FINANCIAL MARKET VARIABLES DO NOT PREDICT REAL ACTIVITY

MARK A. THOMA and JO ANNA GRAY*

The past decade has seen an extensive empirical reassessment of the information content of financial market variables sensitive to monetary policy. Particularly provocative are recent papers suggesting that some interest rates and interest rate spreads contain more information about economic activity than monetary aggregates. This paper reviews important methodological pitfalls in these studies. We then show that none *of the commonly employed measures of monetary policy contain incremental information useful in forecasting real economic activity. Two conclusions are possible. Either monetary policy innovations have no significant real effects, or we (collectively) have failed in our efforts to measure monetary policy.* (*JEL* E52)

I. INTRODUCTION

The past decade has seen an extensive empirical reassessment of the information content of financial market variables believed to be sensitive to monetary policy. Particularly provocative are papers suggesting that some interest rates and interest rate spreads contain more information about economic activity than monetary aggregates. This paper reviews important methodological pitfalls in these studies. We then show that *none* of the commonly employed indicators of monetary policy contain incremental information useful in forecasting real economic activity, as measured by industrial production. Two conclusions are possible. Either monetary policy innovations have no significant real effects, or we (collectively) have failed in our efforts to measure monetary policy.

The money-income causality literature spawned by Christopher Sims in the 1970s has broadened in recent years to encompass a diverse set of measures of monetary policy, more comprehensive and sophisticated empirical assessment strategies, and a richer array of explanations for observed correlations between financial market variables and economic activity. "Early" contributors to this literature, such as Stock and Watson [1989], Friedman and Kuttner [1992], and Bernanke and Blinder [1992], present evidence that particular interest rates and spreads not only dominate monetary aggregates as predictors of economic performance, but are remarkably powerful predictors. Stock and Watson [1989] and Friedman and Kuttner [1992] demonstrate that the spread between the interest rates on commercial paper and treasury bills is highly significant in explaining movements in real activity, while Bernanke and Blinder [1992] made an equally convincing case for the federal funds rate. These conclusions are qualified, however, in subsequent work. Both Bernanke [1990] and Stock and Watson [1993], for example, point out that the predictive power of the paper-bill spread (and interest rates more generally) weakened during the second half of the 1980s and the early 1990s.

The results reported in these and related papers form the empirical core of the debate over the relative merits of various interest rates, spreads and monetary aggregates as indicators of the stance of monetary policy. Subsequent empirical work has seen a shift from "atheoretical" to "structural" vector autoregressions. Strongin [1992], Gordon and

* Author order was determined by a coin toss.

Thoma: Associate Professor, Department of Economics, University of Oregon, Eugene, Phone 1-541-346-1266, Fax 1-541-346-1243
E-mail mthoma@oregon.uoregon.edu

Gray: Professor, Department of Economics, University of Oregon, Eugene, Phone 1-541-346-4667
Fax 1-541-346-1243
E-mail jgray@oregon.uoregon.edu

ABBREVIATIONS

SP: Paper-Bill Spread
FF: Federal Funds Rate
RMSE: Root Mean Square Error

Economic Inquiry
(ISSN 0095-2583)
Vol. XXXVI, October 1998, 522–539

Leeper [1994] and Bernanke and Mihov [1995], for example, focus on more refined measures of monetary policy derived from explicit models of the reserve market and central bank operating procedures. These papers address concerns that go beyond discriminating among variables on the basis of predictive power. They grapple directly with the fundamentally difficult issue of identifying monetary policy shocks by positing and estimating structural models that take the form of identified VARs. The usefulness of these efforts as well as earlier efforts to measure monetary policy is, in turn, questioned by Rudebusch [1996], who finds that the studies' econometric results are fragile and at odds with other evidence on the nature of the Fed's reaction function and monetary policy surprises.

The first part of our paper illustrates some of the methodological pitfalls common in the literature cited above, and complements recent work by Bernanke [1990], Hess and Porter [1993], Thoma and Gray [1995], Rudebusch [1996], and Emery [1996]. We begin by demonstrating the advantages of rolling (recursive) regression techniques that directly and comprehensively address the question of sample sensitivity. These techniques are used to illustrate the extreme sensitivity to sample period of the causality statistics and variance decompositions commonly used to assess the explanatory power of financial market variables. Indeed, in the applications under review here, such in-sample measures of fit are sometimes heavily influenced by *individual* monthly observations. Because it also emerges as an important observation in the second half of the paper, we focus particularly on 1974:12, showing that this one observation accounts for the uniformly superior performance of the paper-bill spread reported in many earlier studies. This portion of the paper concludes with a striking illustration of the drawbacks of in-sample measures of fit, which we show to be misleading indicators of out-of-sample measures (forecast errors). Since out-of-sample measures of fit are generally regarded as "the ultimate test of an equation" (Bernanke [1990, 59]), the demonstration suggests the need for a reassessment of the conclusions drawn in earlier studies that rely primarily on in-sample measures of fit.

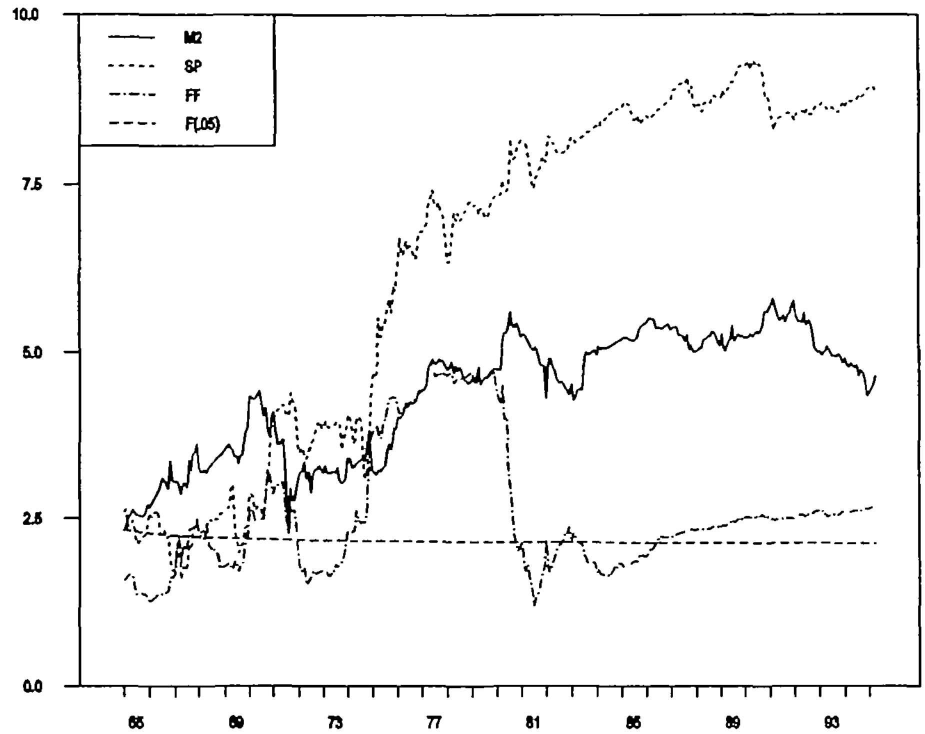
The second half of the paper departs from the common practice of framing empirical exercises as horse races between competing financial market variables. This orientation, along with an emphasis on in-sample measures of fit, has diverted attention from the more fundamental question of whether *any* of the commonly employed measures of monetary policy contain incremental information useful in predicting real activity. We address this question by comparing the out-of-sample forecasting power of a simple autoregressive model of industrial production to the predictive power of models that include the paper-bill spread, the federal funds rate, and M2.

Once again, rolling regression techniques allow us to identify and evaluate the effects of influential observations. None of the financial market variables considered aid systematically in forecasting industrial production, whether the variables are considered alone or in combination. We argue, however, that the outliers present in the data in 1974 could lead one to conclude, incorrectly, that the paper-bill spread contains information generally useful in forecasting real activity. We conclude, as noted at the outset, that either monetary policy innovations have no significant real effects, or we (collectively) have failed in our efforts to measure monetary policy.

II. THE SENSITIVITY OF CAUSALITY STATISTICS TO SAMPLE PERIOD

This section reviews important elements of the recent debate over the relative predictive power of financial market variables. Using familiar model specifications and in-sample causality statistics, we evaluate the explanatory power of three commonly studied financial market variables, the federal funds rate, the paper-bill spread, and M2. In general, we confirm the findings of others: The paper-bill spread is highly significant in explaining industrial production growth over much of the post-war period, producing larger causality statistics than either of the other two financial variables for any sample period ending after the mid-1970s. However, test results suggest that M2 and the federal funds rate also contain information useful in explaining output. Our empirical methodology provides a clear picture of the sensitivity of test-statistics to small variations in sample period and, occasionally, to individual observations. For example, as one would expect, the explanatory power of

FIGURE 1
Tests of the Hypotheses that M2, SP, and FF Do Not Cause Output Growth



Vertical axis: F-statistics and .05 critical value.
Horizontal axis: End-date of the sample.
Estimated equation:

$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + u_t .$$

where y and p are the growth rates of industrial production and the CPI less shelter; f is either M2, SP, or FF; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

the federal funds rate plummets in late 1979 with the shift in Federal Reserve operating procedures. Other notable dates include 1974:12, which is important in evaluating the predictive power of the paper-bill spread and useful in interpreting the results reported in subsequent sections of the paper.

Figure 1 reports F-statistics for tests of the hypotheses that lagged values of M2, the Paper-Bill Spread (SP), and the Federal Funds Rate (FF) do not Granger cause the growth rate of industrial production in the empirical model given by equation (1):

$$(1) \qquad y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + u_t$$

Here y is the growth rate of industrial production, p is the CPI inflation rate, f is one of three financial variables noted above (M2, SP, or FF), and u is a disturbance term.[1] The plots show how the explanatory power of each financial variable changes as the end of the sample period is extended. For each variable, equation (1) is first estimated over the period 1960:2 (after lags and differencing) through 1965:1 and a test of the hypothesis that the variable does not cause output growth is conducted. One month is then added to the data set so that the sample covers 1960:2 through 1965:2 and the estimation and test are repeated. The process of adding one month to the data and repeating the causality test continues until the entire data set, 1960:2 through 1995:4, is used in performing the test. For each financial variable, the result is 364 F-statistics with different degrees of freedom. These F-statistics, along with their critical values, are plotted in Figure 1 with the end-date of the sample that generated each statistic measured on the horizontal axis.

Figure 1 "nests" the results reported in many other studies. For example, Friedman and Kuttner's finding that the paper-bill spread dramatically outperforms M2 in explaining output growth for sample periods ending in the third quarter of 1979 and in the fourth quarter of 1992 is confirmed by the F-statistics indexed 1979:9 and 1990:12, which are much higher for the paper-bill spread than for M2. Figure 1 also documents a significant relationship between the federal funds rate and output throughout most of the 1970s, as reported by Bernanke and Blinder [1992], along with a dramatic decline in the explanatory power of the funds rate in late 1979 when the Federal Reserve shifted its operating procedure from targeting the funds rate to targeting non-borrowed reserves. Finally, as one would expect in a specification that does not include an interest rate, M2 is significant in explaining industrial production throughout the sample.

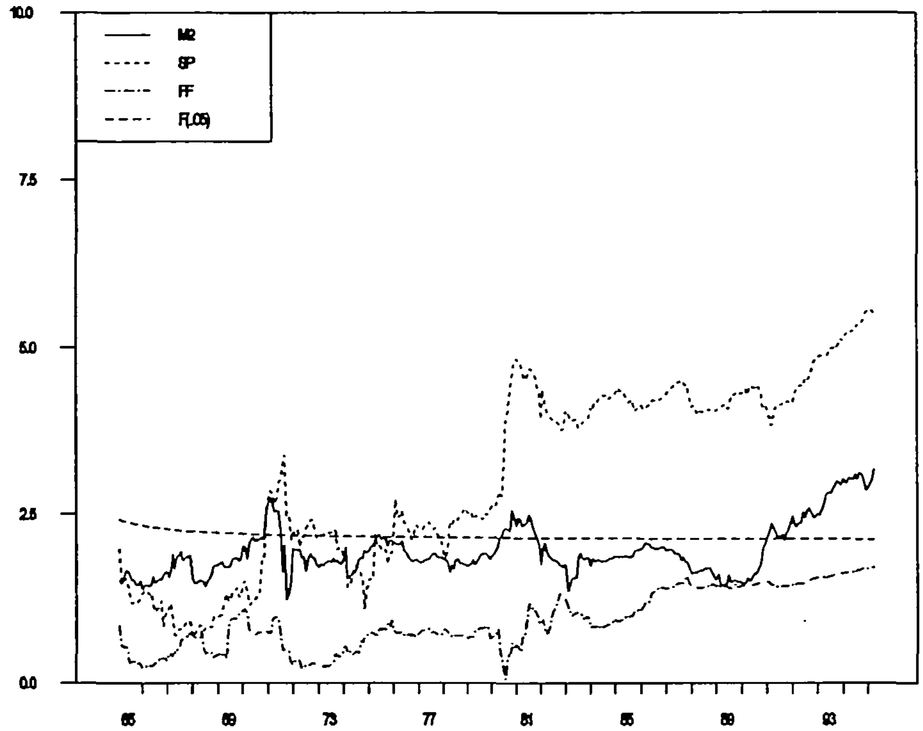
Figure 2 takes these baseline exercises a step further. Here we record the results of the same exclusion tests reported in Figure 1, except that M2, the paper-bill spread, and the federal funds rate are allowed to compete against each other in the same model. That is, equation (1) is expanded to include all three financial variables simultaneously. The high degree of collinearity among the three variables accounts for the relatively low F-statistics recorded in the figure. Only the paper-bill spread stands out in the figure; it is consistently significant in explaining output, and clearly dominates the other two financial variables over the second half of the sample. As one would expect based on the work of Sims and others, the performance of M2 is adversely affected by the presence of interest rate variables in the estimated model. M2 does not cause output growth in this specification over most samples ending before 1990. Perhaps more surprising are the results reported for the federal funds rate. Contrary to the causality statistics reported in Bernanke and Blinder [1992], the federal funds rate performs the most poorly of the three indicators for every sample period considered.[2]

Figures 1 and 2 provide a compact summary of the sample-sensitivity of our test-statistics. We noted earlier, for example, the decline in the explanatory power of the federal funds rate that accompanied the shift in Federal Reserve operating procedures in late 1979. Another date that stands out is 1974:12, which produces large increases in the test statistics depicted in Figure 1 for the paper-bill

1. Our model specifications were chosen to be both representative and parsimonious. In selecting lag length, we confirm the common choice of six monthly lags using both the Akaike and Schwarz information criteria. Indeed, in all models reported in our paper except one, three lags are sufficient. The one exception requires six lags. The usual battery of tests for serial correlation, including Q-tests for Q(1) through Q(36), fails to indicate serial correlation in all cases. All data are from Citibase and include industrial production (IP), the CPI less shelter (PUXHS), the secondary market rate on six-month treasury bills (FYGM6), the six-month prime commercial paper rate (FYCP), M2 (FM2), and the federal funds rate (FYFF). Industrial production, the CPI measure, and the money supply series are all seasonally adjusted. The interest rate series are not seasonally adjusted. The use of the CPI less shelter follows Litterman and Weiss [1985]. Our conclusions are unaffected if the CPI is used in place of the CPI less shelter.

2. Table I of Bernanke and Blinder [1992] reports marginal significance levels for the federal funds rate that are dramatically higher than those for M2. There are numerous differences between our studies, most of which are inconsequential. However, this discrepancy in results is due to a computational error in the Bernanke and Blinder study. While the error significantly effects some of the F-statistics reported by Bernanke and Blinder, it has little effect on the corresponding variance decompositions reported in the paper. We thank Ben Bernanke for providing assistance that allowed us to confirm and correct the error.

FIGURE 2
Tests of the Hypotheses that M2, SP, and FF Do Not Cause Output Growth



Vertical axis: F-statistics and .05 critical value.
Horizontal axis: End-date of the sample.
Estimated equation:

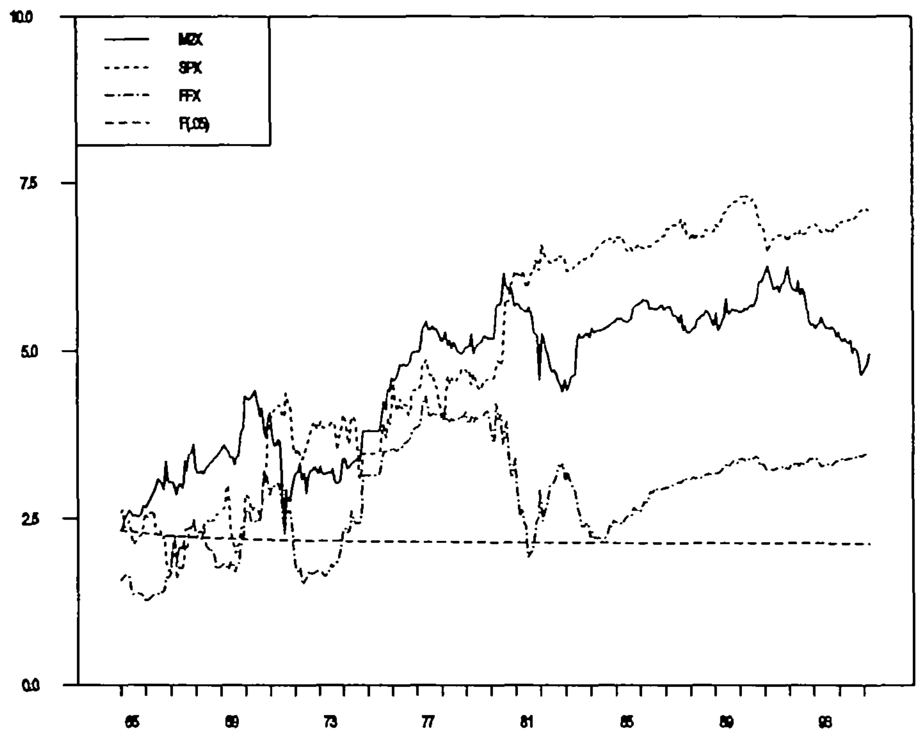
$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i M2_{t-i} + \sum_{i=1}^{6} \eta_i SP_{t-i} + \sum_{i=1}^{6} \rho_i FF_{t-i} + u_t$$

where y and p are the growth rates of industrial production and the CPI less shelter; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

spread and the federal funds rate. This date appears to be particularly important in evaluating the paper-bill spread. Figures 1 and 2 both show that the clear-cut dominance of the paper-bill spread reported by Friedman and Kuttner and others does not appear until the sample is extended to include 1974:12. It is difficult to dismiss as coincidence the striking nature of the data in late 1974. The time series for the growth rate of industrial production exhibits by far its largest negative value (and its largest absolute value) in 1974:12, while the paper bill spread reaches a value almost double any other post-war high in 1974:07. In short, a record high in interest rates and spreads in mid-1974 preceded a record low growth rate in late 1974, raising the question of the extent to which the dominance of the paper-bill spread in explaining output growth can be attributed to a single observation.

This question is addressed in Figures 3 and 4, which record the results of a sequence of causality tests identical to those recorded in Figures 1 and 2 except that they are conducted

FIGURE 3
Tests of the Hypotheses that M2, SP, and FF Do Not Cause Output Growth: 1974:12 Removed



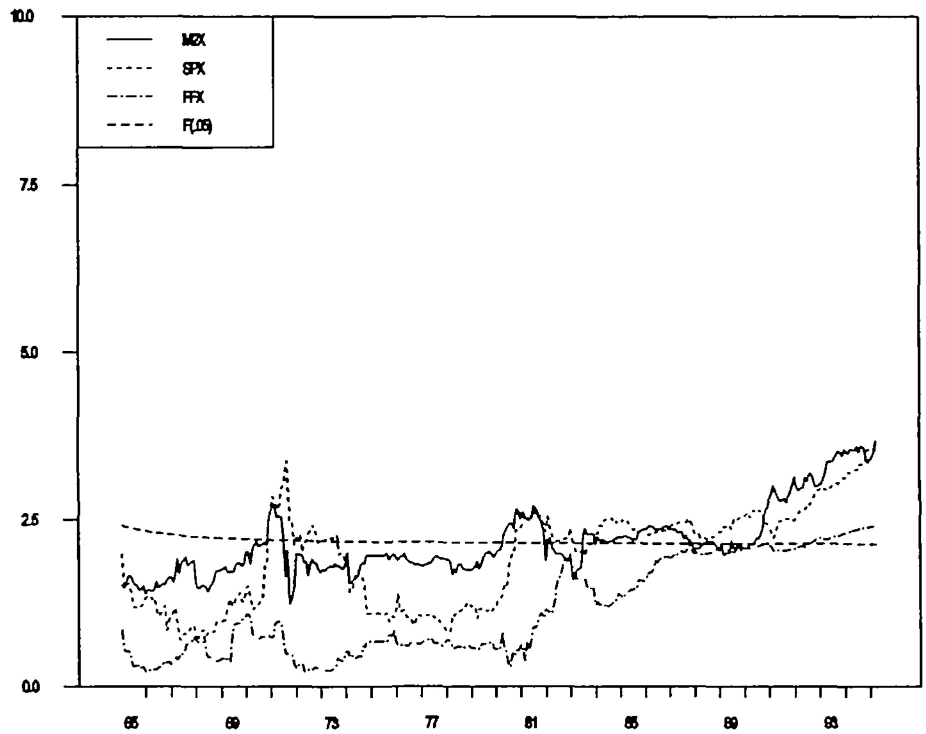
Vertical axis: F-statistics and .05 critical value.
Horizontal axis: End-date of the sample.
Estimated equation:

$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + u_t,$$

where y and p are the growth rates of industrial production and the CPI less shelter; f is either M2, SP, or FF; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

on a data set from which 1974:12 has been excluded. Six lags of all the explanatory variables, including industrial production, are used in the empirical model. Accordingly, eliminating the effects of 1974:12 entirely requires deleting its effects in the six months subsequent to 1974:12 as well. The repercussions of the modification are dramatic. The figures show that, once the effects of 1974:12 are removed, there is little to distinguish M2 and the paper-bill spread as predictors of industrial production over most sample periods. Indeed, in the specification in which M2 and the spread are allowed to compete against each other, the two variables are virtually indistinguishable over the past 15 years. Eliminating 1974:12 has less dramatic effects on the predictive power of the federal funds rate, which continues to perform relatively poorly. We conclude that the systematically superior

FIGURE 4
Tests of the Hypotheses that M2, SP, and FF Do Not Cause Output Growth: 1974:12 Removed



Vertical axis: F-statistics and .05 critical value.
Horizontal axis: End-date of the sample.
Estimated equation:

$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i M2_{t-i} + \sum_{i=1}^{6} \eta_i SP_{t-i} + \sum_{i=1}^{6} \rho_i FF_{t-i} + u_t$$

where y and p are the growth rates of industrial production and the CPI less shelter; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

explanatory power of the paper-bill spread is a post-1974 phenomenon. Furthermore, its dominance beginning in late 1974 depends to a very large extent on one observation, 1974:12.[3]

There are other outliers in the data that have smaller, but still notable, effects on the causality statistics. Eliminating from the data the observations immediately before and after 1974:12, which also involve large negative output growth rates, reduces even further the differences in explanatory power between M2 and the spread after 1974. Observations in

3. Since F-statistics can be misleading indicators of causality in multi-variable systems, we have also examined rolling variance decompositions for the empirical models that produced Figures 2 and 4. As one might expect, ordering matters. The financial variable that appears first in the ordering generally explains more of industrial production than either of the other two financial variables. The variance decompositions, like the causality statistics reported in Figure 2, show sensitivity to 1974:12. Its effect on the proportion of output variance attributable to the federal funds rate is particularly dramatic; the only period over which the federal funds rate dominates the other financial variables is the five years immediately following 1974:12. Excluding 1974:12 from the data produces a situation similar to that shown in Figure 4; in the absence of 1974:12 there is little to distinguish the performance of M2, the paper-bill spread, and the federal funds rate.

mid-1980, when the U.S. experienced another marked decline in the growth rate of industrial production as well as record high-interest rates, also appear to be influential. As Emery [1996] emphasizes, 1980 corresponds to the Carter credit controls, which significantly disrupted credit markets. Eliminating the mid-1980s from the data in addition to the three-month episode in 1974 further reduces any differences in the causality statistics for M2 and the paper-bill spread. Eliminating the mid-1980s and the three-month episode in 1974 also produces a "best case" for the federal funds rate, although the causality statistics for M2 generally remain above those for the federal funds. This paper focuses on 1974:12 because it produces the sharpest illustration of the points we wish to make. However, the importance of other outliers, particularly 1980, are also noted in the course of the analysis.

It is evident that our objective in this section is not a comprehensive overview of the literatures on leading indicators and money-income causality, which span multiple decades. Rather, in Figures 1 and 2 we attempt to characterize the flavor of the many recent contributions to these literatures that focus primarily on in-sample measures of fit and on interest rates and spreads as leading indicators and measures of monetary policy. Similarly, we do not attempt to survey the large literatures dealing with the identification and repercussions of influential observations. However, Figures 3 and 4 demonstrate that outliers can heavily influence—indeed, determine—conclusions drawn from the data in the class of studies with which we are concerned.[4]

III. EVALUATING FIT: IN-SAMPLE VERSUS OUT-OF-SAMPLE METRICS

This section considers the question of how to evaluate the performance of a model. We argue, as have many others, that out-of-sample measures of fit are the correct metric,[5] and that in-sample measures of fit may be misleading indicators of out-of-sample measures. Figures 5 through 7 report measures of the forecasting power of models that include M2 growth, the paper-bill spread, and the federal funds rate. Results are reported for a forecast horizon of three months in Figure 5, for a horizon of six months in Figure 6, and for a horizon of nine months in Figure 7. As in the case of the causality statistics discussed earlier, we use techniques that show how the forecasting power of each model changes as the sample period is expanded.

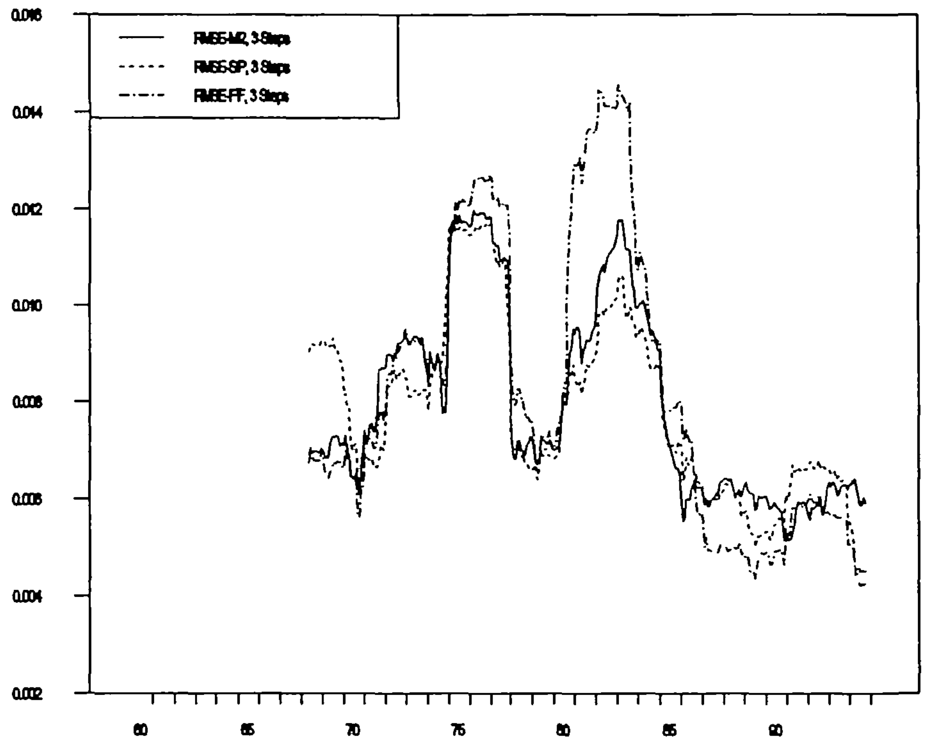
Our measure of forecasting power is the Root Mean Square Error (RMSE) of 36 consecutive forecasts, with the date of the last of the 36 forecast errors used to index the measure. Thus, for example, the first statistics recorded in Figure 5 are dated 1968:4 and are calculated as follows: For each of the three financial variables, the six-lag vector autoregression corresponding to equation (1) is estimated over the period 1960:2 through 1965:2.[6] Using the estimated coefficients from these regressions and dynamic forecasting techniques, we generate a forecast of output in 1965:5. We then update the sample by one observation (1965:3), reestimate the model, and generate a forecast for output in 1965:6. We continue updating and generating 3-step-ahead forecasts until the model is estimated over 1960:2 through 1968:1 and used to generate a forecast for 1968:4. The root mean square error of the resulting 36 forecasts is dated 1968:4 in Figure 5. Similarly, the statistics dated 1968:5 in Figure 5 are calculated from output growth forecasts for 1965:6 through 1968:5, with the forecasts generated as just described, and so on. Of course, low root mean square errors indicate high forecasting ability and large values of the statistic indicate poor predictive power.

4. We have, of course, explored the robustness of our results with respect to a number of variations in specification, data, and assessment criterion. We have estimated models in levels and in first-differences, with time trends, with error-correction terms, with 12 lags rather than six, using the unemployment rate in place of industrial production, with forecast horizons ranging from one to 36 months, and with root-mean-square errors averaged over as few as 12 and as many as 60 months. None of these variations caused us to question the generality of the paper's main conclusions.

5. See, for example, Bernanke [1990, 59], who asserts that "the ultimate test of an equation is the ability to forecast out of sample."

6. Estimation of the complete VAR is required in order to generate forecasts of the explanatory variables in equation (1) for 1965:3 and 1965:4. These forecasts are used in equation (1) in place of the actual values of the explanatory variables (which, of course, would be unknown in 1965:2) to generate the predicted value of output growth in 1965:5.

FIGURE 5
Root-Mean-Square-Errors of Models including M2, SP, or FF:
Three-Month Forecast Horizon



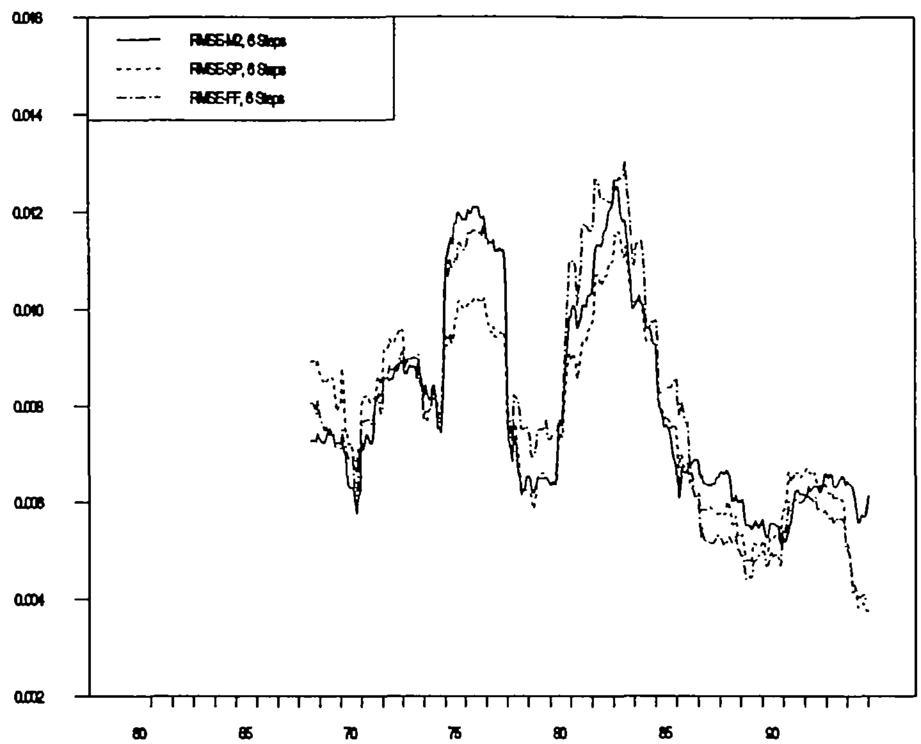
Horizontal axis: Date of last forecast error used in calculating RMSEs.
Vertical axis: RMSE of 36 forecasts generated from the equation:

$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + u_t,$$

where y and p are the growth rates of industrial production and the CPI less shelter; f is either M2, SP, or FF; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

Like the in-sample measures of fit discussed earlier, forecasting power varies considerably with sample period. Furthermore, comparing Figures 5 through 7 to Figures 1 and 2 shows clearly that in-sample measures of fit do not provide reliable indicators of out-of-sample fit. Note, in particular, the dramatic deterioration in the forecasting ability of all three models in late-1974. This contrasts sharply with the large jump in causality statistics shown in Figure 1 for the spread and the funds rate as the sample is extended through late-1974. Likewise, the impact of 1974:12 on the *relative* forecasting power of the three variables cannot be predicted on the basis of the causality statistics in Figure 1. Observe, for example, that the causality statistics for the paper-bill increase a great deal in 1974:12 while those for M2 are not noticeably affected. Nonetheless, the deterioration in out-of-sample forecasting power is very similar for M2 and the paper-bill spread at the three-month horizon, and larger for the paper-bill spread at the nine-month horizon). Only at the six-month forecast horizon does the paper-bill spread predict "better" than M2

FIGURE 6
Root-Mean-Square-Errors of Models including M2, SP, or FF:
Six-Month Forecast Horizon



Horizontal axis: Date of last forecast error used in calculating RMSEs.
Vertical axis: RMSE of 36 forecasts generated from the equation:

$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + u_t,$$

where y and p are the growth rates of industrial production and the CPI less shelter; f is either M2, SP, or FF; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

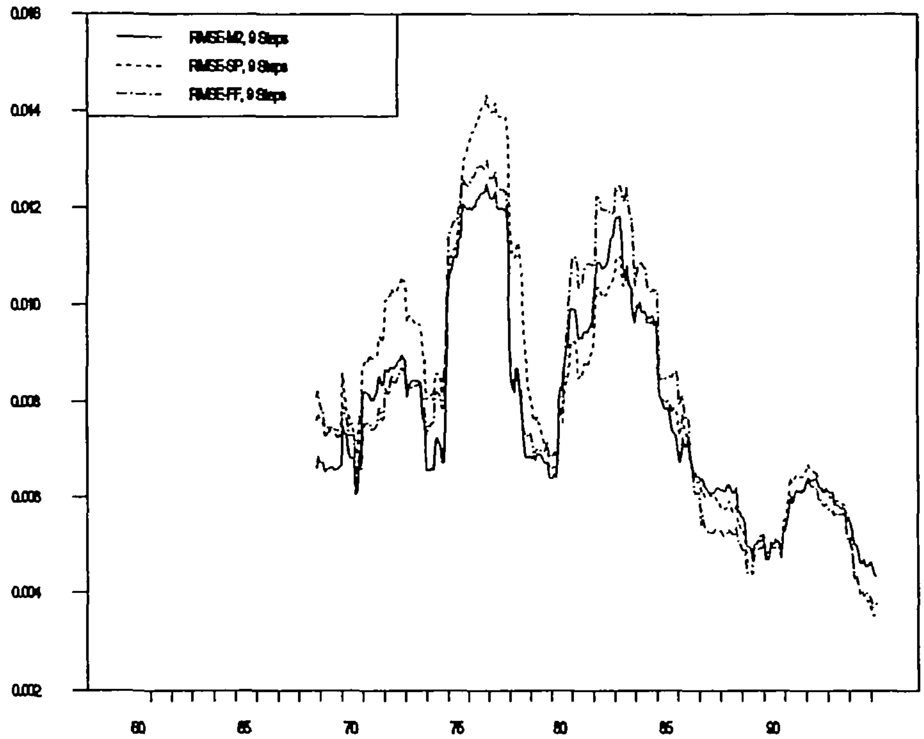
when the sample is extended past 1974, as one might expect based on causality statistics.[7]

Figures 5 through 7 suggest that there is little to distinguish the *average* performance of M2, the paper-bill spread, and the federal funds rates in predicting economic performance over the sample as a whole. This impression can be confirmed by constructing confidence intervals for the pairwise differences in the RMSEs reported in Figures 5 through 7.[8] Significant differences in forecasting power are transitory and tend to be specific to particular forecast horizons and pair-

7. The results reported in Figures 5 through 7 are representative of those produced by a "finer cut." For example, the deterioration in the forecasting abilities of M2 and the paper-bill spread are comparable horizons of one, two, and three months. The deterioration is greater for M2 at horizons of four, five, and six months. The deterioration is comparable once again at a horizon of seven months. Any the deterioration is greater for the paper-bill spread at horizons of seven to 18 months.

8. Two-sided symmetric confidence intervals were constructed using the bootstrap procedure described in Christiano and Ljungqvist [1988]. The results discussed here are based on 95% confidence intervals and 100 draws. We are aware of recently voiced concerns about the theoretical foundations and performance of the bootstrap procedure we have employed, but are unaware of a preferred alternative at this time.

FIGURE 7
Root-Mean-Square-Errors of Models including M2, SP, or FF:
Nine-Month Forecast Horizon



Horizontal axis: Date of last forecast error used in calculating RMSEs.
Vertical axis: RMSE of 36 forecasts generated from the equation:

$$y_t = \alpha + \sum_{i=1}^{6} \lambda_i y_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + u_t,$$

where y and p are the growth rates of industrial production and the CPI less shelter; f is either M2, SP, or FF; M2 is the growth rate of the monetary aggregate M2; SP is the difference between the six-month commercial paper and treasury bill rates; FF is the federal funds rate. The data are described in footnote 1.

wise comparisons. For example, for sample periods ending during the mid-1970s and a forecast horizon of six months, the paper-bill spread produces lower RMSEs than M2 and the federal funds rate, but only its advantage over M2 is statistically significant. The spread's relatively lower RMSEs at the six-month horizon over this period are due primarily to its effectiveness in forecasting industrial production in 1974:12, which influences our moving average measure of forecasting ability for the 36 months dated 1974:12 through 1977:11. Note, however, that the spread has no significant advantage over the other two financial variables during this same time period if the forecast horizon is set at three months or nine months. Indeed, at a forecast horizon of nine-months both M2 and the funds rate outperform the spread, and for the funds rate the difference is significant.

The comparisons of out-of-sample forecasting power reported in this section do turn up an interesting regularity that we will return to in the next section. The RMSEs of models containing the paper-bill spread and the federal funds rate decline to their lowest sample values as the sample is extended through very recent years. This is true at every forecast horizon. The M2 model, however, does not show the same improvement. Indeed, at forecast horizons of three and six months, its performance deteriorates somewhat during the first half of the 1990s. Thus, at shorter forecast horizons, a significant difference between the performance of the interest rate models and the performance of the M2 model emerges during the early 1990s. As we will show in the next section, the reason for the improved performance of the spread and funds rate models is *not* that interest rates have suddenly become more informative in recent years. Rather, it is because the unexplained variation in industrial production has declined; output has become less variable.

IV. DO FINANCIAL VARIABLES REALLY HELP PREDICT OUTPUT?

The previous sections examine the in-sample and out-of-sample explanatory power of several financial market variables commonly used to predict economic activity. We compare the explanatory power of these competing variables using empirical specifications typical of the literature and generally duplicate the findings of other researchers. However, we also demonstrate sample sensitivities and differences in in-sample and out-of-sample measures of fit that raise questions about the interpretation of these findings. In this section, we argue that the focus on "horse races" of the sort illustrated in the sections II and III has distracted most researchers from the more fundamental question of whether *any* of these variables contain incremental information useful in forecasting economic activity out-of-sample. We answer this question in the negative and, in the paper's concluding section, discuss how the methodological shortcomings reviewed in the previous sections may have contributed to leading us (collectively) astray.

Figures 8 through 10 compare the forecast errors of the three-variable models of the previous section to the forecast errors of a simple autoregressive model of industrial production. The first column of each figure plots the RMSEs produced by a three-variable model that includes either SP, FF, or M2 against the RMSEs produced by the autoregressive model. As before, forecast horizons of three, six, and nine months are reported in each figure. The results are striking. Casual inspection suggests that *none* of the financial variables is systematically useful in forecasting industrial production. Over most sample periods, the RMSEs of models that include a financial variable are difficult to distinguish from the RMSEs of the autoregressive model. In cases in which a discrepancy is apparent, it is just as likely that the simpler autoregressive model outperforms the three-variable model as the reverse. That is, including a financial variable in the model reduces forecasting power as often as not. The most systematic and prominent difference in performance occurs in the period immediately following 1974:12. All three financial variables appear to be useful in forecasting industrial production in 1974:12, and the small forecast errors that result influence our measure of forecasting power for the subsequent three years.

The visual impressions noted above can be confirmed by bootstrapping confidence intervals around the differences between the RMSEs of each three-variable model and the autoregressive model. The results are shown in the second column of each figure, where the 95% confidence bands are denoted "Lower" and "Upper." Indeed, 1974:12 stands out as the only instance in which financial market variables make a consistent and (sometimes) statistically significant contribution to out-of-sample forecasting power.[9] Furthermore, there are a number of cases in which the inclusion of a financial variable in the forecasting equation significantly *reduces* forecasting power. As shown in Figure 10, for example, the federal funds rate significantly reduces forecasting power at all horizons for sample periods ending in the early 1980s. Recent experience provides another example. In Figure 8 the autoregressive model outperforms the M2 model at the three-month and six-month forecast horizons over sample periods that include the first half of the 1990s. For the most part, however, differences between the forecasting power of the au-

9. The three-variable model includes lagged values of the dependent variable (the growth rate of industrial production) and the CPI inflation rate as well as a financial variable. We have also estimated two-variable models that include only lagged values of the dependent variable and the CPI inflation rate. The forecasting power of the two-variable model is virtually identical to the forecasting power of the autoregressive model. Thus we are able to attribute any differences between the autoregressive model and the three-variable model to the presence of the financial variable in the latter.

FIGURE 8
Comparing the RMSE's of the Three-variable M2 Model
and the Autoregressive Model

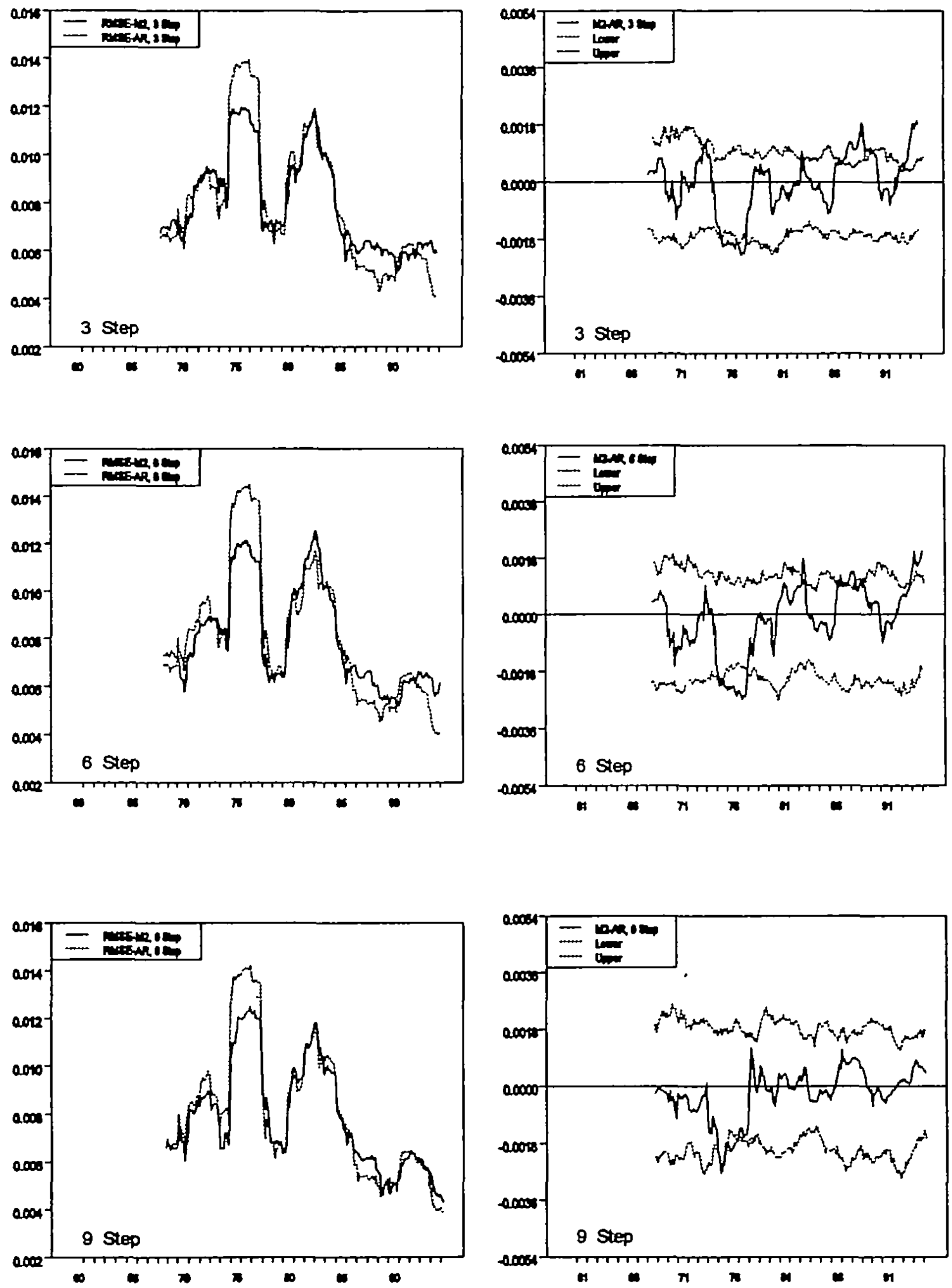


FIGURE 9
Comparing the RMSE's of the Three-variable CP Model
and the Autoregressive Model

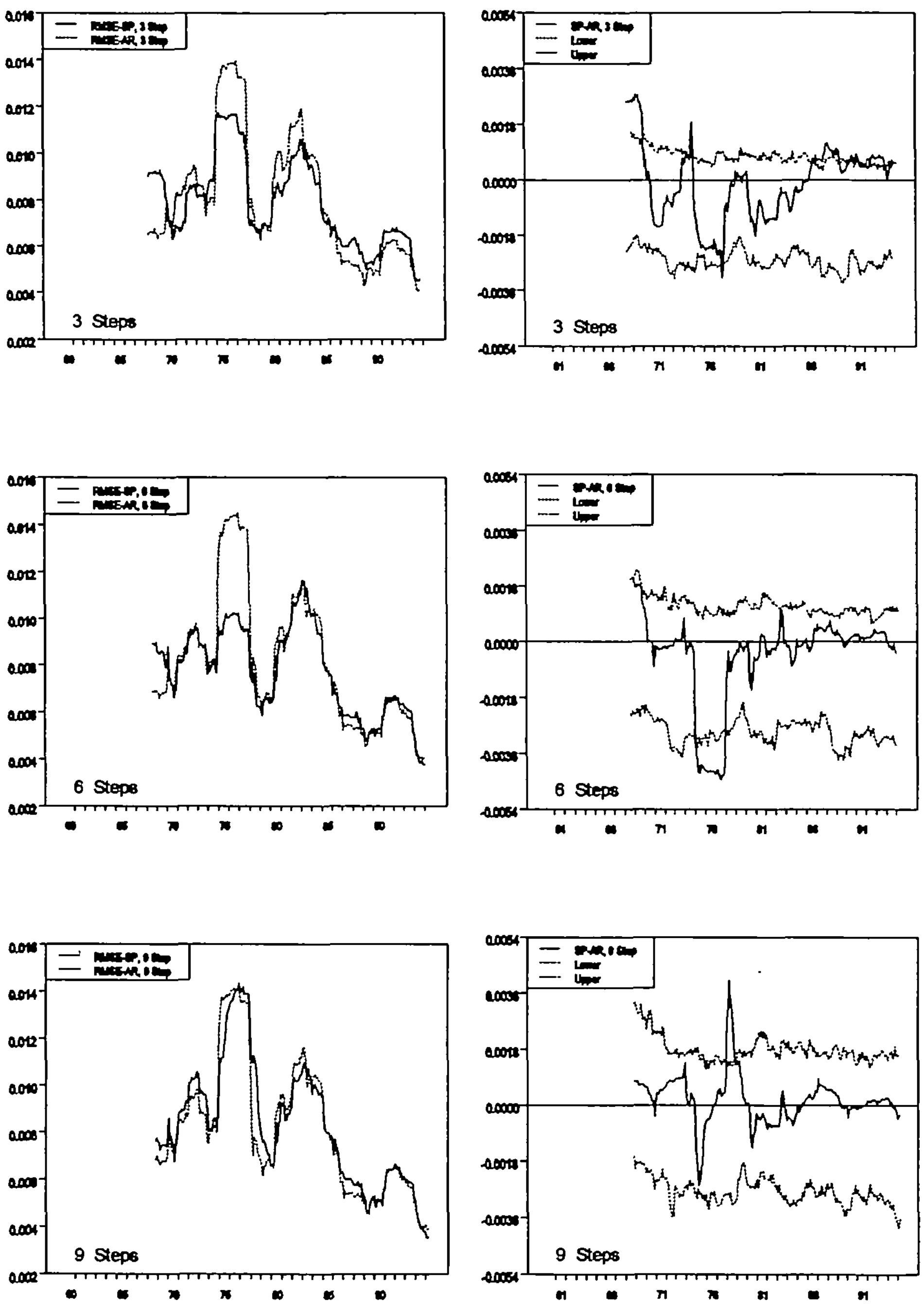
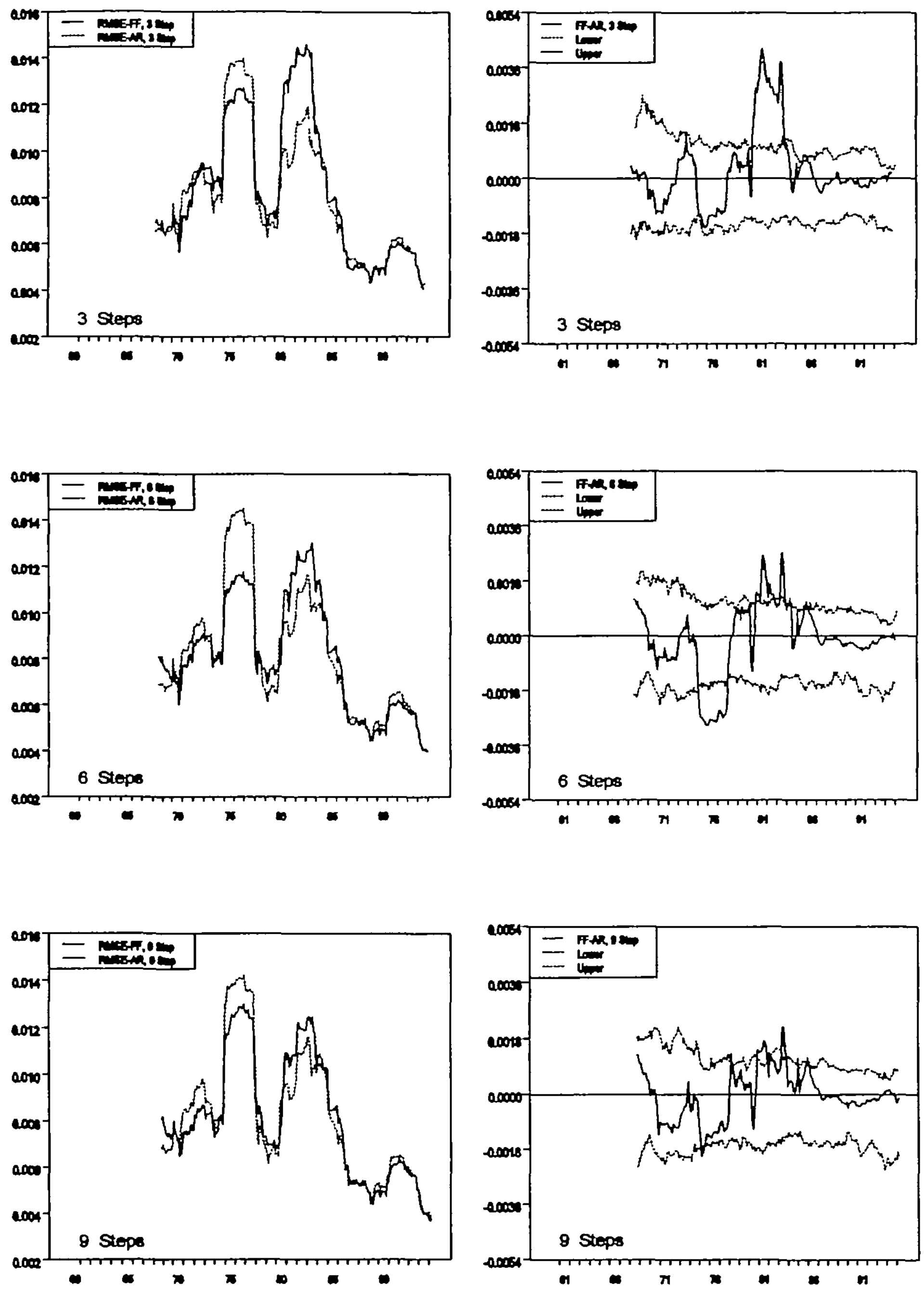


FIGURE 10
Comparing the RMSE's of the Three-variable FF Model
and the Autoregressive Model



toregressive model and models containing a financial variable are insignificant, a finding that carries over to more inclusive empirical specifications.[10] We conclude that none of the financial variables we consider is generally useful in predicting output growth.

Finally, we note that the RMSEs of the autoregressive model drop to their lowest values as the sample is extended through the first half of the 1990s, and that the three-variable models containing the spread and the federal funds rate offer no improvement over the forecasting power of the autoregressive model for these sample periods. This is the basis for the claim made at the end of section III concerning the recently low forecast errors of models that include SP and FF. Figures 9 and 10 show that the low RMSES of the interest rate models in very recent years are due to a decline in the unexplained variation in output rather than an increase in the information content of SP and FF. Indeed, the figures indicate that the information content of SP and FF is negligible throughout this period. The M2 model reported in Figure 8, on the other hand, actually falls short of the autoregressive model in predicting output growth in very recent years, significantly so at the three-month and six-month forecast horizons. One factor that may explain this is the unexpected increase in M2 velocity that occurred in the early 1990s, an increase that some researchers have blamed on financial innovations that caused unpredicted flows out of M2 and into bond and stock mutual funds. (See Orphanides, Reid, and Small [1994]). Our forecasts of industrial production growth are dynamic forecasts that require forecasts of the future values of the explanatory variables that enter the prediction equation. Thus, as M2 becomes less predictable, the quality of the dynamic forecasts of the M2 model decline. This negative effect on the M2 model's forecasting performance appears to have more than offset the positive effect of less variable industrial production growth rates.

10. We have explored a variety of other empirical models, including specifications in which all three of the financial variables featured in this paper to enter simultaneously, and specifications (similar to those employed in Bernanke and Blinder [1992]) that further augment the list of financial variables to include M1, the three-month Treasury bill rate, and the ten-year Treasury bond rate in various combinations. The exercises produce results very similar to those already reported in the paper. None of the models we have examined systematically outperforms the autoregressive model in out-of-sample forecasting exercises.

V. CONCLUSIONS

We have reviewed and illustrated elements of the recent debate over interest rates, interest rate spreads, and monetary aggregates as predictors of economic activity. Consistent with the focus of a number of recent papers, we begin by comparing the in-sample explanatory power of several extensively studied financial market variables and generally confirm the findings of others. In particular, the results of Granger-causality tests suggest that the paper-bill spread is superior to both M2 and the federal funds rate in predicting economic activity. Our empirical approach, however, provides a comprehensive treatment of sample sensitivity that raises questions about the conclusions drawn from these results. For example, the superior performance of the paper-bill spread in explaining industrial production appears to be due primarily to a single observation in late 1974. Because it is generally regarded as the preferred measure of fit, we also compare the out-of-sample forecasting power of the spread, the federal funds rate, and M2, and show that in-sample measures of fit (e.g. Granger-causality tests) can be poor indicators of out-of-sample forecasting power.

The paper's primary contribution lies in its second half, where we argue that an important and unfortunate by-product of the recent focus on "horse races" between competing financial market variables has been to distract researchers from the question of whether *any* of these variables contains incremental information that is systematically useful in forecasting real activity out-of-sample. We answer this questions in the negative, showing that over most sample periods, and certainly recently, a simple autoregressive model of industrial production performs as just well as (indeed, sometimes better than) multi-variate models that include financial market variables believed to be sensitive to monetary policy. We conclude that either monetary policy innovations have no significant real effects, or monetary economists have failed in their efforts to measure monetary policy.

The recursive regression technique employed throughout the paper is, we believe, a powerful diagnostic. As we demonstrate, the technique is well-suited to identifying "atypical" observations and changes in regime that may dramatically affect test statistics and the outcomes of other econometric exercises (e.g. forecasting).[11] This approach to identifying influential observations is likely to be particularly advantageous in times series applications in which the multicollinearity generated by distributed lags makes alternative diagnostics such as tests of parameter instability less attractive. In the applications considered here, 1974:12 stands out as an influential observation. It has dramatic effects on both in-sample and out-of-sample measures of fit. Furthermore, we find that the only instance in which financial market variables have been statistically significant predictors of industrial production growth rates is in 1974:12. Most recent empirical studies of the explanatory power of competing financial market variables perform estimation over sample periods that include 1974. Indeed, samples are often split so that 1974 is included in both halves. (E.g., Friedman and Kuttner [1992] report results for the samples 1960–1979, 1970–1990, and 1960–1990.) Given its striking effects on both in-sample and out-of-sample measure of fit, it seems possible that the presence of this outlier in the data has contributed to the view that financial market variables contain information useful in predicting economic activity.

In what way was late 1974 "atypical" of the processes that jointly determine real activity and financial market variables? We offer the following observations, which point to uncertainty about the state of the international banking system as an important element of the answer to this question.[12] The unprecedented decline in post-war industrial production in December of 1974 was preceded in July of 1974 by unprecedented highs in interest rates and spreads, although not by exceptionally low money growth rates. It is unlikely a coincidence that the mid-1974 peak in interest rates was closely associated with public recognition of the plight of Franklin National Bank, which between May and October of 1974 "became part of the first, and since then the only, crisis of the new international banking system."[13]

The prospective failure of one of the largest banks in the United States created fears of "a banking crisis with inevitable losses to depositors and creditors, disruption of fund transfers and of credit commitments, as well as consequent loss of jobs and income."[14] In the end the failure of Franklin National did not cause an international banking crisis. But the uncertainty that preceded its failure may well have played a role in the unprecedented levels of the federal funds rate and the paper-bill spread that occurred in mid-1974, thereby producing an atypically high negative correlation between these financial variables and output.[15]

11. Influence statistics also offer a means of identifying observations that have a disproportionate influence on regression outcomes. A variety of influence statistics are used as diagnostic tools in cross-sectional econometric studies, but these diagnostics are not widely applied in time series studies. One case that has been studied is the use of the Kalman filter to assess the influence of individual observations on time series regression coefficients (see RATS 4.0, example 13.2). This statistic identified 1974:12 as the most influential observation in the data set for a regression that included y, p, M2, FF, and SP.

12. As a referee pointed out, we are suggesting that identification is at the heart of the difficulties that 1974:12 poses for the use of interest rates and spreads as measures of the stance of monetary policy. The identification problem arises from misattributing banking sector or other privates sector shocks to monetary policy. The 1974:12 observation is an example of this. Thus, one interpretation of our results is that VAR model disturbances typically used to represent policy shocks are a mixture of disturbances arising from policy and from the banking sector or the private sector. This suggests moving away from the normality assumption for the errors in these models. In this respect, our work complements that of Sims [1986], Strongin [1992], Leeper and Gordon [1992], Gordon and Leeper [1994], and Bernanke and Mihov [1995], among others.

13. Spero [1980, 11].

14. See Spero [1980, 114]. Franklin National Bank's failure was triggered by speculation in foreign exchange markets. In May of 1974, knowledge of Franklin's foreign exchange losses became widespread, as evidenced by the precipitous decline in the price of its stock between April 30 and May 10. Only a few weeks later, the German Bankhaus I. D. Herstatt failed as a result of foreign exchange losses. The failure was, in the judgement of many, poorly managed by the German Central Bank and "disrupted the world's foreign exchange markets and actually led to their paralysis for a number of days." (Spero [1980, 111]). The disruption created by the Herstatt failure paled in comparison to the potential damage from Franklin, whose size and foreign exchange exposure dwarfed that of Herstatt. Thus Franklin became a precedent-setting test-case for the new international banking system that emerged after the move to floating exchange rates in the early 1970s.

15. The explanation offered here is a variant of the "default risk" explanation for the predictive power of the spread discussed elsewhere in the literature. In its usual form, this explanation is rejected (correctly, we believe) by Bernanke [1990] on the grounds that defaults on prime nonfinancial commercial paper are extremely rare. Our (implicit) argument is that only an event as momentous as a threat to the international banking system could produce a significant default premium on prime commercial paper, and that 1974:12 is an example—possibly the only post-1960 example—of such a threat.

REFERENCES

Bernanke, Ben S. "On the Predictive Power of Interest Rates and Interest Rate Spreads." *New England Economic Review*, November/December 1990, 51–68.

Bernanke, Ben S. and Alan S. Blinder. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review*, 82(4), September 1992, 901–21.

Bernanke, Ben. S. and Ilian Mihov. "Measuring Monetary Policy." Federal Reserve Bank of San Francisco Working Paper 95-09, March 1995.

Christiano, Lawrence J. and Lars Ljungqvist. "Money Does Granger-Cause Output in the Bivariate Money-Output Relation." *Journal of Monetary Economics*, 22, 1988, 217–35.

Emery, Kenneth M. "The Information Content of the Paper-Bill Spread." *Journal of Economics and Business*, 1966, 1–10.

Friedman, Benjamin M. and Kenneth N. Kuttner. "Money, Income, Prices, and Interest Rates." *American Economic Review*, 82(3), June 1992, 472–92.

Gordon, David B. and Eric M. Leeper. "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification." *Journal of Political Economy*, 102, December 1994, 1,228–47.

Hess, Gregory D. and Richard D. Porter. "Comparing Interest-Rate Spreads and Money Growth as Predictors of Output Growth: Granger Causality in the Sense Granger Intended." *Journal of Economics and Business*, 45, 1993, 247–68.

Leeper, Eric M. and David B. Gordon. "In Search of the Liquidity Effect." *Journal of Monetary Economics*, June 1992, 341–69.

Litterman, Robert B. and Lawrence Weiss. "Money, Real Interest Rates, and Output: A Reinterpretation of Postwar U.S. Data." *Econometrica*, January 1985, 129–56.

Orphanides, Athanasios, Brian Reid and David H. Small. "The Empirical Properties of a Monetary Aggregate that Adds Bond and Stock Funds to M2." *Review*, Federal Reserve Bank of St. Louis, 76(6), November/December 1994, 31–51.

Rudebusch, Glenn D., "Do Measures of Monetary Policy in a VAR Make Sense?" manuscript, Federal Reserve Bank of San Francisco, January 1996.

Sims, Christopher A. "Are Forecasting Models Usable for Policy Analysis?" *Quarterly Review*. Federal Reserve Bank of Minneapolis, Winter 1986, 2–16.

Spero, Joan E. *The Failure of the Franklin National Bank: Challenge to the International Banking System*. New York, New York: Columbia University Press, 1980.

Stock, James and Mark W. Watson. "New Indexes of Coincident and Leading Indicators," in *NBER Macroeconomics Annual*, edited by O. Blanchard and S. Fischer. Cambridge, Mass.: MIT Press, 1989, 351–93.

______. "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience," in *New Research on Business Cycles, Indicators, and Forecasting*, edited by J. H. Stock and M. W. Watson, University of Chicago Press, Chicago, 1993, 95–153.

Strongin, Steven. "The Identification of Monetary Policy Disturbances: Explaining the Liquidity Puzzle." Working paper WP-92-27, Federal Reserve Bank of Chicago, November 1992.

Thoma, Mark A. and Jo Anna Gray. "Aggregates Versus Interest Rates: A Reassessment of Methodology and Results." Manuscript, University of Oregon, August 1995.