

Vector Autoregression Evidence on Monetarism: Another Look at the Robustness Debate

Richard M. Todd
Senior Economist
Research Department
Federal Reserve Bank of Minneapolis
and Associate Director
Institute for Empirical Macroeconomics

Explaining the recurrent fluctuations in prices and quantities known as the *business cycle* is one of the major tasks of macroeconomics. To study these dynamics, macroeconomists construct models, usually systems of equations, which are attempts to describe how key aggregate economic variables like output, the price level, interest rates, and the stock of money evolve and influence each other over time. Among the models that macroeconomists can use to gather evidence on business cycle dynamics is a relatively simple type: *vector autoregressions* (VARs). In these models, each member of a group of random variables is expressed as a linear function of past values of itself, past values of the other members of the group, and nonrandom components, such as constant terms or polynomial functions of time.

Over the last ten years or so, VARs have become popular among economists, especially as tools for routine forecasting. Using VARs to provide evidence on theories of business cycle dynamics is, however, controversial. One reason for this is the belief that empirical results from estimated VARs are not *robust*, or stable.¹ The idea is that VAR statistics are very sensitive to seemingly minor or arbitrary modifications in the VAR's random and nonrandom components. Thus, some believe, a VAR model's evidence on the dynamic relationships among its variables can provide little usable information.

Views on the robustness of VAR evidence have been stimulated in no small part by the work of Christopher Sims (1980b). Using monthly U.S. data for 1947–78,

Sims estimated a VAR with four variables (interest rates, money, the price level, and output) in order to get evidence on the dynamic relationships among these variables, especially the relationship between money and output. One of Sims' main conclusions was unexpected by many: the evidence from his model contradicted a specific version of monetarist theory that was prominent in the 1970s. Sims (1980b, p. 252) highlighted one "strikingly nonmonetarist" result: unpredicted variations associated with the money stock account for only 4 percent of the unpredicted variation in output.

Skeptics have questioned the robustness of Sims' findings and thus VAR evidence in general. Stephen King (1983), David Runkle (1987), and David Spencer

¹ Another reason for the controversy is the belief that some VAR results are only superficially nonmonetarist. To make this point, some economists have developed theoretical and statistical models illustrating how monetary policy could be an important determinant of output even in an economy in which a time series analysis would show little relationship between output and the quantity of money. (See the work of Bennett McCallum, 1983; Thomas Cooley and Stephen LeRoy, 1985; Edward Leamer, 1985; Ben Bernanke, 1986; and Christopher Sims, 1986.)

This line of thinking played an important role in the evolution of VAR analysis in the 1980s. In this paper, however, I have nothing new to say about it. Still, in deference to its importance, let me warn the reader that I will interpret VAR results somewhat naively. That is, I will describe various numerical results as supporting or contradicting a certain strong version of monetarism depending on whether a fairly simple, straightforward interpretation of the VAR results suggests that money's role in price and output determination jives with the predictions of the theory. It might be more scrupulous to describe such results as *superficially* supporting or contradicting the theory. I reject this option as too clumsy and tedious, but invite the reader to supply such qualifiers as needed. I also encourage exploration of the extent to which more sophisticated interpretations of my results would change them.

(1989) particularly questioned the robustness of Sims' evidence against monetarism. Switch a few features of the model from the forms Sims chose to other equally plausible forms, they said, and some of his results weaken considerably or disappear. Modifications shown to help produce at least mild evidence of nonrobustness include adding a time trend, switching from monthly to quarterly data, adding more past values of variables, and switching to alternative measurements of the variables (for example, replacing the producer price index with the consumer price index). Furthermore, the apparent nonrobustness of Sims' nonmonetarist findings has led some critics to speculate that the nonrobustness of VAR evidence on business cycle issues could be a widespread or even universal phenomenon. (See, for example, Spencer 1989, pp. 452–53; Runkle 1987, p. 442.)

Sims has partially accepted the criticisms of the skeptics in the sense that he has not tried to defend his 1980 results to the last decimal place (Sims 1987, p. 443). But he has stood by key parts of his results as well as the usefulness of his VAR method (Sims 1987, 1989). Sims still claims that his low estimate of money's role in output determination is probably fairly accurate and that his 1980 evidence against the particular form of monetarism he had targeted is robust (Sims 1987, pp. 443, 448; 1989, p. 491). Implicitly, then, Sims claims that his 1980 findings provide an example of how even simple VARs can be used to present important evidence on a theoretical issue.

In some respects, this continuing debate between Sims and his critics is anachronistic. The debate has more or less outlived the object of Sims' investigation: the specific form of monetarism he presented evidence against is no longer widely believed. The debate has also outlived Sims' methodology: current VAR analyses are often built upon more elaborate statistical and theoretical assumptions, such as Bayesian priors or nonrecursive identification schemes. So what this appears to be is a vigorous, ongoing debate about evidence on a discarded theory produced by a superseded methodology.

In other respects, however, this debate remains timely and important. The findings of Sims and his critics have been used to support broad claims that continue to color general opinions of not only VAR analyses but also time series analyses in general. Therefore, I here reexamine the conclusions reached by both Sims and his critics. My approach is to estimate hundreds of variations of Sims' model and then try to pick out any statistically reasonable ones that appear to

contradict some of Sims' key results.

I find that both sides of the debate have some merit. About Sims' 1980 results, I find, as did Sims' critics, that several of his estimates of one variable's role in the determination of another, including his low estimate of money's role in output determination, are not robust. That is, I find many alternative specifications that appear to be as plausible as Sims' model yet have significantly different implications about at least one variable's role in determining another. However, I find few models that even come close to supporting the version of monetarism Sims evaluated. In that sense, I find, as Sims did, that his evidence against this specific form of monetarism is robust.

At the more general level, I again find some truth on each side. I agree with the critics that many results from VAR models, at least those constructed with generic macroeconomic variables such as output or prices, may in fact not be robust. However, I also agree with Sims that nonrobustness is not a general property of VAR results and that even simple VARs can sometimes provide useful evidence on economic issues. In short, it is not generally true that all VAR results are robust or that none are. What does seem true—and what this study demonstrates—is that researchers using VARs should check their results for robustness.

A VAR View of Monetarism . . .

Sims (1980b) analyzed the dynamic behavior of the U.S. economy in the postwar period (1947–78).² His focus, at least initially, was on how well the performance of the economy conforms to the predictions of a simple and rather strong version of monetarist theory.

The core of monetarism is the belief that monetary policy is an important cause of fluctuations in the growth of output and the price level. As stated, this belief is too vague to be used by policymakers or tested by economists. Thus, operational versions of monetarism surround the core belief with more specific statements about just what *monetary policy* means and just how important it is.

In his 1980 article, Sims examined a version of monetarism that he later called *monism* (1987, p. 448). This form of monetarism has four key elements. Sims (1980b, p. 250) explicitly listed the two most distinctive elements:

²Sims also analyzed the dynamic behavior of the U.S. economy in the interwar period (1919–41) and contrasted it to the behavior of the postwar economy. This part of his work has received little attention, and I will not reexamine it here.

1. Monetary policy, or its instability, is the primary cause of business cycles.
2. The time path of the quantity of money in circulation is a good indicator of monetary policy.

Other versions of monetarism differ with at least one of these points. Some downplay monetary policy's primacy as the cause of business cycles (without dismissing its importance altogether). And some hold that the quantity of money is, by itself, not necessarily a good indicator of monetary policy.

The other two key elements of monism, though not explicitly listed by Sims, have been clearly expressed by monetarist authors of the 1960s and 1970s, such as Milton Friedman and Anna Schwartz (1963) and William Poole (1978). These elements can be inferred as either implicit in Sims' text or common to most forms of monetarism:

3. Changes in the quantity of money are the primary cause of business cycles because these changes cause, lead, and are positively related to changes in output (at least in the short run).
4. Changes in the quantity of money lead, are positively related to, and are the primary determinants of changes in the price level (at least in the long run).

To test this monist theory, Sims (1980b) used a simple, four-variable VAR and some techniques he had developed for analyzing economic issues with VARs. The four variables, again, are output, prices, interest rates, and money. Sims used data on these variables from the postwar period to try to estimate whether the quantity of money plays a leading role in explaining the fluctuations of output and prices.

To do this, Sims had to specify exactly which data he would use. He chose monthly rather than quarterly or annual data, perhaps in hopes of getting more precise measures of the dynamic interactions among the variables. This meant he could not use the U.S. Commerce Department's data on real gross national product (GNP), since they are not available monthly. Instead, he measured output by industrial production (IP), a commonly used monthly output indicator compiled and published by the Federal Reserve Board of Governors. Sims measured the overall price level by the U.S. Labor Department's producer price index (PPI). To provide an alternative to the money stock as a source of business cycle fluctuations, he drew from empirical studies which suggested that interest rates might be important.

In particular, he used the interest rate on four-to-six-month commercial paper (C-paper).³ To represent the stock of money, he chose the Fed's M1, a standard measure of money which then consisted mainly of currency and non-interest-bearing checking accounts.

For estimating money's impact, Sims specified a vector autoregression of the logarithms of his four variables. The model had one equation for each variable, and each equation had the same form: 49 unknown coefficients that multiplied a constant term and one year of past, or *lagged*, values (12 lags in all, for his monthly data) of each of the four variables. Sims' model, thus, was

$$\begin{aligned}
 (1) \quad r_t &= k_r + \sum_{i=1}^{12} a_{ri} r_{t-i} + \sum_{i=1}^{12} b_{ri} m_{t-i} \\
 &\quad + \sum_{i=1}^{12} c_{ri} p_{t-i} + \sum_{i=1}^{12} d_{ri} y_{t-i} + e_{rt} \\
 (2) \quad m_t &= k_m + \sum_{i=1}^{12} a_{mi} r_{t-i} + \sum_{i=1}^{12} b_{mi} m_{t-i} \\
 &\quad + \sum_{i=1}^{12} c_{mi} p_{t-i} + \sum_{i=1}^{12} d_{mi} y_{t-i} + e_{mt} \\
 (3) \quad p_t &= k_p + \sum_{i=1}^{12} a_{pi} r_{t-i} + \sum_{i=1}^{12} b_{pi} m_{t-i} \\
 &\quad + \sum_{i=1}^{12} c_{pi} p_{t-i} + \sum_{i=1}^{12} d_{pi} y_{t-i} + e_{pt} \\
 (4) \quad y_t &= k_y + \sum_{i=1}^{12} a_{yi} r_{t-i} + \sum_{i=1}^{12} b_{yi} m_{t-i} \\
 &\quad + \sum_{i=1}^{12} c_{yi} p_{t-i} + \sum_{i=1}^{12} d_{yi} y_{t-i} + e_{yt}
 \end{aligned}$$

where t is time; r , m , p , and y are the logarithms of interest rates, money, prices, and output, respectively; the k , a , b , c , and d terms are coefficients that determine how the variables interact; and the e 's are the error terms, which capture the monthly unexplained or surprise movement in each variable.

Sims estimated the unknowns in his model by applying ordinary least squares regression to each equation separately for the sample period from January 1948 to December 1978. Besides estimates of the 49 coefficients in each equation, this gave him estimates of each equation's historical error term. Sims used these to compute Σ , the variance-covariance matrix of the error terms, a measure of the correlation among the surprise movements in each variable.

Sims used the estimated version of his model to measure the dynamic interactions among the variables in two different ways. One summary measure, the

³This somewhat unusual choice for interest rates was in part dictated by Sims' goal of comparing the interwar and postwar periods. More commonly used postwar measures of interest rates were not available for the interwar period.

forecast error variance decomposition (FEVD), analyzes the errors the model would tend to make if it were used to forecast its variables. The FEVD is meant to show how much of the average squared forecast error which the model would tend to make is caused by surprise movements associated with each of the variables in the model. The FEVD of a variable thus can suggest that forces associated with one variable are major influences on the evolution of another variable. If, for example, money's share of the FEVD of output were relatively large, then the fundamental factors determining money would seem to be a relatively important source of fluctuations in output. Sims' other summary measure, the *impulse response*, shows how one variable responds over time to a single surprise increase in itself or in another variable. This, too, can suggest evolutionary influences for each variable. Monetarists, for example, might expect that a surprise jump in the quantity of money (if caused by an unanticipated easing of monetary policy) would, over the next several years, cause output to rise gradually but significantly above where it would have been otherwise and then fall back to its original path.

Before Sims could calculate these summary measures, however, he had to add one set of assumptions to his estimated model. Both of the measures require that the notion of a *surprise movement* associated with each of the model's variables be precisely defined. The error vector

$$(5) \quad e_t = [e_{rt} \ e_{mt} \ e_{pt} \ e_{yt}]^T$$

jointly summarizes the surprise movements in the model at time t . But economic theory and estimated covariance matrices suggest that the individual components of this vector are interrelated. Thus, it would not make sense to, for example, treat e_{rt} and e_{mt} as though they were independent surprise movements in, respectively, interest rates and money. Sims solved this problem by assuming a simple model of how the correlations among the components of the estimated e_t 's could be derived from an underlying set of uncorrelated random shocks to the economy. Making assumptions like this is called *identifying* the model; this process uniquely links observed variables that have ambiguous theoretical interpretations (such as the components of e_t) to unobserved variables with clearer theoretical interpretations.

Sims identified his model by assuming a recursive chain of causality among the surprises in any given month. (See the Appendix.) Although he did not say so

explicitly, he seems to have begun by positing a nonmonist policy regime. In this regime, the quantity of money is not a good indicator of monetary policy; instead, monetary policymakers set interest rates. Sims here relabeled the discrepancy between the expected and the actual interest rate at time t , or e_{rt} , as u_{rt} and interpreted it as a shock to a horizontal intramonth money supply function. This shock is random from the point of view of the VAR model, but is thought of as being generated by the nonrandom decisions of the monetary authority. In effect, Sims assumed that, in this nonmonist regime, monetary policy operates by shifting a horizontal intramonth money supply curve so as to peg interest rates for the month.

Next Sims implicitly posited a downward-sloping intramonth money demand curve. This implies that the surprise movement in money, e_{mt} , has two components: $\alpha_m u_{rt}$, which reflects the intramonth effect of moving along the money demand curve in response to the monthly policy shock to interest rates, and u_{mt} , which represents the underlying random shift in money demand at time t .

Then Sims turned to the surprise movement in the price level. He assumed that e_{pt} has three components: $\alpha_p u_{rt}$, to reflect the intramonth effect on prices of the monetary policy shock to interest rates; $\beta_p u_{mt}$, to reflect the intramonth effect on prices of the random shift in money demand; and u_{pt} , to represent the remaining surprise movement in prices at time t . Sims didn't say what model he had in mind here, but it doesn't matter much for the purpose of evaluating monism.

Similarly, the exact interpretation of all of the four components of the surprise movement in output, e_{yt} , is not too important. The components of interest for Sims' purposes, $\alpha_y u_{rt}$ and $\beta_y u_{mt}$, reflect effects on intramonth output stemming from, respectively, the random policy shock to interest rates and the money demand shock. The component $\gamma_y u_{pt}$ is associated in some unspecified way with factors determining prices. The remaining component, u_{yt} , represents a random shock to some of the remaining intramonth factors that affect output.

In matrix notation, the intramonthly model can be concisely written as

$$(6) \quad e_t = Au_t$$

where

$$(7) \quad e'_t = [e_{rt} \ e_{mt} \ e_{pt} \ e_{yt}]$$

$$(8) \quad u'_t = [u_{rt} \ u_{mt} \ u_{pt} \ u_{yt}]$$

$$(9) \quad A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_m & 1 & 0 & 0 \\ \alpha_p & \beta_p & 1 & 0 \\ \alpha_y & \beta_y & \gamma_y & 1 \end{bmatrix}$$

The triangular pattern of the matrix A , with ones on the diagonal and zeros above, reflects the causal pattern Sims (1980b) assumed. Effects flow only downward, from variables earlier in the causal chain to those later.

To make this assumed intramonthly model operational, Sims used the elements of Σ , the variance-covariance matrix of the e_t 's, to compute the α , β , and γ coefficients as population regression coefficients. This means, for example, that α_m is picked to equal the ratio between the historical covariance of e_{rt} and e_{mt} and the historical variance of e_{rt} . It also means that, on average, all of the component common to e_{rt} and e_{mt} is attributed to u_{rt} alone.

. . . With Controversial Results

Sims' VAR Evidence

Sims used his model to compute statistics that summarized the dynamic interactions among the variables—with unexpected results. Sims' statistics appeared to contradict monism, the simple form of monetarism he was examining. They implied that shocks associated with the money stock played a very small role in the determination of postwar output. The results had implications beyond monism, though. The model said money's role in output determination was so small that it seemed to challenge a wide variety of models that embodied the general monetarist belief that fluctuations in the money stock influenced output to at least some degree.

Shocks associated with the money stock were clearly not the dominant source of output fluctuations when Sims' model was identified with his nonmonist policy regime. There, remember, the causal chain ran from interest rates to money to prices to output. In analyzing the errors that that model would tend to make when forecasting output four years ahead, Sims estimated that the cumulative effects of money demand shocks to the money stock (the u_{mt} 's, not the e_{mt} 's) would account for only 4 percent of the average squared error. Monetary policy shocks to interest rates seemed to be much more important, accounting for 30 percent of that error.

Sims checked whether the unimportance of shocks associated with the quantity of money stemmed primar-

ily from his assumption of a nonmonist policy regime, where money shocks come from the demand side rather than the supply or policy side. To do this, he repeated his analysis with a more monist regime, embodied in a different causal chain: from money to interest rates to prices to output. (Again, see the Appendix.) He re-labeled e_{mt} as u_{mt} and identified it as a monetary policy shock to a vertical intramonth money supply curve. In this chain, e_{rt} was interpreted as having a component that represented movements along the money demand curve in response to money supply shocks and a component, u_{rt} , that represented random shifts in money demand. The rest of the chain was as before, with u_{mt} and u_{rt} playing the roles previously assigned to u_{rt} and u_{mt} , respectively. This second identification, where all surprise movements in money are assumed to reflect policy decisions, produced nearly the same results as the first. Sims took this as evidence that his original findings were not caused by a simple confounding of policy and nonpolicy shocks to the money stock.

Sims had found that shocks associated with the quantity of money were far from the primary determinant of output, regardless of whether the monetary policy variable was interest rates or the money stock. This he considered strong evidence against monism, the simple form of monetarism he had identified as his target. That version of monetarism asserted, again, that changes in the quantity of money were the best measure of monetary policy and monetary policy was the main cause of output fluctuations.

Sims was not as confident about his evidence that money had almost no effect on output, but he did explore its implications. He attempted to reinterpret the fact, documented by monetarist scholars, that fluctuations in the money stock were positively correlated with subsequent fluctuations in output. One reinterpretation was statistical. Sims noted that shocks associated with interest rates accounted for over 50 percent of the four-year FEVD of the money stock. Furthermore, his impulse responses showed that both output and money responded to shocks to interest rates in roughly the same way: with a smooth, slowly building but sustained decline. Sims proposed the common response of output and the money stock to interest rate shocks as a statistical alternative to monism for explaining the empirical correlation between fluctuations in money and output. Sims also offered a theoretical alternative. He sketched a Keynesian investment model in which anticipated declines in the real return to capital raise current interest rates and, with a delay, depress production. The anticipated decline in output causes the

money stock to fall, gradually and somewhat ahead of the actual decline in output, by smoothly reducing the demand for money.

The Critical Response

Sims' evidence against monetarism naturally received a lot of attention. This was especially true of his most surprising finding, that the money stock plays essentially no role in output determination. A prominent line of criticism during the 1980s was that Sims' particular empirical results, and possibly VAR results in general, are not robust.

Robustness, again, means that the results do not change significantly when the model is modified in an arbitrary way. *Arbitrary*, in turn, means that there is no strong economic or statistical argument that either model is clearly superior to or conceptually different from the other. To criticize Sims' results as not robust thus involves proposing a seemingly innocent modification of his model and showing that the modified model gives results that are sufficiently different from those of his model to call into question the validity of his results.

Several researchers have presented evidence against the robustness of Sims' results. As Table 1 shows, they find that several simple modifications of Sims' specification lead to rather different numbers. King (1983), for example, added three lags and a time trend to Sims' model, changed the interest rate data to the rate on U.S. Treasury bills (T-bills), and then estimated the model on a slightly later time period. He found that these changes resulted in money's share in the four-year-ahead FEVD of output rising from Sims' 4 percent to 24 percent. Similar results were obtained for models involving other modifications—quarterly data, two-year lags, and different measures of the price level and output. (See Table 1.)

Criticism of Sims' findings has tended to focus on money's and interest rates' shares in the FEVD of output at the expense of other indicators of the monetarist or nonmonetarist properties of estimated models. In part, this is probably because the small share Sims found for money and the much larger share he found for interest rates were his most surprising results. For some, though, another factor was involved. King (1983, p. 7) noted, for example, that Sims' interpretation of his results "influenced a number of authors to attempt to reformulate business cycle theories away from the exclusive concentration on the money supply." By challenging the robustness of Sims' numbers, especially the robustness of his small money share, critics like Spencer (1989, p. 453) meant to question the necessity

of reformulating business cycle theory. They were not necessarily concerned with defending monism, and perhaps for this reason they generally did not check the implications of their respecified models for all four elements of monism.

Some critics also drew broader conclusions from their study of Sims' results. The sensitivity of Sims' numbers to simple changes in specification, as well as the considerable statistical imprecision of estimated VAR impulse responses and FEVDs, led them to question in general the usefulness of evidence from whole classes of VARs. Runkle's (1987, p. 442) results suggested to him that "it may be difficult to draw firm conclusions about economic relations from unrestricted VAR's," though he withheld judgment about VARs restricted by Bayesian priors or by identifying assumptions more sophisticated than Sims' causal chains. Spencer (1989, p. 453) affirmed Runkle's generalization and added remarks that seem to question any inferences drawn from VARs, restricted or not.

Another VAR View . . .

Sims (1987) has objected that some of the changes critics have proposed for his model are not arbitrary. If correct, that would mean the large role in output determination that the critics' models found for money cannot be considered a legitimate challenge to the robustness of Sims' money number. Sims has also pointed out that both his model and many of his critics' found that interest rates' role in output determination is larger than money's role. This suggests that at least this part of Sims' nonmonist conclusion is robust. Neither Sims nor his critics, however, have thoroughly explored beyond this point in the debate.

I attempt to do that. I check whether Sims' results hold up for hundreds of variations of his specification—most of the possible combinations of the modifications shown in Table 1 plus others. I also use statistical tests to help evaluate whether the proposed modifications can be viewed as arbitrary and whether the results they produce are really strong evidence of nonrobustness.

The least controversial of the modifications, in the sense that neither Sims nor his critics have raised theoretical or statistical objections to them, involves replacing one data series by another that in principle measures the same concept over the same interval of time. I try this change on all of the variables in Table 1 except Martin Eichenbaum and Kenneth Singleton's (1986) real interest rates, which I regard as conceptually distinct from Sims' choice of nominal rates. As measures of the price level, I replace Sims' PPI with the

Table 1

How Simple Changes in Sims' Model Have Changed His Results

☐ Specification Different From Sims'

| Studies | Model Characteristics | | | | | | | Model Results | |
|---------------------------------|-----------------------------|---------------------|---------------------|-------------------------|-------------------|----------------------|----------------|------------------------------------|----------------|
| | Variables ¹ | | | Number of Years of Lags | Constant or Trend | Monthly or Quarterly | Sample Periods | (% Share of 4-Year FEVD of Output) | |
| | Interest Rates ² | Prices ³ | Output ⁴ | | | | | Money | Interest Rates |
| Sims (1980b) | C-Paper | PPI | IP | 1 | C | M | 48:1-78:12 | 4 | 30 |
| King (1983) | T-Bills | PPI | IP | 1.25 | C,T | M | 50:1-81:6 | 24 | 29 |
| | T-Bills | Deflator | GNP | 2 | C | Q | 52:1-81:II | 18 | 66 |
| Eichenbaum and Singleton (1986) | T-Bills | CPI-S | IP | 1 | C,T ⁵ | M | 59:2-79:12 | 19 | 27 |
| Runkle (1987) | C-Paper | PPI | IP | 1 | C,T | Q | 48:1-78:IV | 22 | 34 |
| | T-Bills | PPI | IP | 1 | C,T | Q | 48:1-78:IV | 28 | 27 |
| Spencer (1989) | T-Bills | CPI-S | IP | 1 | C,T | M | 48:1-78:12 | 19.5 | n.a. |
| | T-Bills | CPI-S | IP | 1 | C | Q | 48:1-78:IV | 19.5 | n.a. |
| | T-Bills | CPI-S | IP | 1 | C,T | Q | 48:1-78:IV | 27.4 | n.a. |
| | T-Bills | CPI-S | IP | 2 | C | M | 48:1-78:12 | 19.9 | n.a. |
| | T-Bills | CPI-S | IP | 2 | C,T | M | 48:1-78:12 | 23.3 | n.a. |
| | T-Bills | CPI-S | IP | 2 | C | Q | 48:1-78:IV | 28.9 | n.a. |
| | T-Bills | CPI-S | IP | 2 | C,T | Q | 48:1-78:IV | 27.1 | n.a. |

FEVD = forecast error variance decomposition

n.a. = not available

¹All of these studies used the Federal Reserve Board's M1 as their money variable. Most variables are logged, and all except interest rates are seasonally adjusted.

²Most interest rate variables are net and nominal. Sims' is the rate on six-month commercial paper; Eichenbaum and Singleton's, the ex post inflation-adjusted rate on one-month Treasury bills; and the others', the rate on three-month Treasury bills. King's are gross rates.

³The PPI is the producer price index for all items; the CPI-S, the consumer price index for all items except shelter. Both are U.S. Labor Department series. The deflator is the measure that the U.S. Commerce Department uses to adjust the gross national product for inflation.

⁴For output, most studies used only the Federal Reserve Board's index of industrial production. King also used the U.S. Commerce Department's measure of real gross national product.

⁵Before estimating their model, Eichenbaum and Singleton subtracted a constant and a linear trend from all their variables except interest rates.

Commerce Department's implicit GNP price deflator and the Labor Department's consumer price index for all items less shelter (CPI-S). As measures of output, I use real GNP as well as Sims' IP. For interest rates, I use those on three-month T-bills as well as those on six-month C-paper. Finally, I examine two alternative measures of the money stock—the monetary base (MB) and M2—that Sims and his critics have examined only briefly or not at all. Since the monetarist theory Sims examined was specified in terms of generic concepts like money and interest rates rather than specific data series, such swaps among alternative measures of these concepts seem to be valid tests of the robustness of Sims' results.⁴

I also consider the other modifications shown in Table 1. For lag length, I try both one and two years plus one period, that is, 13 and 25 lags for monthly models and 5 and 9 lags for quarterly models.⁵ Since economic theory doesn't pin down the appropriate lag length for Sims' VAR, varying lag lengths is an appropriate test of the robustness of his results provided the alternative lag lengths are not implausible on statistical grounds.

The remaining modifications in Table 1 are more controversial. Sims has objected to the addition of linear trends, even when they are statistically significant. His argument (Sims 1987, p. 444) is that because the responses of both money and output to surprise movements in interest rates are low-frequency phenomena (that is, their responses are long, smooth, and slowly building), adding a low-frequency variable with no clear economic interpretation, such as a linear trend, just adds uncertainty to the estimation of long-run effects. Sims' critics obviously feel differently about the validity of checking robustness by adding trends since almost all have done it. In my study, I try to highlight when trend terms have an important effect on the results.⁶

Sims has not objected in print to checking his results with quarterly rather than monthly data. In fact, he has done this sort of temporal aggregation himself (Sims 1987). Nonetheless, the statistical theory of time series analysis shows that, in general, temporal aggregation can cause the estimates of the relationships among variables to become quite misleading. Lawrence Christiano and Eichenbaum (1987) have argued that this may be a particularly serious problem in models with money and output. Therefore, I try to test whether the changes produced by switching to quarterly versions of Sims' model reflect nonrobustness or merely errors induced by temporal aggregation.

As did some of Sims' critics, I use a somewhat

different time period for estimating my models. Sims used data from 1947 to 1978. After allowing for one year of lagged values, he was able to fit his model to data from the beginning of 1948 through the end of 1978. Some of my data were not available before 1948. Furthermore, I work with up to nine quarters of lagged data, so I can only begin estimating the model in the second quarter of 1950. To keep the results of my many models comparable, I use one sample period in which the estimation of all of my models begins at that time (or in April 1950 for monthly models) and ends at the end of 1978.

A problem with my data set, as well as with the data sets used by Sims and most of his critics, is that it may mix data from two or more significantly different periods of economic history. In particular, both historical events and statistical evidence suggest that for Sims' model a structural break (a change in the relationships among the model's variables) may have occurred around 1951. Early in that year, the Federal Reserve System signed a famous accord with the U.S. Treasury officially stating that the Fed would no longer adhere to the policy, adopted during World War II, of fixing the price of long-term government bonds. Both King (1983) and Sims (1987) have suggested that this change in Fed policy might have significantly changed the relationships among money, interest rates, and the rest of the economy. In addition, in testing various dates for a possible shift in the growth trend of GNP, Christiano (1988) has found the strongest evidence of a shift at about this time.

⁴As did Sims and his critics, I use seasonally adjusted versions of all data series except those on interest rates. Several of my alternative series are necessarily constructions. The official GNP and deflator series, for example, are available only quarterly. In monthly models, therefore, I use unofficial monthly series constructed as described by Hossain Amirizadeh (1985). Similarly, some official money series are not available for some time periods, so had to be constructed. My M2 data for the period before 1959 were provided by William Roberds. For details on their construction, see the paper by Charles Whiteman and Roberds (1989).

⁵I add the extra period (for example, the 13th month), even though Sims and most of his critics have used even year lags, for three reasons. Extra lags can sometimes capture seasonal effects not removed by seasonal adjustment of the data. These extra coefficients are often highly statistically significant. And Sims (1987) has recently adopted this specification in reexamining his earlier work.

⁶Based on the work of Martin Eichenbaum and Kenneth Singleton (1986), David Spencer (1989), and others, as well as on some experiments of my own, I ignore models with differenced data since they almost never seem to yield monetarist results. For the same reason, I will not report here results from models estimated with quadratic trends, even though I estimated hundreds of them. Quadratic trend models can, indeed, raise the importance of money. But this is almost always because they make the relationship between money and the rest of the economy nonmonotonic by, for example, making money's impact on output a strongly negative one.

My tests for a structural change in Sims' model in the 1950–78 period agree with these studies: my results are also strongly significant by conventional standards. By contrast, my tests of the period 1953–79 show little evidence of a break.⁷ Therefore, I will report all of my results not only for models estimated over the 1950–78 period, but also for models estimated over the period starting in the fourth month (second quarter) of 1953 and ending in the ninth month (third quarter) of 1979.

Finally, I will report results for one causal chain, with intramonth influences flowing from policy shifts in a vertical money supply curve to interest rates (or money demand) and then on to prices and output. This reverses money's and interest rates' positions from those in Sims' causal chain, but I adopt it for two reasons. First, as the work of Sims (1980b) and Spencer (1989) would seem to predict, my results are rarely sensitive to which of the two orderings is used. Second, my ordering seems to conform more closely to monism, the theory Sims was evaluating, because it identifies monetary policy with intramonth shocks to the quantity of money.

By concatenating all these variations of Sims' specifications, I get hundreds of modified versions of Sims' four-variable model. For quarterly models, on each of my two sample periods, I can specify 48 modifications to use with each of the available monetary aggregates (MB, M1, and M2). (I have two interest rate series; three price series; two output series; two types of lags, short or long; and two trend conditions, with or without.) For monthly models, I can do the same except that monthly M2 data are not available. Thus, I get 144 quarterly models and 96 monthly models—240 models in all.

For each of these models, I look not only at FEVDs, but also at paths of the dynamic responses of output and prices to random shocks in money supply and demand. Although most previous studies have downplayed or ignored these impulse responses, I find them quite helpful in picking out nonmonist models.

... With More Conclusive Results?

The Critics Are Right . . .

My results partly support Sims' critics. Many of the modifications of Sims' (1980b) model that can be viewed as arbitrary do produce statistically significant changes in Sims' numbers, including his estimate that money's role in output determination is near zero. This estimate, that is, does not seem to be robust.

□ *Robustness Tests*

My version of Sims' model is comparable to his. When I estimate it on the 1950–78 period, I get results that are

Table 2

My Version of Sims' Model

% Shares That Each Shock Contributes to Each Variable's 4-Year FEVD*

| Variables and Models** | Sources of Shocks | | | |
|------------------------|-------------------|----|-----|----|
| | C-Paper | M1 | PPI | IP |
| C-Paper | | | | |
| Sims | 50 | 19 | 4 | 28 |
| Todd | 58 | 12 | 5 | 24 |
| M1 | | | | |
| Sims | 56 | 42 | 1 | 1 |
| Todd | 54 | 43 | 2 | 2 |
| PPI | | | | |
| Sims | 2 | 32 | 60 | 6 |
| Todd | 1 | 18 | 56 | 25 |
| IP | | | | |
| Sims | 30 | 4 | 14 | 52 |
| Todd | 40 | 2 | 14 | 44 |

*FEVD = forecast error variance decomposition

**Estimation periods: Sims, January 1948–December 1978;
Todd, April 1950–December 1978

Source: Sims 1980b

mostly quite close to those Sims (1980b) reported, as shown in Table 2. The biggest differences are in the determination of prices and output. My results give money shocks less of a role and output shocks more of a role in price determination. And my results give interest

⁷I used Sims' variables and either 5 quarterly or 13 monthly lags to conduct, respectively, quarterly and monthly versions of these tests. In all these tests, I computed standard *F*-statistics for the null hypothesis that all the coefficients of the model were constant throughout the sample period versus the alternative hypothesis that all the coefficients changed at a certain date during that period. I varied the possible dates of the break from the earliest end-of-year (fourth quarter or December) period for which I could estimate the model before the break to the latest end-of-year period for which I could estimate the model after the break. The *F*-statistics for each of these dates were compared to two different distributions. One was the standard, asymptotically valid, theoretical *F*-distribution. The other was a small-sample *F*-distribution estimated by Lawrence Christiano's (1988) Monte Carlo method. To implement this method, Sims' specification was fitted to the data and used to generate 1,000 alternative sample paths of the four variables. *F*-statistics for each possible break were then computed for the 1,000 alternative paths and sorted to produce small-sample distributions for each possible break date.

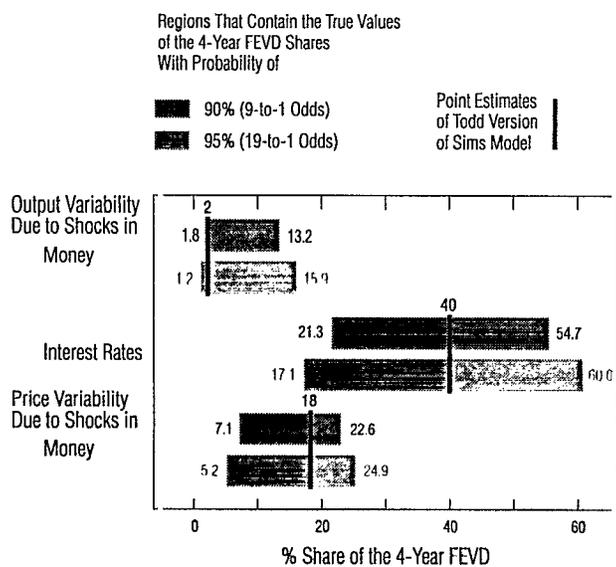
rate shocks more of a role and output shocks less of a role in output determination. The later beginning of my sample period accounts for most of these differences. The rest are primarily due to my use of 13 instead of 12 lags. The differences in causal chain ordering and revisions to data are unimportant. Since my version of Sims' model is so like his, I will evaluate the robustness of my findings, not his.

Estimates such as those in Table 2 are statistically much less precise than they appear. To quantify this imprecision, I have estimated 90 and 95 percent probability regions surrounding the key results of Sims' model, as shown in the accompanying figure.⁸ The odds that the true values of the FEVD shares are actually in each region are 9-to-1 for the 90 percent region and 19-to-1 for the 95 percent region. As Runkle (1987) has emphasized, these regions are fairly broad. For example, my estimate of money's 2 percent share of the output FEVD provides only moderately strong statistical evidence against shares as high as 15 percent.

I use the 95 percent probability regions to establish criteria for the robustness of the numbers I estimate for Sims' model. In particular, I will regard a statistic estimated by Sims' model to be nonrobust if

1. I can find a modification of Sims' model, from the

A Basis for Measuring Robustness



FEVD = forecast error variance decomposition

class of modifications outlined above, for which the same statistic is estimated to lie outside the 95 percent regions shown in the figure.

2. I cannot find a strong theoretical or statistical argument for disregarding the results of the modified model.

□ Test Results

Some of Sims' results clearly satisfy my first criterion. Table 3 shows that many modifications of Sims' model produce significantly different FEVD shares when estimated on the 1950-78 period. For money's share in output determination, quarterly models are especially likely to give significantly higher results. In addition, linear trend terms often significantly boost money's share when used with M1 or MB, but not when used with M2. For interest rates' share in output determination, models without trends are more likely to give significantly lower results than models with trends, and M2 also seems to cut interest rates' share significantly. For money's share of price determination, the main boosting factor is choice of the data series on prices, with the CPI-S and the deflator almost always producing significantly higher results than Sims' choice of the PPI. Table 4 shows that my first criterion for statistical nonrobustness is also frequently met when the models are estimated on the 1953-79 period.

No elaborate tests or arguments are needed to see that my second criterion for statistical nonrobustness is also met for some results for each period. Just look at the parts of Tables 3 and 4 that display the results for models with M1, the money variable preferred by Sims as well as his critics. No matter whether the specification includes quarterly or monthly data, one- or two-year lags, trend terms or no trend terms, some of the M1 models have at least some statistics outside of the 95 percent probability regions surrounding Sims' results. Neither Sims nor anyone else has offered any reason why swapping among these nonmoney variables is not a valid way of checking robustness. Thus, I conclude that at least some of Sims' (1980b) results are statistically not robust.

For Sims' most famous result—the very small share of money in the four-year-ahead FEVD of output—my

⁸The Bayesian, Monte Carlo method of Thomas Doan (1988, p. 10-4) was used to compute 1,000 points in the posterior distribution of each result. These were sorted into a histogram. The probability regions were constructed basically by throwing out the lowest cells of the histogram, from either tail, until the remaining cells contained 90 or 95 percent of the mass of the distribution. The resulting highest posterior density regions are thus not necessarily bounded away from zero for FEVDs.

Tables 3 and 4

Some Evidence of Nonrobust Results . . .

Number of Versions of Sims' Model That Produce FEVD Shares Outside the 95% Probability Regions†

Table 3 Models Estimated on the 1950–78 Period*

Table 4 Models Estimated on the 1953–79 Period**

| Model Characteristics | | | | Shares of Variability in | | | Shares of Variability in | | |
|-----------------------|---------|----------|----------|----------------------------|-------------------|-------------------------------|----------------------------|-------------------|-------------------------------|
| | | | | Output Due to Shocks in | | Prices Due to Shocks in | Output Due to Shocks in | | Prices Due to Shocks in |
| | | | | Money | Interest Rates | Money | Money | Interest Rates | Money |
| Quarterly | M1 | 5 Lags | No Trend | 4 | 6 | 8 | 2 | 4 | 6 |
| | | | Trend | 9 | 0 | 9 | 4 | 0 | 5 |
| | | 9 Lags | No Trend | 2 | 7 | 8 | 1 | 7 | 4 |
| | | | Trend | 11 | 2 | 8 | 1 | 0 | 4 |
| | M2 | 5 Lags | No Trend | 12 | 12 | 0 | 12 | 10 | 0 |
| | | | Trend | 12 | 4 | 0 | 4 | 2 | 0 |
| | 9 Lags | No Trend | 8 | 9 | 0 | 2 | 9 | 0 | |
| | | Trend | 2 | 3 | 0 | 1 | 3 | 0 | |
| MB | 5 Lags | No Trend | 2 | 6 | 10 | 0 | 2 | 4 | |
| | | Trend | 10 | 0 | 8 | 0 | 0 | 2 | |
| | 9 Lags | No Trend | 7 | 5 | 8 | 0 | 2 | 5 | |
| | | Trend | 12 | 1 | 8 | 0 | 0 | 6 | |
| Monthly | M1 | 13 Lags | No Trend | 0 | 4 | 10 | 7 | 3 | 5 |
| | | | Trend | 0 | 0 | 11 | 7 | 0 | 2 |
| | | 25 Lags | No Trend | 1 | 3 | 8 | 4 | 5 | 2 |
| | | | Trend | 8 | 0 | 8 | 3 | 0 | 2 |
| | MB | 13 Lags | No Trend | 0 | 4 | 9 | 3 | 2 | 4 |
| | | | Trend | 2 | 0 | 9 | 0 | 0 | 4 |
| | 25 Lags | No Trend | 3 | 5 | 7 | 0 | 2 | 1 | |
| | | Trend | 12 | 0 | 7 | 0 | 0 | 0 | |

FEVD = forecast error variance decomposition

†To get each number in these tables, 12 versions of Sims' model—differing only in their data series—were checked. Each number counts the versions, out of the 12, that say money's share of output variability is at least 15.9%, interest rates' share of output variability is at most 17.1%, or money's share of price variability is at least 24.9%.

*This period is from April 1950 through December 1978.

**This period is from April 1953 through September 1979.

second criterion for statistical nonrobustness is met clearly for the 1953–79 period, but not so clearly for the 1950–78 period. Table 4 shows that even many of the models most like Sims’—monthly M1 models with about one year of lagged values—give money more than a 15.9 percent share in the four-year-ahead FEVD of output when they are estimated over the 1953–79 period. For the 1950–78 period, however, obtaining shares greater than 15.9 percent requires either further lags, temporal aggregation, or replacement of M1 by MB plus addition of a time trend. Sims (1987) regards addition of time trends as inappropriate, and there are potential statistical objections to the rest of these modifications as well.

I find little statistical evidence against the addition of time trends or extra lags, however. Statistical tests of several of the monthly M1 and MB models that give money a relatively large role in output determination reject the hypotheses that the true model has 13 lags, no time trend, or both of these features when the alternative is a model with 25 lags and a linear time trend.⁹ In addition, simulations suggest that the differences between trend and no trend versions of these models can be explained as the effect of erroneously fitting a no trend model to data generated by a trend model at least as well as these differences can be explained as the effect of erroneously fitting a trend model to data generated by a no trend model. For lag lengths, the selection criterion of Hirotugu Akaike (1974) supports longer lag lengths more strongly than shorter lag lengths. The only statistical evidence against the long lag specification comes from Gideon Schwarz’s (1978) alternative lag length selection procedure, which for monthly (but not quarterly) models favors lag lengths shorter than one year. Though Schwarz’s criterion has some theoretical advantages over Akaike’s, as well as over the other statistical tests I used, overall the statistical evidence does not indicate that the monthly models with time trend or 25 lags are unreasonable.

These tests suggest that Sims’ finding of a nearly zero share for money in the four-year-ahead FEVD of output is not robust over the 1950–78 period either. This seems clear if trend terms are accepted as arbitrary variations of Sims’ model. Even without trend terms, however, some of the monthly M1 and MB models with 25 lags boost money’s share above the 95 percent probability region of Sims’ model.

... *But So Is Sims*

Still, the results of my tests also partly support Sims. Few of the statistically reasonable alternatives even

come close to overturning his general conclusion that the data contradict monetarist—or monist—theory. Unlike some of his numbers, that is, Sims’ nonmonist conclusion is robust.

It may seem obvious that if Sims’ key numbers are individually unreliable, then any conclusions drawn from his analysis must be too. This is not true. Sims’ results were more than just slightly nonmonist, so fairly large changes in his numbers are required to overturn his nonmonist conclusion. Also, the monist theory Sims presented some evidence against makes very strong claims about several aspects of the dynamic relationships among Sims’ variables that Sims did not report on. Nonetheless, a first round of tests shows that some versions of Sims’ model do appear to be at least somewhat monist when estimated on the 1950–78 period. A second round shows, however, that there are statistical reasons for disregarding many of these and that the remaining cases have, at best, weakly monist properties. Finally, a double-check shows that Sims’ conclusion gets strong support from the more statistically homogeneous period 1953–79.

□ *Monism Testing: Round 1 . . .*

Recall the monist view that monetary policy can be gauged by movements in the money stock and that these movements lead, are positively related to, and are the principal determinants of movements in both output and prices. In a version of Sims’ VAR with shocks to the money equation identified as unanticipated shifts in a vertical intramonth money supply curve, this implies that the impulse responses of both output and prices to money shocks should be predominantly positive, especially at short-to-medium horizons for output and at medium-to-long horizons for prices. Furthermore, money’s shares in the FEVD of output and prices should be large, at least by medium horizons for output and by medium-to-long horizons for prices.

But just how large should they be? At least some prominent monetarists of the 1960s and 1970s seem to think that money’s share of the FEVD of output should be well over 50 percent at business cycle horizons (Poole 1978, especially p. 64, but also pp. 1, 2, 97, and 104; Friedman 1969, p. 146). Friedman and Schwartz (1963, p. 695) suggest that in an era of relatively mild cycles, such as the one studied here, money’s share might be lower, only about 50 percent. They also sug-

⁹This is based on both standard and small-sample-adjusted likelihood ratio statistics computed for the monthly models in Table 5. For the small-sample adjustment, see the work of P. Whittle (1953) and Sims (1980a).

gest that size for money's share of the FEVD of prices.

To be on the safe side, I will set much lower standards to try to pick out some monist models. Under what I'll call my *four-year test*, I will regard a model as nonmonist unless it meets all of these criteria:

1. Money's shares of the four-year-ahead FEVDs of output and prices equal or exceed 15 percent.
2. Money's share of the four-year-ahead FEVD of output exceeds the share attributed to interest rates.
3. Money's relatively large shares in the FEVDs of output and prices are not caused by negative relationships between money and output or prices.

If a model has all of these characteristics, that is, I will consider it a candidate for a monist model.

The third, nonnegativity criterion is based on an examination of the money-output and money-price impulse responses. I judge a model to be nonmonist by this criterion if all of these situations are true:

- a. At a certain horizon H (for example, two years ahead), the effect on output or prices of a surprise movement in money becomes negative.
- b. After H , the cumulative effect on output or prices of a surprise movement in money, measured by the sum of the money-output or money-price impulse response coefficients beyond H , is negative.
- c. Money's share in the FEVD of output or prices does not exceed 15 percent until H or later.

In other words, I regard a model as nonmonist if money's importance in determining output or prices seems to result from a negative relationship between money and output or prices.¹⁰

I will also work with another test of monism. The four-year test uses the horizons that Sims and his critics have examined most often. But monetarists might think that money's influence over output should be examined at a shorter horizon. For what I'll call my *peak-response test*, I will require instead that the peak in money's share of the FEVD of output exceed both 15 percent and the peak in interest rates' share of the FEVD of output.¹¹ The third criterion of the four-year test will still apply in the sense that the peak output share for money will be defined as the largest share (among horizons less than or equal to four years) not attributable to a negative response of output to money (in the sense defined above). However, I will continue to check the money-price relationship just at 48 months or 16 quarters, because it is almost never true here that money's share

of the FEVD of prices rises above 15 percent for an early horizon and later slips below 15 percent.

When I estimate my 240 models on the 1950–78 period, I find that a sizable minority, displayed in Table 5, pass one or both of my tests for possible monism. Among the 48 quarterly M1 models, for example, 11 pass the four-year test and 6 others pass the peak-response test. Although Sims' conclusion holds up for most of the models, at a glance the exceptions seem too numerous to regard his conclusion as robust.

A glance may not be good enough, though. For Table 5 also shows some noteworthy patterns among the models. The sizable minority of models that pass either of the tests for monism tend to share certain characteristics. For example, among models with M1 or MB as money, quarterly models produce monist results more frequently than do monthly models. Yet, among models with M2, which I only have in quarterly form, none of the 48 estimated models passes either of my tests. This is primarily because the response of prices to M2 is generally small or negative. Among the monthly M1 and MB models, 25 lags and interest rates represented by T-bills seem to be necessary, but not always sufficient, to produce monist results.

□ . . . And Round 2

These patterns suggest that before I judge Sims' conclusion to be nonrobust, I should evaluate whether there is any reason to rule out the modifications that seem critical for that judgment. In particular, temporal aggregation—the switch from a monthly to a quarterly model—seems to be the one modification most likely to overturn Sims' conclusion. When the monthly and the quarterly models disagree, as some do here, which should we believe?

Claims have been made for both types of models. On the one hand, time series analysts have shown, precisely

¹⁰In implementing this third, nonnegativity criterion, I actually only check the FEVD of output at horizons 1, 1.5, 2, 3, and 4 years. I also only check the sum of the impulse response coefficients between those horizons, such as the sum of quarters 1 through 4 or of months 37 through 48. For a quarterly model, my criterion would be the following. Let $I = \{4, 6, 8, 12, 16\}$ and $J = \{1, 2, 3, 4, 5\}$. Define $H: \{0\} \cup J \rightarrow \{0, 4, 6, 8, 12, 16\}$ as $H(j) = 0$ if $j = 0$ and, for $j \in J$, $H(j)$ is the j th element of I . Then a model fails my criterion if three things are true: (a) For some $j \in J$, the sum of the coefficients in the response of output to a shock to money from quarter $H(j-1) + 1$ through quarter $H(j)$ is negative; (b) the subsequent partial sums—from $H(j) + 1$ to $H(j+1)$, $H(j+1) + 1$ to $H(j+2)$, ..., $H(4) + 1$ to $H(5)$ —are all negative; and (c) for $i = 1, 2, \dots, j-1$, the share of money in the FEVD of output $H(i)$ quarters ahead is less than 15 percent.

¹¹Actually, as indicated in the last note, I only check the FEVDs for the horizons corresponding to 12, 18, 24, 36, and 48 months. Since money's influence on output may peak sharply as a function of horizon, whereas interest rates' influence is likely to be smoother, my spot-checking may be biased against monism.

Table 5

... And a Robust Conclusion?

The Versions of Sims' Model That Pass the Monism Tests

| Estimation Periods* | Model Characteristics | | | | | | | Model Results % Shares of Variability in | | | | | | | |
|------------------------|-------------------------|-----------|-------------------|----------|---------|-------------------|----------------------|---|----------------------------|-------------------|-------------------------------|--------|-------------------|------|------|
| | Monthly or Quarterly | Variables | | | | Number of Lags | Constant or Trend | Tests | Output Due to Shocks in | | Prices Due to Shocks in | | | | |
| | | Money | Interest Rates | Prices | Output | | | | Money | Interest Rates | | Money | | | |
| 1950-78 | Quarterly | M1 | T-Bills | PPI | GNP | 5 | C,T | 4-Year | 28.6 | 24.2 | 23.8 | | | | |
| | | | T-Bills | CPI-S | IP | 5 | C | | 23.5 | 14.4 | 48.8 | | | | |
| | | | T-Bills | CPI-S | GNP | 5 | C | | 19.2 | 4.9 | 60.5 | | | | |
| | | | T-Bills | CPI-S | GNP | 5 | C,T | | 30.8 | 23.9 | 61.4 | | | | |
| | | | T-Bills | CPI-S | IP | 9 | C | | 17.3 | 17.2 | 56.4 | | | | |
| | | | T-Bills | CPI-S | GNP | 9 | C | | 15.3 | 8.9 | 56.1 | | | | |
| | | | T-Bills | CPI-S | GNP | 9 | C,T | | 31.5 | 16.2 | 58.5 | | | | |
| | | | T-Bills | Deflator | GNP | 5 | C | | 16.6 | 9.0 | 64.6 | | | | |
| | | | T-Bills | Deflator | GNP | 5 | C,T | | 28.0 | 28.0 | 65.8 | | | | |
| | | | T-Bills | Deflator | GNP | 9 | C | | 16.0 | 6.7 | 54.4 | | | | |
| | | | T-Bills | Deflator | GNP | 9 | C,T | | 24.2 | 21.3 | 55.3 | | | | |
| | | | T-Bills | PPI | GNP | 5 | C | | Peak- Response | 16.1 | 8.9 | 25.0 | | | |
| | | | T-Bills | PPI | IP | 9 | C,T | | | 34.2 | 30.6 | 18.5 | | | |
| | | | T-Bills | CPI-S | IP | 9 | C,T | | | 36.4 | 35.5 | 60.0 | | | |
| | | | C-Paper | CPI-S | GNP | 9 | C | | 28.5 | 13.1 | 52.0 | | | | |
| | | C-Paper | CPI-S | GNP | 9 | C,T | 40.0 | 32.0 | 56.1 | | | | | | |
| | | C-Paper | Deflator | GNP | 9 | C | 27.4 | 17.3 | 54.7 | | | | | | |
| | | | | M2 | None | | | | 4-Year | | | | | | |
| | | | | | None | | | | | Peak- Response | | | | | |
| | | | | MB | T-Bills | PPI | IP | 5 | C | 4-Year | 20.9 | 15.9 | 17.0 | | |
| | | | | | T-Bills | PPI | GNP | 5 | C,T | | 24.5 | 21.0 | 18.3 | | |
| | | | | | T-Bills | CPI-S | IP | 5 | C | | 22.5 | 15.5 | 32.0 | | |
| | | | | | T-Bills | CPI-S | GNP | 5 | C,T | | 25.1 | 20.5 | 42.5 | | |
| | | | | | T-Bills | CPI-S | IP | 9 | C | | 25.3 | 15.6 | 32.0 | | |
| | | | | | T-Bills | CPI-S | IP | 9 | C,T | | 38.5 | 29.5 | 33.0 | | |
| | | | | | T-Bills | CPI-S | GNP | 9 | C,T | | 49.2 | 16.7 | 38.8 | | |
| | | | | | C-Paper | CPI-S | GNP | 9 | C,T | | 43.1 | 29.5 | 40.5 | | |
| | | | | | T-Bills | Deflator | IP | 5 | C | | 15.5 | 14.4 | 43.6 | | |
| | | | | | T-Bills | Deflator | IP | 9 | C | | 24.7 | 20.2 | 39.4 | | |
| | | | | | T-Bills | Deflator | GNP | 9 | C | | 18.5 | 6.7 | 41.1 | | |
| | | | | | T-Bills | Deflator | GNP | 9 | C,T | | 33.9 | 23.5 | 41.9 | | |
| | | | | | None | | | | | | Peak- Response | | | | |
| | | | Monthly | | M1 | T-Bills | CPI-S | IP | 25 | | C | 4-Year | 19.9 | 19.4 | 49.6 |
| | | | | | | T-Bills | PPI | GNP | 25 | | C,T | | Peak- Response | 27.8 | 22.5 |
| | | | | T-Bills | | CPI-S | GNP | 25 | C | 17.5 | 8.2 | | 51.8 | | |
| | | | | T-Bills | | CPI-S | GNP | 25 | C,T | 28.7 | 26.4 | | 56.5 | | |
| | T-Bills | Deflator | | GNP | | 25 | C | 16.7 | 14.8 | 50.6 | | | | | |
| | MB | T-Bills | | CPI-S | IP | 25 | C | 4-Year | 24.3 | 10.5 | 34.1 | | | | |
| | | T-Bills | | CPI-S | IP | 25 | C,T | | 34.4 | 22.2 | 32.7 | | | | |
| | | T-Bills | | CPI-S | GNP | 25 | C,T | | 33.7 | 19.0 | 40.9 | | | | |
| | | None | | | | | | | Peak- Response | | | | | | |
| | | None | | | | | | | | | | | | | |
| 1953-79 | None | | | | | | | | | | | | | | |

*The first period is from April 1950 through December 1978; the second, from April 1953 through September 1979.

and in detail, that temporally aggregating a model can significantly distort the estimation of dynamic relationships among variables. This suggests that the monthly models are more likely to give an accurate picture of the economy's dynamic properties. (See Christiano and Eichenbaum 1987 for a summary of some of these results as well as for comments on the implications of temporal aggregation for analysis of the money-output relationship.) On the other hand, some of Sims' critics have suggested that quarterly versions of Sims' model are more trustworthy because the monthly data are more contaminated by measurement error (King 1983, pp. 9–10). These suggestions have not been backed up, however, by a precise and detailed analysis of why measurement error would make estimates of the economy's dynamic properties taken from quarterly models more reliable than estimates taken from monthly models. So, on theoretical grounds, the argument for believing the monthly results is stronger. Still, claims for the quarterly models are buttressed by Sims' apparent position. Sims is familiar with the time series literature on the potentially misleading effects of time aggregation. But in his published responses to his critics he has never complained about their use of quarterly models. And he has himself discussed both annual and quarterly variations of his VAR (Sims 1980b, p. 254; 1987, p. 446).

I try to shed some statistical light on this issue by testing whether temporal aggregation can explain the differences between monthly and quarterly versions of models like Sims'.¹² For the specifications of Sims' model which are quarterly and meet my criteria for possible monism, I fit the comparable monthly version to the data from the 1950–78 sample period and compute the variance-covariance matrix, Σ , of the model's fitted residuals. I then use the fitted monthly model and its variance-covariance matrix to construct 250 data sets spanning from January 1948 to December 1978.¹³ That is, the fitted monthly model is the true data-generating mechanism for 250 sample data sets. Each of these data sets is temporally aggregated from monthly to quarterly. Then, to all of the quarterly data sets, I fit one quarterly model corresponding to the monthly model that was used to construct the data. In effect, I pretend to be a lucky econometrician who is able to observe history replayed in 250 random variations. I use all 250 of these replays to obtain one precise estimate of the quarterly model that best fits data generated by the hypothetically true monthly model.¹⁴ Finally, for this precise estimate of the quarterly model that would result from estimation with time-aggregated

data from the monthly model, I compute both money's and interest rates' shares in the four-year-ahead FEVD of output.

My tests suggest that temporal aggregation does distort the estimates of relationships among variables—and thus may explain much of the differences between the monthly and quarterly versions of Sims' model. For almost all of the specifications tested, temporal aggregation tends to boost the share of money and cut the share of interest rates in the four-year-ahead FEVD of output. After I adjust for these tendencies, only three quarterly models still appear to pass my tests for monism.¹⁵ Furthermore, the monthly versions of those three models already give monist results, so little is added by considering them in quarterly form.

This suggests I can ignore quarterly models. A precise and well-known effect, temporal aggregation, seems to be able to account for most of the differences between the monthly and quarterly models. The corresponding effects of data measurement errors are not well understood in this context and have not been explicitly modeled. Therefore, I conclude that the quarterly models' evidence against Sims is suspect and not a valid check on the robustness of his conclusion about monism.¹⁶

Eliminating the quarterly models from Table 5 still

¹²I thank Larry Christiano for the basic idea for these tests.

¹³For each data set, the fitted coefficients are initially applied to the actual data from January 1948 to March 1950 to get a projected money/interest rates/prices/output vector for April 1950. To this projection, I add a random disturbance vector drawn from a normal distribution with mean zero and variance-covariance matrix Σ to get the constructed data for April 1950. Then the procedure moves forward a period: the coefficients are applied to the data through April 1950 to project data for May 1950, and another random disturbance vector is independently drawn from a $N(0, \Sigma)$ distribution and added to the projection to finish constructing the data for May 1950. This procedure is repeated through December 1978.

¹⁴This is done by Kalman-filtering the 250 histories with one model. Ordinary least squares regression, corresponding to Kalman-filtering with a flat prior, is used to fit the model to the first history. For history $j, j \geq 2$, the prior coefficients and the prior $X'X^{-1}$ matrix containing information on the variances and covariances of the coefficients are set equal to the final estimates of the coefficients and the $X'X^{-1}$ matrix from the filtering of history $j-1$. After history 250, when the coefficient estimates have stabilized, fitted residuals for all 250 histories are computed using the final coefficients. These residuals are used to compute the variance-covariance matrix of the quarterly model's residuals.

¹⁵The adjustment is computed as follows. For each horizon, FEVD shares from the monthly data-generating model are subtracted from FEVD shares from the quarterly model fit to the 250 generated data sets. These differences are then subtracted from the FEVD shares of the quarterly model fit to actual quarterly data.

¹⁶My temporal aggregation adjustments don't imply, however, that temporal aggregation can fully account for the tendency of quarterly models to boost money's share in the four-year-ahead FEVD of output above 15.9 percent. This is further evidence that Sims' low estimate of money's share in this FEVD is not robust.

leaves eight monthly models that pass at least one of my tests for possible monism. All eight of these specifications include 25 rather than 13 lags, and four of the eight include a linear time trend. However, as I have discussed, I find little statistical evidence for rejecting models with these features.

Of the eight monthly models that pass my tests for possible monism, the weakest challenges to Sims' conclusion are the five involving M1. For these models, money's share in the four-year-ahead FEVD of output is not much larger than interest rates' share. In an absolute sense, too, money's share in these models is not very large. After all, according to monism, variations in the money stock are the predominant cause of business cycles.

The three monthly MB models are much harder to dismiss. Unless further analysis reveals some serious defect in these models, I would conclude that they are evidence that Sims' nonmonist conclusion is not perfectly robust for the 1950–78 period. I do view Sims' conclusion as close to robust for this period, though. His conclusion holds for almost all of the monthly models, and the few exceptions require the use of very particular price (CPI-S) and interest rate (T-bills) series.

□ *Double-Checking*

I also analyze the robustness of Sims' results for the slightly later but more statistically homogeneous 1953–79 period. This can be viewed as simply another check on Sims' results or as a check on the robustness to time period of my analysis of his results. From either point of view, this alternative time period seems appropriate for at least two reasons. It is close in time to Sims' 1948–79 period as well as to my 1950–78 period. And unlike periods that include data either before 1951 or after 1980, the 1953–79 period is relatively free of evidence that the structure of the economy changed during it.

For this slightly later time period, my analysis strongly supports the robustness of Sims' nonmonist conclusion. None of the 240 models I estimated, including the quarterly models, passes my tests for possible monism. Just as they did for the earlier period, M2 models for this period tend to give money a high share in the four-year-ahead FEVD of output. However, they also tend to give money either a weak or a negative relationship to the price level. M1 and MB models tend either to give money a small share in the four-year FEVD of output or to give money a large share that results from a strong negative response of output to money.

Sims' conclusion thus appears perfectly robust for this slightly different time period regardless of temporal aggregation, lag length, detrending, or choice of variables. Combined with the small number of acceptable models that produced even moderately strong monist results for the 1950–78 period, my unambiguous results for the 1953–79 period are strong support for Sims' claim that the evidence his VAR provides against monism is robust.

Concluding Remarks

According to my analysis, each side of the debate between Sims and his critics has made valid points. The critics appear to be right that many apparently reasonable modifications of Sims' model produce results significantly different than those Sims computed, including those on money's share of output variability. Sims, however, seems to be almost always right that these numerical instabilities don't overturn an important conclusion: the data don't support the version of monetarism he called *monism*.

Consideration of the validity of each side's point of view leads me to a few generalizations about VAR and other time series analyses:

- As did Sims (1987, p. 448), I question whether, in general, theoretical models that use generic variables like the money stock, interest rates, the price level, or output provide an adequate foundation for empirical work and whether, in model estimation, these generic variables can be more or less equally well represented by a variety of different data series.
- As did Spencer (1989, pp. 452–53), I conclude that researchers using VARs (and other time series techniques) should examine the robustness of their results with respect to the kinds of changes in model specification examined by Sims' critics.
- In judging the robustness of substantive conclusions, however, researchers often should look at a broad range of the properties of the estimated models. Focusing on just one statistic, such as money's share in a measure of the variability of output, to the exclusion of a more comprehensive examination of the models' dynamic properties, may be misleading.
- Modifications proposed to test the robustness of a result should be screened on theoretical and statistical grounds to confirm that the modifications are neither implausible nor inherently biased against the result.

Practically, of course, all this robustness-checking is not easy—or cheap. But technological advances will no doubt minimize those problems eventually. And researchers could well have deeper problems if they don't check the robustness of their results.

Appendix Identifying Monetary Policy Shocks With Causal Chains

In the work of Christopher Sims (1980b) and in the preceding paper, the fitted error vectors of vector autoregression models are estimates of reduced-form error vectors. These vectors are in turn assumed to be functions of unobserved underlying shocks to monthly supply and demand relationships. In this Appendix, I will describe and illustrate the causal chains Sims and I use to relate the shocks to the theoretical relationships.

Sims assumes a causal chain running from interest rates to money and then to prices and to output. This assumption implies that the reduced-form vectors are related to the underlying shocks according to this model:

$$(A1) \quad e_{rt} = u_{rt}$$

$$(A2) \quad e_{mt} = \alpha_m u_{rt} + u_{mt}$$

$$(A3) \quad e_{pt} = \alpha_p u_{rt} + \beta_p u_{mt} + u_{pt}$$

$$(A4) \quad e_{yt} = \alpha_y u_{rt} + \beta_y u_{mt} + \gamma_y u_{pt} + u_{yt}$$

Here the e 's denote reduced-form errors; the u 's, underlying shocks to intramonth supply and demand; and the α , β , and γ 's, coefficients linking the underlying shocks to the reduced-form errors. The subscripts t , r , m , p , and y refer to time, interest rates, money, prices, and output.

This model could be interpreted in at least two ways.

One is the interpretation Sims probably had in mind, which is shown in Figure A1. Here the intramonth money supply curve, equation (A1), is horizontal and the intramonth money demand curve, equation (A2), is downward-sloping. Monetary policy can be tightened by setting u_{rt} to a positive number. This policy shock to interest rates shifts up the money supply curve, which leads to an unpredicted decline in the equilibrium quantity of money as consumers and firms move up along the intramonth money demand curve. Note that this interpretation requires that the coefficient α_m in (A2), the slope of the money demand curve, be negative.

The other interpretation of the model is illustrated in Figure A2. Here the curves' positions are more or less reversed. Now equation (A1) is a horizontal intramonth money demand curve, and equation (A2) is an upward-sloping intramonth money supply curve. Under this interpretation, monetary policy would be tightened by setting u_{mt} in (A2) to a positive number, producing an unanticipated upward shift in the supply curve. This would cut the equilibrium quantity of money, but leave interest rates

Figures A1–A4

Two Ways to Interpret Tighter Monetary Policy
Using Two Causal Chains

When Money Demand Slopes Down

When Money Supply Slopes Up

Figures A1–A2
Chain
From
Interest Rates
to Money

Figure A1 Money Supply
Horizontal

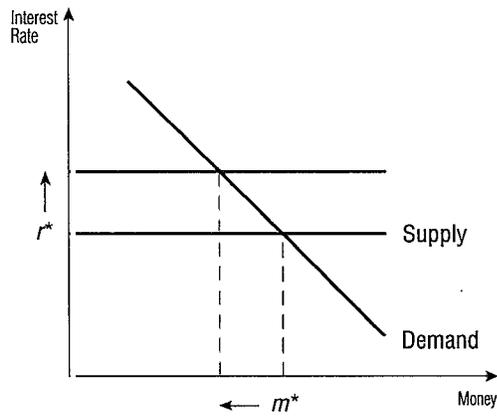
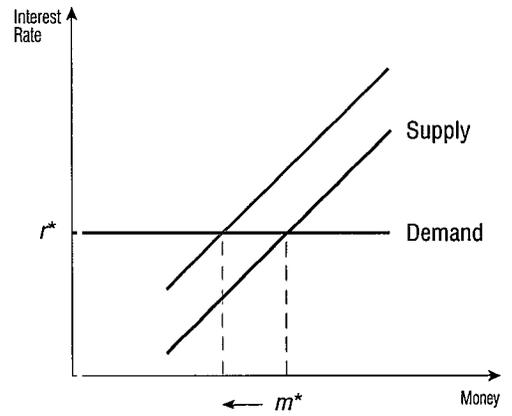


Figure A2 Money Demand
Horizontal



Figures A3–A4
Chain
From
Money
to Interest Rates

Figure A3 Money Supply
Vertical

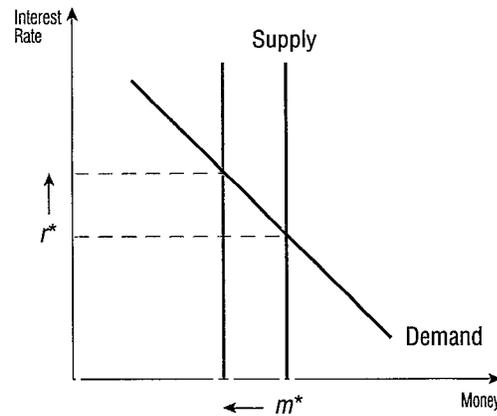
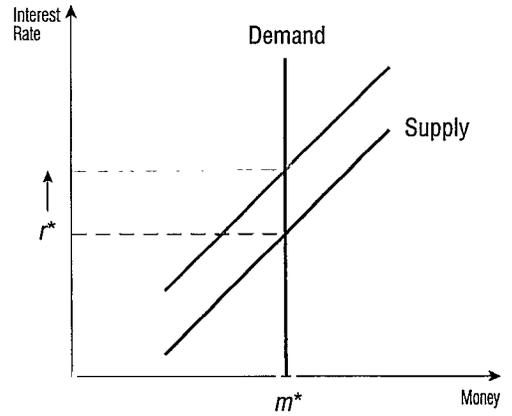


Figure A4 Money Demand
Vertical



unchanged. Month-to-month surprise movements in interest rates would be caused only by shocks to money demand. These shocks would induce unpredicted intramonth changes in the quantity of money, as the monetary authority slid along its supply curve in response to the unpredicted interest rate movement. This interpretation may seem less plausible than the other. But it is the only interpretation consistent with equations (A1)–(A4) when α_m , now the slope of the money supply curve, is positive.

In the preceding paper, I use a different causal chain that runs from money to interest rates. This different order requires me to replace equations (A1) and (A2) with

$$(A5) \quad e_{mt} = u_{mt}$$

$$(A6) \quad e_{rt} = \alpha_r u_{mt} + u_{rt}$$

where the symbols all have the same meanings.

Figure A3 illustrates my causal chain when α_r , the slope of the money demand curve, is negative, as it typically is in the models I estimate. Equation (A5) then represents a vertical intramonth money supply curve, while equation (A6) represents a downward-sloping intramonth money demand curve. Monetary policy would be tightened by setting the monetary policy shock, u_{mt} , to a negative number, which would shift the intramonth supply curve to the left. This unpredicted decrease in the quantity of money would tend to simultaneously produce an unpredicted rise in interest rates, as consumers and firms slid up along their intramonth demand curves.

If α_r were instead the positive slope of the money supply curve, my causal chain would imply the money supply and demand relationships shown in Figure A4. Here, equation (A5) is represented by a vertical intramonth money demand curve; equation (A6), by an upward-sloping intramonth money supply curve. Monetary policy would be tightened by setting positive values of u_{rt} . These would produce unanticipated upward shifts in the money supply curve and higher interest rates, but no change in the equilibrium quantity of money.

References

- Akaike, Hirotosugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19: 716–23.
- Amirizadeh, Hossain. 1985. Technical appendix for the Bayesian vector autoregression model of the U.S. economy. Research Department, Federal Reserve Bank of Minneapolis.
- Bernanke, Ben S. 1986. Alternative explanations of the money-income correlation. In *Real business cycles, real exchange rates, and actual policies*, ed. Karl Brunner and Allan H. Meltzer. Carnegie-Rochester Conference Series on Public Policy 25 (Autumn): 49–99. Amsterdam: North-Holland.
- Christiano, Lawrence J. 1988. Searching for a break in GNP. Working Paper 2695. National Bureau of Economic Research.
- Christiano, Lawrence J., and Eichenbaum, Martin. 1987. Temporal aggregation and structural inference in macroeconomics. In *Bubbles and other essays*, ed. Karl Brunner and Allan H. Meltzer. Carnegie-Rochester Conference Series on Public Policy 26 (Spring): 63–130. Amsterdam: North-Holland.
- Cooley, Thomas F., and LeRoy, Stephen F. 1985. Atheoretical macroeconomics: A critique. *Journal of Monetary Economics* 16 (November): 283–308.
- Doan, Thomas A. 1988. User's manual: RATS (Regression analysis of time series). Version 3.00. Evanston, Ill.: VAR Econometrics.
- Eichenbaum, Martin, and Singleton, Kenneth J. 1986. Do equilibrium real business cycle theories explain postwar U.S. business cycles? In *NBER Macroeconomics Annual 1986*, ed. Stanley Fischer, 1: 91–135. Cambridge, Mass.: MIT Press.
- Friedman, Milton. 1969. *The optimum quantity of money and other essays*. Chicago: Aldine.
- Friedman, Milton, and Schwartz, Anna Jacobson. 1963. *A monetary history of the United States, 1867–1960*. Princeton, N.J.: Princeton University Press.
- King, Stephen R. 1983. Real interest rates and the interaction of money, output, and prices. Manuscript. Northwestern University.
- Leamer, Edward E. 1985. Vector autoregressions for causal inference? In *Understanding monetary regimes*, ed. Karl Brunner and Allan H. Meltzer. Carnegie-Rochester Conference Series on Public Policy 22 (Spring): 255–303. Amsterdam: North-Holland.
- McCallum, Bennett T. 1983. A reconsideration of Sims' evidence concerning monetarism. *Economic Letters* 13: 167–71.
- Poole, William. 1978. *Money and the economy: A monetarist view*. Reading, Mass.: Addison-Wesley.
- Runkle, David E. 1987. Vector autoregressions and reality. *Journal of Business and Economic Statistics* 5 (October): 437–42.
- Schwarz, Gideon. 1978. Estimating the dimension of a model. *Annals of Statistics* 6 (March): 461–64.
- Sims, Christopher A. 1980a. Macroeconomics and reality. *Econometrica* 48 (January): 1–48.
- . 1980b. Comparison of interwar and postwar business cycles: Monetarism reconsidered. *American Economic Review* 70 (May): 250–57.
- . 1986. Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review* 10 (Winter): 2–15.
- . 1987. Comment. *Journal of Business and Economic Statistics* 5 (October): 443–49.
- . 1989. Models and their uses. *American Journal of Agricultural Economics* 71 (May): 489–94.
- Spencer, David E. 1989. Does money matter? The robustness of evidence from vector autoregressions. *Journal of Money, Credit, and Banking* 21 (November): 442–54.
- Whiteman, Charles H., and Roberds, William. 1989. Monetary aggregates as monetary targets: A statistical investigation. Manuscript. University of Iowa and Federal Reserve Bank of Atlanta.
- Whittle, P. 1953. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society, Series B*, 15 (No. 1): 125–39.

The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.