used in those calculations assume the disturbances in the DSs are normally distributed, whereas in fact they are probably fat-tailed. Taking these factors into account would simply have reinforced my conclusion that the difference in the forecasts between short- and long-period DSs is not statistically significant.

[12] Technically, the confidence interval widens because the VAR I use specifies the first difference of inflation to be covariance stationary.

Chart 5

A Difference Stationary Model's Errors in Forecasting Inflation

One-Month-Ahead Forecast Errors of DS(MB) Model Estimated With January 1970–September 1979 Data*



*A forecast error is the actual value less the predicted value.

Table 3

A Closer Look at the Difference Stationary Model's Errors in Forecasting Inflation

	Short-Period DS(MB) Model's Errors in Forecasting*			
	January 1971–September 1979		October 1979–November 1985	
Forecast Horizon	Mean	Standard Error	Mean	Standard Error
1	0**	2.02	.14	3.48
2	0	2.13	.33	3.80
3	0	2.29	.34	3.82
4	0	2.51	.34	4.12

*All figures are percentage points, at annual rates.
**The mean at horizon 1 is zero because the estimation method used (ordinary least squares) sets it to zero.

(The results for the other DS models are similar.) According to the table, the mean forecast error in the period after September 1979 rises slightly with the forecast horizon: from 0.14 to 0.34 percentage points, at an annual rate. However, relative to the average inflation rate over this period, and to the standard errors, the shift is negligibly different from zero. (Of course, the increased variance of the inflation forecast errors observed in Chart 5 is also reflected in Table 3.) Thus, the inflation forecast errors show no evidence of a shift in the relation of inflation to the other four macroeconomic variables, although the magnitude of the shocks impinging on inflation may have increased for a while in the early 1980s.[13]

For a last test of stability, Charts 6–11 each report two impulse response functions. These summarize the historical average response of inflation to a shock in money growth, as implied by the relevant short- and long-period DS models.[14] The shock is defined as a one standard deviation disturbance in money that is unpredictable given past values of all the variables of the model and contemporaneous values of industrial production and inflation.[15] The charts also show a 70 percent confidence interval based on the long-period model.

The results of this test split the DS models into two groups. Those using three of the definitions of money—M1, MS1, and MS2—show substantial evidence of instability. In each of Charts 6–8, the impulse response based on the short period is much higher or lower than that for the long period, so much so that it is outside the long-period's confidence interval.[16] But models using the other three definitions—MB, M2, and MQ—show substantial evidence of stability. In Charts 9–11, the short-period impulse responses are within the long-period's confidence interval, and for DS(M2) the two periods' responses are virtually identical.

[13]Formally, what I have in mind is the following. The DS model can be written as $Y_t = A_0 + E\epsilon_t + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_6 Y_{t-6} + (\epsilon_t - E\epsilon_t)$, where Y_t is a vector of my five (suitably differenced and logged) macroeconomic variables; A_0 is a five-element vector; and A_i is a (5×5)-element matrix for $i = 1, \dots, 6$. The dynamic interaction between the Y's is controlled by A_i, $i = 0, 1, 2, \dots, 6$, and $E\epsilon_t$. If none of these shift—and the data appear consistent with this view—then I say the dynamic interaction between the macroeconomic variables has not changed; in particular, the relation between money and the economy has not changed. Unfortunately, constancy of the variance of ϵ_t is a maintained hypothesis underlying the systemwide stability tests in Table 1 and the Chow tests in Table 2. I am uncertain how the possible failure of this maintained hypothesis might have affected those tests.

[14]For each DS model, the impulse response function and its 70 percent confidence interval are computed using Monte Carlo simulation methods. This involves drawing 2,000 times from the estimated asymptotic normal distribution of the DS model parameters. Details of the procedure are essentially as described in example 17.1 in Doan and Litterman 1984. I define the 70 percent confidence interval for an impulse response at a given lag as that interval that leaves 15 percent of the probability in each tail. A complication is the fact that the price term in the DS models is the change in inflation rather than its level. To get the impulse response function from money shocks to inflation, I first get the function for changes in inflation and then cumulate them.

[15]For a discussion of a shock in VAR models and the variety of ways it can be defined, see Litterman 1984.

[16]Note that the impulse response function is significantly different from zero at virtually all lags. This contrasts with the impulse response function relating a shock in money to the change in inflation, which I do not report here. That function is not significantly different from zero for DS(M1) estimated over the long period. The reason for the difference is that adjacent elements in the function are negatively correlated. Consequently, the cumulative sum of these objects is relatively precisely estimated. The cumulative sum is just the impulse response of inflation to money shocks.

Charts 6–11

Responses of Inflation to a Money Shock in the Difference Stationary Models†

Short-Period Model* | 70% Confidence Interval for Long-Period Model** | Long-Period Model*

Chart 6 M1

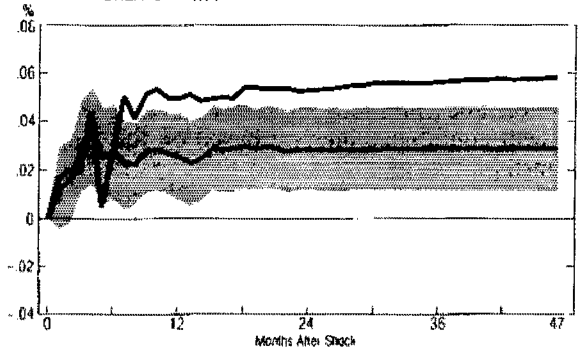


Chart 9 MB

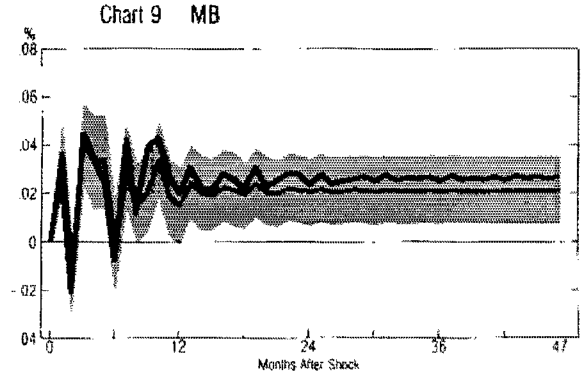


Chart 7 MS1

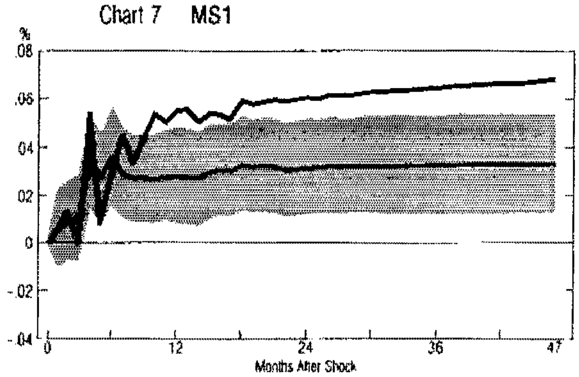


Chart 10 M2

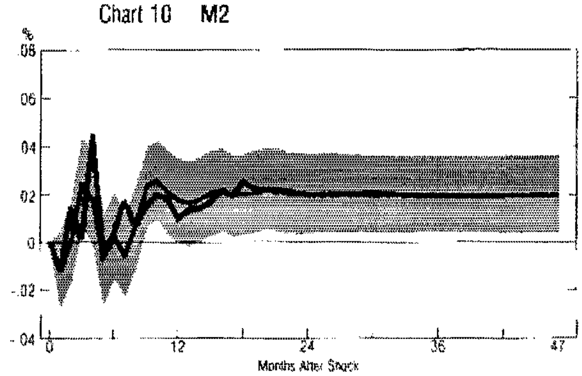


Chart 8 MS2

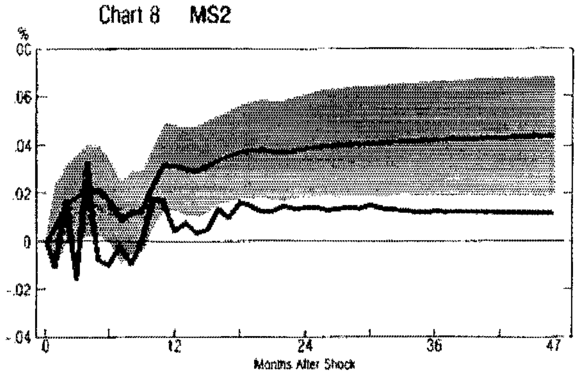
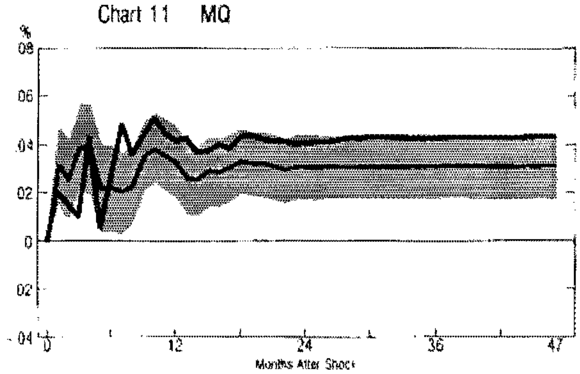


Chart 11 MQ



† The shock is a one standard deviation disturbance in money that is unpredictable from past values of all the model's variables and contemporaneous values of industrial production and inflation.

*The short-period model is estimated with data from January 1970 to September 1979; the long-period model, with data from January 1970 to November 1985

**According to the long-period model, the odds are 7 in 10 that the inflation response will fall in this interval

Which View is Right?

Thus, statistical tests essentially confirm what the informal examination of Charts 1 and 2 suggested. In particular, if data on money and the economy are interpreted using a DS model, then the dynamic relationship between these variables is surprisingly similar in the 1980s to what it was in the 1970s. Moreover, two monetary aggregates, MB and MQ, show no evidence of a change in their relationship to other variables between the 1970s and 1980s. However, interpreting the data using a TS model leads to the conclusion that the relationship between money and the

Table 4

Ratio of Stationary Model Forecasting Errors: Trend/Difference*

	Forecasted Variable: Monthly Growth in			Forecasted Variable: Monthly Level of		
Monetary Aggregate	Output	Prices	Money	Interest Rate	Log of Exchange Rate	Horizon
MB	1.03	.99	1.05	1.05	1.03	1
	1.07	.96	1.00	1.08	1.08	2
	1.19	1.00	1.01	1.09	1.10	3
	1.16	.96	.98	1.10	1.10	4
	1.10	1.04	1.03	1.12	1.12	5
	1.22	1.06	1.15	1.09	1.13	6
	1.31	1.08	.97	1.06	1.15	7
	1.25	1.07	1.05	1.06	1.15	8
	1.21	1.09	1.05	1.07	1.18	9
	1.14	1.12	1.02	1.03	1.23	10
	1.15	1.08	1.26	1.00	1.28	11
	1.21	1.12	1.07	1.00	1.31	12
M1	1.07	1.04	1.00	1.08	1.02	1
	1.08	1.02	1.07	1.05	1.08	2
	1.11	1.05	1.12	.99	1.11	3
	1.12	1.07	1.08	.96	1.11	4
	1.25	1.08	1.02	.93	1.12	5
	1.27	1.19	1.02	.93	1.14	6
	1.21	1.13	1.03	.94	1.16	7
	1.09	1.08	1.08	.95	1.17	8
	1.11	1.14	1.07	.97	1.17	9
	1.09	1.12	1.05	.98	1.18	10
	1.03	1.13	1.08	1.01	1.19	11
	1.08	1.11	1.05	1.05	1.19	12
M2	1.12	1.00	1.01	1.02	1.03	1
	1.21	1.02	1.07	1.03	1.10	2
	1.29	1.07	1.06	1.02	1.15	3
	1.26	1.06	1.02	1.04	1.17	4
	1.28	1.11	1.01	1.08	1.19	5
	1.28	1.20	1.03	1.09	1.21	6
	1.25	1.20	1.09	1.12	1.21	7
	1.14	1.27	1.17	1.14	1.21	8
	1.06	1.29	1.19	1.18	1.21	9
	1.05	1.30	1.22	1.20	1.20	10
	1.06	1.29	1.28	1.21	1.18	11
	1.05	1.25	1.25	1.21	1.16	12
MQ	1.07	1.03	.98	1.05	1.03	1
	1.07	.99	1.03	1.08	1.07	2
	1.06	.99	1.07	1.08	1.10	3
	1.06	1.03	1.05	1.09	1.10	4
	1.13	1.02	1.02	1.08	1.11	5
	1.20	1.10	1.00	1.07	1.12	6
	1.15	1.05	1.00	1.08	1.14	7
	1.01	1.00	1.02	1.10	1.16	8
	1.02	1.05	1.00	1.12	1.17	9
	1.01	1.06	1.00	1.13	1.19	10
	.98	1.02	1.02	1.14	1.21	11
	1.07	1.05	1.01	1.15	1.22	12
MS1	1.14	1.09	1.03	1.08	1.04	1
	1.12	1.03	1.08	1.07	1.12	2
	1.07	1.08	1.10	1.04	1.16	3
	1.12	1.11	1.09	1.03	1.17	4
	1.17	1.15	1.05	1.02	1.18	5
	1.20	1.24	1.04	1.03	1.19	6
	1.18	1.16	1.03	1.02	1.20	7
	1.07	1.14	1.06	1.00	1.21	8
	1.07	1.13	1.08	1.00	1.22	9
	1.06	1.15	1.04	1.00	1.23	10
	1.03	1.09	1.03	1.02	1.26	11
	1.07	1.18	1.02	1.06	1.26	12
MS2	1.15	1.06	1.02	.99	1.02	1
	1.25	1.05	1.05	.99	1.09	2
	1.31	1.09	1.02	.98	1.14	3
	1.35	1.07	.98	1.00	1.17	4
	1.35	1.11	1.00	1.03	1.20	5
	1.32	1.20	1.00	1.04	1.21	6
	1.32	1.23	1.05	1.08	1.21	7
	1.23	1.26	1.12	1.14	1.19	8
	1.19	1.36	1.20	1.22	1.16	9
	1.24	1.41	1.25	1.29	1.12	10
	1.18	1.45	1.29	1.36	1.09	11
	1.19	1.43	1.30	1.42	1.04	12

*This is the ratio of the models' root mean square errors.

economy is quite different in the 1980s from what it was in the 1970s.[17] Which of these conclusions is closer to the truth?

An appealing way to answer this question is to compare the two models' out-of-sample forecasting performance. Here, DS does better than TS, suggesting that it better captures the dynamics in the data.[18]

For my forecasting test, I compute ratios of out-of-sample root mean square errors (RMSEs) for forecasts at horizon $k = 1, 2, 3, \ldots, 12$. I then calculate the ratio of the RMSEs for the TS model to those for the DS model. In this way, a value for this ratio greater than one represents superior performance for the DS model.

Before presenting the results, I need to clarify my procedure. Let ${}_tx(k, m)$ denote the error in forecasting the date $t + k$ value of some variable x as of date t using a model (either DS or TS) estimated with data from January 1970 to month t. Here, x can take on the following possible values: ip, infl, R, and exch, where *ip* denotes monthly industrial production index growth; *infl,* monthly inflation, or growth in the consumer price index; *R,* the three-month Treasury bill rate; and *exch,* the log of the trade-weighted value of the dollar. Thus, ${}_{82:1}\text{infl}(3, \text{M1})$ is the error in forecasting March 1982 inflation as of January 1982 using a model with M1 estimated with data from January 1970 to January 1982. These forecast errors are computed for $t = 74{:}12$ to $t = 85{:}10$. I thus obtain for each variable $(132-k)$ observations on k-step-ahead forecast errors, for $k = 1, 2, \ldots, 12$. The RMSEs are calculated as

$$\text{RMSE}(x, k, m) = (132-k)^{-1}\left[\sum_{t=74:12}^{t=85:11-k} {}_tx(k, m)\right]^{1/2}$$

for x = ip, infl, R, exch; $k = 1, 2, \ldots, 12$; m = MB, M1, M2, MQ, MS1, MS2. Table 4 reports the ratio of RMSEs for the TS model to those for the DS model for all values of x, k, and m.

According to Table 4, the DS model dominates the TS model in forecasting at virtually all horizons and for all variables considered. Moreover, in some cases the improvements are substantial. For example, forecasts of infl and R using M2 are around 25 percent more accurate at long horizons if the DS model is used rather than the TS model. The TS model dominates the DS model in only a few places, and there the improvement is always very small, never more than 7 percent. This evidence suggests that the DS model represents a better characterization of the data than the TS model. I therefore take seriously the implications of the DS model and discount the very different implications of the TS model.

Summary and Implications

Informal and formal statistical analyses reveal that the dynamic interactions of money and several other macroeconomic variables are surprisingly similar in the 1980s to what they were in the 1970s. In fact, for two measures of money, MB and MQ, no test uncovers evidence of instability. These conclusions depend on using a model from the difference stationary class; the results are very different when a trend stationary class model is used. Since the difference model is better at out-of-sample forecasting, it may be a better characterization of the data. Therefore, its results with regard to money and the U.S. economy should perhaps be preferred to those of the trend model.

This evidence in favor of difference stationary models may be useful to researchers interested in forecasting macroeconomic variables and constructing models of the macroeconomy. U.S. policymakers may also find the results interesting. In fact, some may want to infer from them that the Fed should set ranges for MB or MQ instead of M1. The evidence does not warrant such an inference, however. The pitfalls of inferring policy implications from statistics such as these without the aid of a structural macroeconomic model have been extensively discussed. (See, for example, Lucas and Sargent 1979 and Sargent 1980.) No such model is presented here.

---

[17]Eichenbaum and Singleton (1986) also report evidence that the TS specification leads to instability, whereas the DS specification does not.

[18]Devising ways of discriminating between DS and TS models is an area of active research. Meese and Geweke (1984) note that existing methods fall into two categories based on the type of goodness-of-fit criteria: within-sample or out-of-sample. Nelson and Plosser (1982) pursue the former; Meese and Geweke (1984) and I, the latter. Meese and Geweke find in favor of TS over DS models, whereas Nelson and Plosser reach the opposite conclusion—as do I. That Meese and Geweke and I disagree is particularly surprising since we use virtually the same out-of-sample goodness-of-fit criteria.

References

Amirizadeh, Hossain. 1985. Technical appendix for the Bayesian vector autoregression model of the U.S. economy. Research Department, Federal Reserve Bank of Minneapolis.

Barnett, William A. 1980. Economic monetary aggregates: An application of index number and aggregation theory. *Journal of Econometrics* 14 (September): 11–48.

Barnett, William A., and Spindt, Paul A. 1979. The velocity behavior and information content of divisia monetary aggregates. *Economics Letters* 4: 51–57.

Barnett, William A.; Offenbacher, Edward K.; and Spindt, Paul A. 1984. The new divisia monetary aggregates. *Journal of Political Economy* 92 (December): 1049–85.

Batten, Dallas S., and Thornton, Daniel L. 1985. Are weighted monetary aggregates better than simple-sum M1? *Review* 67 (June/July): 29–40. Federal Reserve Bank of St. Louis.

Box, George E. P., and Jenkins, Gwilym M. 1970. *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Dhrymes, Phoebus J. 1978. *Introductory econometrics.* New York: Springer-Verlag.

Doan, Thomas A., and Litterman, Robert B. 1984. Regression analysis of time series: User's manual, RATS version 4.30. Minneapolis: VAR Econometrics.

Eichenbaum, Martin, and Singleton, Kenneth J. 1986. Do equilibrium real business cycle theories explain post-war U.S. business cycles? Manuscript. National Bureau of Economic Research.

Federal Reserve Board of Governors (FR Board). Various dates. *Federal Reserve Bulletin.*

Gilbert, R. Alton. 1983. Two measures of reserves: Why are they different? *Review* 65 (June/July): 16–25. Federal Reserve Bank of St. Louis.

__________. 1985. New seasonal factors for the adjusted monetary base. *Review* 67 (December): 29–33. Federal Reserve Bank of St. Louis.

Gould, John P., and Nelson, Charles R. 1974. The stochastic structure of the velocity of money. *American Economic Review* 64 (June): 405–18.

Lindsey, David E., and Spindt, Paul A. 1986. An evaluation of monetary indexes. Special Studies Paper 195. Board of Governors of the Federal Reserve System.

Lissitz, Robert W. 1972. Comparison of the small sample power of the chi-square and likelihood ratio tests of the assumptions for stochastic models. *Journal of the American Statistical Association* 67 (September): 574–77.

Litterman, Robert B. 1984. Forecasting and policy analysis with Bayesian vector autoregression models. *Federal Reserve Bank of Minneapolis Quarterly Review* 8 (Fall): 30–41.

Lucas, Robert E., Jr., and Sargent, Thomas J. 1979. After Keynesian macroeconomics. *Federal Reserve Bank of Minneapolis Quarterly Review* 3 (Spring): 1–16.

Meese, Richard, and Geweke, John. 1984. A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business & Economic Statistics* 2 (July): 191–200.

Nelson, Charles R., and Plosser, Charles I. 1982. Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10 (September): 139–62.

Sargent, Thomas J. 1979. Estimating vector autoregressions using methods not based on explicit economic theories. *Federal Reserve Bank of Minneapolis Quarterly Review* 3 (Summer): 8–15.

__________. 1980. Rational expectations and the reconstruction of macroeconomics. *Federal Reserve Bank of Minneapolis Quarterly Review* 4 (Summer): 15–19.

Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica* 48 (January): 1–48.

Spindt, Paul A. 1985. Money is what money does: Monetary aggregation and the equation of exchange. *Journal of Political Economy* 93 (February): 175–204.

Whittle, P. 1953. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society,* B, 15 (No. 1): 125–39.