William M. Lupoletti
University of Chicago

Roy H. Webb
Federal Reserve Bank of Richmond

Defining and Improving the Accuracy of Macroeconomic Forecasts: Contributions from a VAR Model*

This paper is an empirical study of unconditional forecasting performance. One object of the paper is to review the performance of several well-known forecasters. Since there is little agreement on the definition of a successful forecast, we introduce a vector autoregressive (VAR) model as a benchmark for predictive accuracy. In addition, the VAR model is able to capture relevant information that is omitted from other forecasts. We can therefore demonstrate the value of VAR forecasts as an input to composite forecasts.

Thirty years ago it appeared that the best strategy for improving economic forecasts was to build bigger, more detailed models. As the cost of computing plummeted, considerable detail was added to models, and more elaborate statistical techniques became feasible. Yet dissatisfaction with conventional macroeconometrics has grown steadily in recent years.[1] One outgrowth of this dissatisfaction has been increasing interest in atheoretical forecasting techniques, which, in a sense, represent a return to "measurement without theory." Proponents might prefer the label "measurement without pretense," however, since they do not accept the idea that conventional models actually embody theory that is consistent with the behavior of optimizing individuals in a stochastic, dynamic environment. Instead, they believe that conventional models contain ad hoc representations that are manipulated until they become consistent with historic time series.

* The authors are indebted to Bennett McCallum and this *Journal's* referee, Victor Zarnowitz, for helpful comments and to Roger Brinner, Donald Graling, Robert Litterman, Charles Nelson, and Robert Wescott for providing series of forecasts. Of course, any views and opinions expressed in this paper should be attributed to the authors and not to any of the individuals named above or to the Federal Reserve Bank of Richmond. This paper was written while Lupoletti was employed at the Federal Reserve Bank of Richmond.

1. See Sims (1980) for a lucid critique of the conventional strategy for constructing macro models.

(*Journal of Business*, 1986, vol. 59, no. 2, pt. 1)

0021-9398/86/5902-0004$01.50

Whereas much of the debate has concentrated on theoretical benefits and drawbacks to various approaches, this paper is an empirical study of unconditional forecasting performance. Although other products can be obtained from an econometric model, it is hard to imagine a model producing inaccurate unconditional forecasts but also producing accurate conditional forecasts and hypothesis tests. Thus one object of this paper is to review the performance of several well-known forecasters. It is apparent, however, that there is little agreement over what constitutes a successful forecasting record. Therefore we introduce a vector autoregressive (VAR) model as a benchmark for predictive accuracy. In addition to its value as a benchmark, the VAR model is able to capture relevant information that is omitted from other forecasts. Consequently, it is valuable as an input to composite forecasts.

The paper is organized as follows. First, the model and data are described. Next, the model's unconditional forecasting accuracy is compared to the records of well-known producers of macroeconomic forecasts. The VAR model's positive contribution to composite forecasts constitutes the final topic.

Model and Data Description

The Model

The VAR model presented in this section provides a striking contrast to conventional structural macroeconomic models. First, there are no exogenous variables—each variable is predicted by its own lagged values and by lagged values of other variables. Second, the only use of economic theory was in choosing the variables to be included.[2] Accordingly, the parameters of this model are given no structural interpretation. Thus there is no guarantee that exercises such as conditional forecasting will produce meaningful results. There is also the danger that a change in the structure of the economy could invalidate the pattern of correlations implicit in the model's estimated parameters. Of course, the concern that a model may be vulnerable to structural change is not unique to VAR models since ad hoc dynamics and inade-

2. Money demand equations have traditionally included a nominal monetary aggregate divided by a price index, a measure of either real income or output, and a short-term interest rate. We chose four of our five variables on the basis of that theoretical and empirical literature. Our only departure from that traditional list of variables was to add the capacity utilization rate in manufacturing industries. The purpose was to include a variable that would represent the stage of the business cycle, following preliminary empirical work that indicated a cyclical influence on forecast errors. We considered other variables, such as the unemployment rate or unfilled orders, but rejected them a priori because of their lack of stationarity. Since the model with capacity utilization produced better forecasts than the four-variable model did without it, we were not inclined to search for alternative cyclical indicators.

quately modeled expectation formation may invalidate a conventional macro model's conditional forecasts.

Given its simplicity and lack of theory, it may seem unlikely that a small VAR model could produce accurate forecasts. However, as Litterman (1984) has suggested, it may be helpful to consider any macroeconomic forecasting model as a filter that attempts to extract information from noisy data. The design of a model involves a trade-off between oversimplification and overparameterization: omitting a relevant variable closes a channel that might contain more signal, while including an irrelevant variable adds more noise. Since it is impossible to determine the ex ante marginal signal-to-noise ratio with respect to any particular variable, we find it difficult to choose a forecasting model on a priori grounds. Instead, we believe that the specification of a forecasting model should be based on empirical evidence.

What kinds of empirical evidence should be considered? Economists have often noted the usefulness of studying both a model's in-sample and its postsample performance. In-sample testing allows use of the familiar tools of statistical inference; however, goodness of fit can easily be improved without improving forecasting ability. Although postsample results are a better measure of a model's predictive power, we are not aware of any formal tests of the significance of differences between series of forecast errors. At the risk of producing results of uncertain significance, we confine ourselves to the study of postsample performance.

The model presented in this paper consists of five variables: the monetary base, real GNP, the GNP implicit price deflator, the manufacturing capacity utilization rate, and the 90-day Treasury bill rate (the first three are expressed as percentage changes). The model is identical to the model in Webb (1984*b*) except for this paper's use of the Treasury bill rate in place of the commercial paper rate.

The model can be expressed as

$$\mathbf{X} = \mathbf{C} + B(L)\mathbf{X} + \mathbf{E}, \tag{1}$$

where $\mathbf{X}$ is the vector of endogenous variables; $\mathbf{C}$ is a vector of constant terms; B is a polynomial in the lag operator, L, in this case representing lags 1–6; and $\mathbf{E}$ is a vector of error terms. The model thus consists of five regression equations, one for each variable. The right side of each equation contains exactly the same terms: a constant, six lagged values of each variable, and the error term. Because the right sides of each equation are identical, there is no efficiency gain in moving from ordinary least squares estimation to simultaneous equation methods.

Since each equation in the model is treated independently, it is appropriate, when examining potential overparameterization, to focus on 31 estimated parameters per equation rather than on the model's 155 parameters. The sample size, which ranges from 74 observations (1952:2–1970:3) to 126 (1952:2–1983:4), is in fact small enough to raise the concern of too many parameters being estimated, leading to less accurate coefficient estimates and, consequently, less accurate forecasts. Therefore it is of interest to note (*a*) whether the model produces accurate forecasts and (*b*) whether the forecasting accuracy tends to increase over time (which could indicate that the initial sample size was too small for the number of parameters estimated).

The Data

To derive a series of postsample forecasts from the model, the equations described by (1) were estimated from 1952:2 to 1969:4, and the estimated coefficients were used to forecast each variable in 1970:1. Those values, in turn, were used to produce a forecast of 1970:2. The procedure was repeated until we had a set of forecasts from 1970:1 through 1971:2, all based on data comparable to that possessed by a real-time forecaster in 1969:4. We then constructed a similar set of forecasts for each quarter through 1983:4. We thus produced a series of forecasts with horizons ranging from 1 to 6 quarters past the latest data that would have been available to a forecaster at the time of the forecast.

In the following section the VAR model's forecasts are first compared with published forecasts from the best-known commercial forecasting services in the United States—Chase Econometrics, Data Resources, and Wharton Econometric Forecasting Associates. We were able to obtain their forecasts, beginning with those published in 1970:3, for real GNP, the GNP deflator, and the Treasury bill rate. Thus we have 1-quarter-ahead forecasts and realized values beginning in 1970:4, 2-quarter forecasts (expressed as average compounded annual rates) beginning in 1971:1, and so forth, up to 6-quarter forecasts beginning in 1972:1. The only exception is Wharton's series of interest rate forecasts, which begins in 1973:2. All comparisons end in 1983:4. In addition, we compare the VAR model's forecasts with two sets of published time series forecasts: Charles Nelson's (1984) ARIMA forecasts of GNP and the deflator beginning in 1976:2 and Robert Litterman's (1985) forecasts from a VAR model beginning in 1980:3.

In most cases the published forecasts were released at the end of a particular quarter.[3] Accordingly, the forecasters would have possessed

3. Such forecasts, usually released within the last 2 weeks of the quarter, are often referred to as late-quarter forecasts. The commercial forecast data in this study are mostly from late-quarter forecasts. The only exceptions, due to data availability, are the pre-1976 Wharton forecasts, which are from mid-quarter. As McNees (1975) has noted, Wharton would thus be at a disadvantage prior to 1976 since they had access to less data for the quarter in which each forecast was prepared than did the other forecasters. It should be noted that our nomenclature differs from that followed by other writers, notably, McNees. He labels a forecast made at time *t* for an outcome also at time *t* as a 1-

most of the information needed to estimate quarterly average values of the current quarter's interest rate and monetary base, typically, 11 of 13 weekly values in a particular quarter. In addition, they would have had two of three monthly values for the capacity utilization rate. But for most of the period there were no official releases of GNP and the deflator until early in the next quarter. Moreover, these data series are subject to routine revision during the year, annual revisions, and possible further change as a result of benchmark revision after several years. The monetary base and, to a lesser extent, the capacity utilization rate are also subject to later revision.

Since we employed the latest revisions of data when estimating the model and producing postsample forecasts, we were concerned that our procedures might exaggerate the relative accuracy of the VAR model. To test the sensitivity of our results to data accuracy, we designed the following experiment. For the first lagged value in each of our regression equations we replaced the latest revision with the preliminary data release for real GNP and the deflator. We also used originally released data for the monetary base and the capacity utilization rate. These changes all tended to reduce the VAR model's informational advantage over real-time forecasters. We then used the preliminary data to estimate the VAR model's coefficients and to produce postsample forecasts. We expected that this procedure would yield uniformly less accurate forecasts than would the identical procedure with the most recent values for those dates, but we were interested in knowing how much accuracy would be lost.

Surprisingly, in two cases the model produced substantially more accurate 4-quarter-ahead forecasts based on the preliminary data. Table 1 shows representative forecasts from the basic model and several variations, including the forecasts that were based on preliminary data. For the Treasury bill rate 4 quarters ahead, the average forecast error fell from 2.8% to 1.9%, and the error for the deflator fell from 2.1% to 1.4%.[4] On the other hand, the corresponding real GNP forecast

quarter-ahead forecast. On the other hand, we define a 1-quarter-ahead forecast as one made at time t for an outcome at time $t + 1$. An example of the reason for changing McNees's convention is that it may be confusing, in his terminology, that a late-quarter forecast of a quarter's average interest rate could be produced after 11 weeks of data were observed. Our practice is analogous to that of Zarnowitz (1979) with respect to annual forecasts.

4. In attempting to explain these unexpected results, we noted that, in regressions from 1952:2 to 1983:4, F-tests showed that the six lagged inflation terms were significant at the 98% and 99% levels in the interest rate and inflation equations, respectively. However, lagged inflation terms were not significant at conventional levels in any other equations. Therefore the preliminary release of the deflator became a prime suspect. A visual inspection of that series indicated a tendency of the preliminary release of the 1-quarter inflation rate to understate the actual value, especially in the early 1970s. A more formal test, from 1970:1 to 1983:4, also indicated that the preliminary data series was a biased predictor of the final series. In a regression of the final value on the preliminary

error rose from 2.8% to 3.0%. In the other comparisons there was little or no difference in forecasting accuracy. All in all, this experiment does not reveal any gross bias favoring the VAR model because of its access to the latest data revisions.

Another concern when comparing real-time and simulated forecasts is the opportunity provided to manipulate a model's specification until an unrealistically accurate fit is obtained. Three important changes have been made since the model was conceived that did improve forecast accuracy. The monetary base was substituted for M1 because of our suspicion that financial deregulation distorted the demand for M1 in 1981–82. Also, the capacity utilization rate was added to a four-variable model that produced forecast errors that seemed to follow the business cycle. Finally, the starting date 1952:2 was chosen because of its use in traditional money demand studies, such as Goldfeld (1976).

Further changes in the model's specification appear to produce minimal benefits, however. Table 1 contains several comparisons that illustrate the potential returns to small changes in the model's specification. For example, substituting the commercial paper rate for the Treasury bill rate has mixed results of a fairly small magnitude. Also, several variables were tried as a sixth variable. One of the most successful (in-sample) was the producer price index. For postsample forecasts, however, it generally decreased forecast accuracy. Increasing the lag length to 8 quarters and extending the sample period both resulted in less accurate forecasts. In short, the returns to model respecification are neither large nor obvious.

Accuracy Comparisons

This section summarizes tables 2–8, which describe the accuracy of the six models over various time periods. Accuracy was measured in each case by the root-mean-square value of the forecast error, which, in turn, is the difference between actual and predicted values.

Comparison with Forecasting Services

Table 2 contains the results for the entire period 1970–83. For all variables except the interest rate, the VAR model's worst relative and absolute performance occurs 1 quarter ahead. At 4- and 6-quarter horizons, however, the VAR forecast lies between the best and the worst

value, the joint hypothesis that the constant term was equal to zero and that the slope coefficient was equal to one was rejected at the 5% level. It appears that the bias in the preliminary release toward understating change offset an opposite tendency toward overpredicting change in the VAR forecasts of inflation and interest rates. Accordingly, much of the gain in forecasting accuracy seems to result from a fortuitous coincidence. We see no reason to believe that that coincidence will persist and, therefore, would not expect the use of preliminary data to improve the accuracy of forecasts now being made.

TABLE 1 **Variations in Model Specification**

	Real GNP		GNP Deflator		Nominal GNP		Interest Rate	
	1-Quarter Forecast	4-Quarter Forecast	1-Quarter Forecast	4-Quarter Forecast	1-Quarter Forecast	4-Quarter Forecast	1-Quarter Forecast	4-Quarter Forecast
Reported model	5.2	2.8	2.3	2.1	5.9	3.3	1.2	2.8
Variation 1:								
Preliminary data	5.2	3.0	2.2	1.4	6.0	3.3	1.2	1.9
Variation 2:								
Commercial paper rate	5.3	2.8	2.4	2.0	6.1	3.2	1.5	3.0
Variation 3:								
Sixth variable (PPI)	5.1	3.1	2.2	2.0	6.1	3.5	1.6	3.2
Variation 4:								
Longer lags (8 quarters)	6.0	3.2	2.6	2.3	6.9	3.5	1.7	3.4
Variation 5:								
Earlier start (1948:4)	4.9	2.8	2.8	2.5	6.2	4.0	1.5	2.9

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts, measured in the same units as the variables themselves. One-quarter RMSEs were calculated over the range 1970:1–1983:4, and 4-quarter RMSEs were calculated over the range 1970:4–1983:4. All variables except the interest rate were expressed as percent changes at annual rates.

TABLE 2 Average Forecast Errors from All Models, 1970–83

Variable: Forecast Horizon (Quarters)	Chase	DRI	Wharton	VAR
Real GNP:				
1	4.1	4.0	4.2	5.3
2	3.1	3.1	2.9	4.1
4	2.5	2.5	2.2	2.8
6	2.3	2.3	1.9	2.4
GNP deflator:				
1	1.8	2.0	1.9	2.3
2	1.9	2.0	1.9	2.0
4	2.2	2.1	2.0	2.1
6	2.5	2.4	2.2	2.4
Nominal GNP:				
1	5.1	4.6	4.9	6.0
2	4.1	3.6	3.8	4.7
4	3.5	3.0	3.0	3.3
6	3.3	2.7	2.6	3.1
Treasury bill rate:				
1	1.5	1.4	. . .	1.3
2	2.2	2.1	. . .	2.1
4	2.9	2.6	. . .	2.8
6	3.5	3.2	. . .	3.5

Note.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are 1-quarter forecasts: 1970:4–1983:4; 2-quarter forecasts: 1971:1–1983:4; 4-quarter forecasts: 1971:3–1983:4; and 6-quarter forecasts: 1972:1–1983:4.

commercial forecast for all variables except real GNP. The results thus contradict the opinion expressed by Klein (1984, p. 91): "None of these time series methods [VAR and ARIMA models] stands up to the use of macroeconometric models with constant adjustments. . . . [T]ime series models do perform about as well as the adjusted macroeconometric model in forecasting very short time horizons, say up to three months or possibly up to six months."

Since Klein's view appears to be shared by many economists, we would like to suggest some explanations for the VAR model's improved relative forecasting ability at longer horizons. First, it should be noted that a VAR model is significantly different from an ARIMA model because of the VAR model's multivariate interaction. For example, allowing growth in the monetary base to enter the inflation equation should allow the inflation equation to adjust to changes in monetary policy, assuming that growth in the base can be predicted with reasonable accuracy.

Second, much of the forecasting services' advantage in 1-quarter forecasts seems to derive from the variety of ad hoc factors that affect quarterly data and the forecasters' good judgment in accounting for those factors on a near-term basis. For example, strikes and federal transfers of agricultural commodities have temporarily distorted growth rates of real GNP and the deflator in the past. Also, knowing the pattern of growth within the previous quarter can improve the prediction of the next quarterly average value. And in certain components of GNP subject to substantial measurement error, such as inventories and imports, knowledge of an extreme value in one quarter can lead to the expectation that some measurement error is likely to be reversed in the next quarter. These ad hoc factors can greatly aid the forecast of the next quarter but are of little value at longer horizons.

Finally, some of the improvement in the VAR model's longer term performance is due to our use of average compound growth rates for multiquarter forecasts of GNP and the deflator. That procedure tends to smooth both the underlying series and the VAR forecasts by allowing either anomalies in the data or forecast errors of opposite signs in consecutive quarters to cancel out. Thus, in comparing 1- and 4-quarter forecasts of real GNP, nominal GNP, and the deflator, the forecast error is reduced both by the lower variance of the actual series and by the canceling of errors in forecasts of adjacent quarters. The latter effect is particularly evident in much lower values of the Theil *U*-statistic for each series at the 4-quarter horizon.[5]

Tables 3–6 contain summary statistics for each forecaster in several subperiods. In the earliest period the VAR model had its worst relative and absolute performance for real and nominal GNP. Both series were highly volatile in that period and reflected many ad hoc factors. Perhaps most dramatic were the imposition and removal of comprehensive wage-price controls and the rise of energy prices. Those two extraordinary events may have worsened the mechanical VAR model's relative performance since they could have been taken into account by the forecasting services' judgmental adjustments. On the other hand, the relatively poor performance may result from the VAR model's profligate parameterization (31 estimated coefficients per equation) and may indicate that 20 years simply did not provide enough observations to estimate that many coefficients accurately.

The dates for the middle period were chosen to cover a complete business cycle, trough to trough. The VAR model's predictions for real and nominal GNP were more accurate than in the earlier period, although they were still substantially worse than predictions from the most accurate forecasting service. Also, the VAR model's inflation forecasts were less accurate in the middle period, whereas the three forecasting services improved the accuracy of their inflation forecasts. In contrast, the interest rate forecasts from the VAR model were more accurate than those of the three forecasting services.

The VAR model exhibited its best relative performance in the 1980–

5. For real GNP, the deflator, and nominal GNP, the Theil *U*-values for 1-quarter forecasts are 0.95, 1.12, and 0.95, respectively. The corresponding *U*-values for 4-quarter forecasts are 0.64, 0.95, and 0.84.

TABLE 3 Average Forecast Errors from the VAR Model

Variable: Forecast Horizon (Quarters)	Full Period	Start–1975:1	1975:2–1980:3	1980:4–1983:4	Around Turning Points
Real GNP:					
1	5.3	6.1	4.9	4.8	5.8
2	4.1	4.6	3.7	4.1	4.3
4	2.8	3.7	2.3	2.4	2.8
6	2.4	2.9	2.1	2.2	2.6
GNP deflator:					
1	2.3	1.8	2.3	2.9	2.6
2	2.0	1.7	2.0	2.3	2.3
4	2.1	2.1	2.3	2.0	2.4
6	2.4	2.5	2.5	2.2	2.6
Nominal GNP:					
1	6.0	6.6	5.6	5.9	6.4
2	4.7	5.2	4.3	4.8	4.7
4	3.3	4.1	2.9	3.1	2.7
6	3.1	3.8	2.4	3.4	2.7
Treasury bill rate:					
1	1.3	1.0	1.0	1.9	1.6
2	2.1	1.7	1.6	3.1	2.6
4	2.8	2.5	1.9	4.2	3.2
6	3.5	3.3	2.5	5.0	3.6

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are as in table 2.

TABLE 4 Average Forecast Errors from Chase Econometrics

Variable: Forecast Horizon (Quarters)	Full Period	Start–1975:1	1975:2–1980:3	1980:4–1983:4	Around Turning Points
Real GNP:					
1	4.1	2.8	4.3	5.0	4.6
2	3.1	2.2	2.8	4.4	3.5
4	2.5	2.4	2.1	3.3	3.0
6	2.3	2.3	2.1	2.4	2.7
GNP deflator:					
1	1.8	2.2	1.5	1.8	2.0
2	1.9	2.3	1.5	1.8	2.0
4	2.2	3.0	1.5	2.1	2.4
6	2.5	3.0	2.2	2.4	2.7
Nominal GNP:					
1	5.1	3.1	5.6	6.2	5.5
2	4.1	2.6	4.0	5.6	4.4
4	3.5	1.6	3.3	5.2	3.9
6	3.3	1.8	2.9	4.8	3.6
Treasury bill rate:					
1	1.5	1.0	1.6	1.8	1.9
2	2.2	1.5	1.9	3.1	2.6
4	2.9	1.6	2.8	3.9	3.0
6	3.5	2.1	3.4	4.5	3.2

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are as in table 2.

TABLE 5 **Average Forecast Errors from Data Resources, Inc.**

Variable: Forecast Horizon (Quarters)	Full Period	Start–1975:1	1975:2–1980:3	1980:4–1983:4	Around Turning Points
Real GNP:					
1	4.0	3.5	3.9	4.6	4.5
2	3.1	3.1	2.3	4.2	3.7
4	2.5	2.9	1.8	3.1	3.2
6	2.3	2.5	2.0	2.5	2.8
GNP deflator:					
1	2.0	2.5	1.7	1.9	2.2
2	2.0	2.6	1.5	1.5	2.1
4	2.1	3.0	1.5	1.8	2.3
6	2.4	3.3	1.8	2.2	2.5
Nominal GNP:					
1	4.6	3.2	4.8	5.7	5.1
2	3.6	2.6	3.1	5.1	4.0
4	3.0	2.0	2.4	4.5	3.5
6	2.7	1.5	1.7	4.5	3.4
Treasury bill rate:					
1	1.4	.9	1.4	1.9	1.8
2	2.1	1.4	1.6	3.2	2.5
4	2.6	1.7	2.1	4.0	2.9
6	3.2	2.0	2.6	4.7	3.3

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are as in table 2.

TABLE 6 **Average Forecast Errors from Wharton Econometric Forecasting Associates**

Variable: Forecast Horizon (Quarters)	Full Period	Start–1975:1	1975:2–1980:3	1980:4–1983:4	Around Turning Points
Real GNP:					
1	4.2	4.0	4.5	3.8	4.8
2	2.9	2.6	2.4	3.8	3.4
4	2.2	2.0	1.7	3.0	2.7
6	1.9	2.1	1.4	2.4	2.3
GNP deflator:					
1	1.9	2.0	1.8	1.9	1.8
2	1.9	2.2	1.7	1.8	1.9
4	2.0	2.6	1.5	1.9	2.1
6	2.2	3.0	1.6	2.1	2.3
Nominal GNP:					
1	4.9	4.1	5.2	5.3	5.5
2	3.8	2.8	3.4	5.2	4.3
4	3.0	1.9	2.1	4.8	3.6
6	2.6	2.0	1.2	4.4	3.3
Treasury bill rate:					
1	. . .	. . .	1.8	1.5	. . .
2	. . .	. . .	2.0	2.7	. . .
4	. . .	. . .	2.2	3.6	. . .
6	. . .	. . .	2.7	4.3	. . .

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are as in table 2.

83 period. It had the most accurate forecasts for real GNP 4 and 6 quarters ahead and for nominal GNP 2, 4, and 6 quarters ahead. Its interest rate forecasts, however, worsened considerably from earlier periods and were less accurate than those of the forecasting services.

The final comparison involves forecasts made within 2 quarters of a business cycle turning point. As might be expected, forecasts were generally less accurate at such times, regardless of the source of the forecast. Unexpectedly, the VAR model was more accurate for nominal GNP at 4- and 6-quarter horizons than for the whole period and was also more accurate than the other forecasters. For real GNP, the VAR model showed less deterioration in accuracy than did the forecasting services; the opposite result holds for the inflation rate.

In conclusion, it has been shown that for short forecast horizons—1, and possibly 2, quarters ahead—the VAR forecasts were substantially less accurate than forecasts produced by major forecasting services. At longer horizons, however, the VAR forecasts were competitive with those of the forecasting services. This result is robust in the sense that it holds at different time periods and for forecasts produced near cyclical turning points. The result is of particular interest when the forecasting methodologies are contrasted: the forecasting services each consider several hundred time series, whereas the VAR model considers only five; the forecasting services use models that embody several hundred restrictions supposedly derived from economic theory, whereas the VAR model uses practically no theory; and the forecasting services devote considerable resources toward modifying their forecasts to account for current information that is not a formal part of their models, whereas the VAR forecasts are produced mechanically with no adjustment. On the other hand, the VAR model does have an information advantage, the subperiods studied are fairly small, and the statistical significance of the difference between forecast series is undefined. Taking those caveats into account, along with our experiment on the value of the VAR model's informational advantage, it appears that a simple five-variable VAR model provides a useful benchmark for the evaluation of macroeconomic forecasts, particularly at the longer horizons.

A provocative explanation of why a simple VAR model can provide accurate forecasts is given by Hakkio and Morris (1984). First, they advance the interpretation that a VAR model is a reduced form that serves as a flexible approximation to the reduced form of any member of a wide variety of structural models. Next, they define a true structure and construct consistent data sets. From that constructed data they then estimate a key parameter using a VAR model, a correctly specified structural model, and several misspecified structural models. Not only did the VAR model provide a point estimate that was more accurate than a misspecified model, but the VAR model's estimate was also slightly more accurate than the correctly specified structural model. Considering the likelihood that large structural models are misspecified, it is intuitively appealing to view VAR models as flexible approximations to the reduced form of a correctly specified model of the actual structure of the economy. Under that interpretation, it may not be surprising that VAR models can provide competitive forecasts.

Comparison with Other Time Series Forecasts

Time series methods other than unrestricted VARs have been suggested as alternatives or adjuncts to the conventional macroeconomic forecasting technique. Nelson (1972, 1984) has proposed that ARIMA forecasts are a suitable benchmark for measuring the accuracy of forecasts; he has asserted further that ARIMA forecasts capture information not present in structural forecasts. Litterman (1979, 1984) has developed a strategy for specifying VARs for forecasting using Bayesian prior distributions as constraints on model coefficients. This section presents a comparison of Nelson's and Litterman's forecasts to those generated by the unrestricted VAR model.

Nelson has published ARIMA forecasts of nominal GNP, the GNP price deflator, and real GNP since 1976. He forecasts quarter-to-quarter percentage changes in the three variables for horizons of 1–4 quarters. One-quarter-ahead forecasts start in 1976:2; two- and four-step forecasts begin in 1976:3 and 1977:1, respectively. All forecasts end in 1982:4.[6]

Table 7 presents a comparison of the forecasting performance of the VAR model to that of Nelson's ARIMA technique. The results suggest that the techniques are fairly evenly matched: the ARIMA forecasts produce lower errors in five of the nine cases, while the VAR forecasts are superior four times. Examination of the results over different forecast horizons shows that the ARIMA forecasts beat the VAR forecasts in all three one-step comparisons and in two of three two-step cases, while the VAR model is superior on all 4-quarter-ahead forecasts. This result is consistent with the finding reported above that the VAR model does not produce very good forecasts at short horizons, but it improves relative to other forecasting techniques as the horizon increases. It is also consistent with the belief that the relative forecasting ability of an ARIMA model declines with increasing horizon.[7]

6. Nelson has used preliminary data for quarter t to forecast values of GNP and the deflator in quarter $t + 1$ and beyond. Since the preliminary GNP data are released a few weeks after the end of quarter t, he would thus have had more information than a forecaster would have had at the end of the quarter. Using our terminology, he could be described as a very-late-quarter forecaster of quarter t; in McNees's terms, he is an early-quarter forecaster of quarter $t + 1$.

7. This result also holds for a comparison of the VAR forecasts with forecasts produced by univariate autoregressions.

TABLE 7 Comparison of VAR and Nelson's ARIMA Forecasts

Variable: Forecast Horizon (Quarters)	VAR	Nelson's ARIMA
Real GNP:		
1	4.8	4.5
2	3.8	3.4
4	1.9	2.4
GNP deflator:		
1	2.3	1.8
2	1.9	1.7
4	1.7	1.8
Nominal GNP:		
1	6.0	5.9
2	4.6	4.7
4	3.3	3.8

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are 1-quarter forecasts: 1976:2–1982:4; 2-quarter forecasts: 1976:3–1982:4; and 4-quarter forecasts: 1977:1–1982:4.

These results, though far from conclusive, suggest that, while the ARIMA and VAR techniques have different strengths and weaknesses, they are both useful benchmarks for the analysis of forecasts. The ARIMA model produces better short-term forecasts; however, for 4-quarter-ahead forecasts and probably for longer horizons as well, the VAR technique is superior. Multivariate interaction probably helps keep the VAR model on track at longer horizons while adding noise to 1-quarter forecasts. An ARIMA model uses data more parsimoniously; on the other hand, the production of ARIMA forecasts requires specialized computer software and judgment that can be developed only with practice. Among the advantages of the VAR benchmark are that it requires almost no judgment and that its forecasts can be produced with elementary regression-analysis software.

Litterman has published forecasts from his VAR model every month since May 1980 for six variables: real GNP, the GNP deflator, nonresidential fixed investment, the M1 measure of money, the unemployment rate, and the rate on 90-day Treasury bills.[8] For this comparison, one-, two-, four-, and six-step late-quarter forecasts are used; they begin, respectively, in 1980:3, 1980:4, 1981:2, and 1981:4, and they all end in 1983:4.

It is obviously difficult to draw conclusions on the basis of such a small sample, but table 8 shows that Litterman's VAR model has produced excellent forecasts of real GNP. (Recall that the unconstrained

8. In 1984 Litterman reformulated his model by adding additional variables. Therefore his current forecasts are not strictly comparable to the forecast record examined in this paper.

TABLE 8 Comparison of Unconstrained VAR and Litterman's VAR Forecasts

Variable: Forecast Horizon (Quarters)	Unconstrained VAR	Litterman's VAR
Real GNP:		
1	4.8	4.0
2	4.1	3.2
4	2.5	1.9
6	2.7	1.1
GNP deflator:		
1	2.8	3.4
2	2.3	3.4
4	2.1	3.5
6	2.5	3.8
Nominal GNP:		
1	6.0	6.0
2	4.8	5.6
4	3.2	4.7
6	4.0	3.6
Treasury bill rate:		
1	1.9	1.9
2	3.1	2.9
4	4.6	4.1
6	4.7	4.3

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts. Ranges for RMSEs are 1-quarter forecasts: 1980:3–1983:4; 2-quarter forecasts: 1980:4–1983:4; 4-quarter forecasts: 1981:2–1983:4; and 6-quarter forecasts: 1981:4–1983:4.

VAR model was more accurate than the consulting services during the 1980s.) Unfortunately, Litterman's successful real GNP forecasts are offset by his weak inflation forecasts; as a result, his nominal GNP forecasts are a bit worse than those produced by the unrestricted five-variable VAR 2 and 4 quarters ahead. The differences between the models' forecasts may reflect the influence of omitted variables. For example, Litterman's use of business fixed investment may help explain his success in predicting real GNP, while the unrestricted VAR model's inclusion of the capacity utilization rate probably aided its inflation forecasts.

These results are also interesting because of the claims that are sometimes made in support of imposing Bayesian prior restrictions on VAR models. For example, Todd (1984) has asserted, "Economists rarely have enough data to construct [unrestricted] VAR models, especially ones with more than a few variables, that forecast as well as models that supplement the data with more informative prior beliefs." Since Todd neither presented evidence to support that claim nor pointed to the evidence of others, it is of particular interest to view the results presented in table 8 as bearing on the value of imposing prior beliefs on VAR models. The results presented above are certainly not decisive evidence because of the short time span and the different set of variables included in each model. However, there is surprisingly little published evidence that allows one to examine the differential postsample forecasting ability of shrewdly chosen unrestricted and restricted VAR models. Since our results seem to indicate that the models are fairly evenly matched, the value of complex prior restrictions is not obvious.

Composite Forecasts

Since producers of forecasts are continually attempting to improve their models, the specification of a model is rarely fixed. Despite the extensive adjustments that macroeconomic forecasters have made in the past 30 years, sizable forecast errors persist. An alternative technique for decreasing forecast errors is available to the forecast consumer. From the universe of available forecasts, it may be possible to construct a portfolio forecast that outperforms any of its components. It is conceivable that, of a set of forecasts, any element could include useful information that is not contained in the others. A forecast does not have to be more accurate than its competitors in order for it to have value in a combination; it only has to add information to the combination.

The notion of combining forecasts provides a framework for the consideration of some interesting questions. For example, it has been shown above that a five-variable VAR model can produce forecasts that are competitive with those made by commercial forecasting services. By combining the VAR and a commercial forecast, it can be determined whether the VAR model contains some information that is missing from the large-scale structural models. Similarly, any two forecasting techniques can be compared: two techniques are different from one another to the extent that a combination of the two is better than its components.

Techniques for Combining Forecasts

The simplest way to combine a group of forecasts is to compute the average forecast. For a group of n forecasters, a mean forecast can be produced by assigning each forecast a weight of $1/n$. Alternatively, the median or mode of the group could serve as the summary forecast measure. The American Statistical Association (ASA) and the National Bureau of Economic Research (NBER) issue their joint survey of forecasters in median form. The biggest drawback to using an averaging technique is that it ignores any information about the relative quality of the forecasters. With information about forecasters' past performance, one can compute weights that could decrease the error of the combined forecast if relative forecasting performance remained stable.

To derive weights,[9] suppose P_1 and P_2 are unbiased one-step-ahead predictions of the actual value, A, with uncorrelated forecast errors. Consider the weighted average forecast

$$A^* = wP_1 + (1 - w)P_2 \tag{2}$$

and its associated error

$$e = A - A^* = w(A - P_1) + (1 - w)(A - P_2). \tag{3}$$

The best choice for w can be determined only by reference to the cost of an error, with w chosen so as to minimize cost. A convenient assumption is that cost is proportional to the square of the error. Keeping in mind that the errors have zero means (so variance is identical to mean squared error) and are uncorrelated, it can be shown that the best choice for w is a ratio of the sums of squared errors calculated from time series of A, P_1, and P_2:

$$\hat{w} = \frac{\Sigma(A - P_2)^2}{\Sigma(A - P_1)^2 + \Sigma(A - P_2)^2}. \tag{4}$$

This result can be extended to the case of a combination of more than two forecasts and to the case of multistep forecasts.

It should be noted that it is quite possible that the relative abilities of forecasters will change over time. It is therefore an empirical question whether the square of the combined error will be minimized by calculating w over the entire period for which information is available or whether some subperiod should be used.

The selection of a squared-error cost function suggests the use of least squares as an alternative technique for estimating weights. Estimation by ordinary least squares of

$$A = w_0 + w_1P_1 + w_2P_2 + u \tag{5}$$

allows the assumption of unbiased forecasts and uncorrelated forecast errors to be relaxed. In general, if the forecast errors have nonzero means, $\hat{w}_0$ will not equal zero, and the weights $\hat{w}_1$ and $\hat{w}_2$ will not sum to one. Although OLS estimation thus relaxes two assumptions needed to derive the ratio technique of formula (4), it has its disadvantages as well. Only one parameter is estimated in (4), while (5) requires three parameters to be estimated; therefore (4) may be preferred when historical data is scarce. Furthermore, efficient estimation of (5) requires that u be white noise. Errors from an n-step-ahead forecast are commonly thought to follow an MA($n - 1$) process, so a generalized least squares procedure must be employed for multistep forecasts. And, in

9. The material in the following two paragraphs is treated in greater detail in Granger and Newbold (1977, pp. 269–72) and in Granger (1980, pp. 158–61).

common with the ratio technique, the least squares technique leaves open the question of which time period should serve as bounds for the estimation.

Although the least squares technique has an intuitive appeal because of its similarity to the basic methodology of empirical economics, Granger and Newbold (1977, pp. 271–78) describe results that indicate that ratio techniques, such as (4), outperform more complex forms that are similar to (5).[10] For this paper, a limited set of tests was examined for particular variables and forecasters. Postsample forecast combinations, employing only the information that would have been available to the forecast combiner at the time of the combination, were constructed using three forecasters, three variables, and both combination techniques. Three different sets of boundaries were used in conjunction with both techniques: a moving range of the last 8 quarters, a moving range of the last 16 quarters, and the range that included all available data. In addition, to capture serial correlation in the least squares regressions, combinations were computed five ways: under OLS and with AR(1), AR(2), MA(1), and MA(2) corrections. One-quarter-ahead forecasts are available starting in 1970:3; 2- and 4-quarter forecasts begin in 1970:4 and 1971:2, respectively.

Table 9 presents the lowest average error for each technique, pair of forecasters, and variable, computed over the period 1976:1–1983:4. Results are reported only for the complete estimation range since, for the least squares technique, the two shorter bounds produced inferior results and results remained essentially the same for the ratio technique under the various boundaries. The results support the findings of Granger and Newbold: for the variables and forecasters under consideration here, the simpler ratio technique consistently outperforms the more complicated least squares technique.[11]

While a detailed discussion of these results is beyond the scope of this paper, some possible explanations for the superiority of the ratio technique may be advanced. Granger and Newbold assert that "population correlation coefficients are generally not well estimated in small samples" (1977, p. 272); in other words, the covariance present in estimates of formula (5) may only be known with a large variance. The same may well be true of bias, as measured by the constant term in (5); note that only the magnitude of $\hat{w}_0$ enters (5), not its standard error. Additionally, all econometricians are aware that a parameter that is significant in-sample may not be useful for postsample prediction. For-

10. Granger may have reconsidered, however; see Granger and Ramanathan (1984).

11. Einhorn and Hogarth (1975) have suggested that mean weighting may be preferable to any more complex method. In our tests, mean weighting was superior to least squares and no better or worse on average than the ratio technique.

TABLE 9 Two Techniques for Combining Forecasts

	Nominal GNP (1 Quarter)	GNP Deflator (2 Quarters)	Real GNP (4 Quarters)
DRI	5.2	1.6	2.2
Wharton	4.9	1.7	2.2
VAR	5.7	2.1	2.3
Least squares technique:			
DRI + VAR	5.4*	1.7†	2.5†
Wharton + VAR	5.1†	1.8†	2.4
DRI + Wharton	5.4‡	1.9§	2.4†
Ratio technique:			
DRI + VAR	5.0	1.6	1.9
Wharton + VAR	4.8	1.6	1.9
DRI + Wharton	5.0	1.6	2.1

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts and combinations. RMSEs were calculated over the period 1976:1–1983:4. Estimation was performed over all available data. The best least squares combinations were generated with OLS, except as noted below.
* AR(2) correction employed.
† MA(1) correction employed.
‡ AR(1) correction employed.
§ MA(2) correction employed.

mula (4) may outperform formula (5) simply because in-sample bias has no predictive power.[12]

On a more mundane level, formula (5) may simply be overparameterized. There is some evidence for this speculation: the shorter boundaries (8 and 16 quarters) produced results noticeably inferior to those generated by using the full data set. On the other hand, the results for the ratio technique were robust with respect to the estimation bounds. It is also true that our practice of employing the same serial correlation correction over the entire sample period in formula (5) may be a poor approximation of reality since the actual error process may have changed with time. Finally, since the forecasts are often similar (this is especially true of the commercial forecasters), multicollinearity is likely to be a problem in a regression like (5).

The Combination of Structural and VAR Forecasts

The results presented in table 9 provide strong evidence that, for the variables and forecasters relevant to this study, the ratio technique for combining forecasts is superior to the least squares technique. Table 10 compares forecasts produced by the two commercial forecasting services for which forecasts for the complete period were available (for brevity, they are referred to below as structural models), a VAR

12. For this reason, in-sample results of forecast combinations, such as those presented by Nelson (1984), are unpersuasive. An entire literature has developed that challenges the rational expectations hypothesis on the basis of similar in-sample results. Webb (1984*a*) demonstrates that postsample results can produce opposite conclusions.

TABLE 10 The Combination of Structural and VAR Forecasts

Variable: Forecast Horizon (Quarters)	Individual Forecasts			Combined Forecasts				
	DRI (1)	Chase (2)	VAR (3)	1 + 2	1 + 3	2 + 3	1 + 2 + 3	ASA/NBER
Real GNP:								
1	4.2	4.5	4.7	4.3	4.1*	4.3	4.3	4.1
2	3.1	3.5	3.7	3.3	2.9*	3.1	3.5	3.2
4	2.2	2.5	2.3	2.3	1.9*	2.2	2.2	2.3
6	1.8	1.9	2.2	1.8	1.5*	1.7	1.6	. . .
GNP deflator:								
1	1.8	1.6*	2.4	1.7	1.8	1.7	1.6*	1.7
2	1.6	1.6	2.1	1.5*	1.6	1.6	1.5*	1.5*
4	1.6*	1.8	2.1	1.7	1.6*	1.7	1.7	1.6*
6	1.8*	2.1	2.4	1.9	2.0	2.0	2.0	. . .
Nominal GNP:								
1	5.2	5.9	5.7	5.5	5.0*	5.4	5.6	5.0
2	4.0	4.8	4.5	4.3	3.8*	4.2	4.5	4.2
4	3.4	4.2	3.1	3.7	3.0*	3.5	3.8	3.6
6	3.1	3.8	3.0*	3.3	3.0*	3.4	3.4	. . .
T-Bill rate:								
1	1.7	1.7	1.4*	1.7	1.4*	1.4*	1.5	. . .
2	2.4	2.5	2.3	2.4	2.2*	2.3	2.3	. . .
4	3.1	3.4	3.1	3.2	3.0*	3.1	3.1	. . .
6	3.7	4.0	3.7	3.7	3.6*	3.8	3.8	. . .

NOTE.—Data are root-mean-square errors (RMSEs) from postsample forecasts and combinations. RMSEs were calculated over the period 1976:1–1983:4. The ratio technique was used for combinations, with estimation performed over all available data.

* Best forecast for each variable and horizon.

model, and the ASA/NBER median survey[13] to forecasts combined by using the ratio technique. The result most obvious from table 10 is that combining structural and VAR forecasts is a useful exercise. Of the 16 examples in the table (four variables at four horizons each), the combination of the best structural forecaster and the VAR forecast is most accurate 13 times, including ties. Since forecasts from Data Resources (DRI) are superior to those of Chase in nearly all instances over this period, it is not surprising that the DRI-VAR combinations dominate the Chase-VAR combinations.

Of the various types of combined forecasts presented in table 10, the structural-VAR team seems to be superior. The DRI-VAR combination beats the DRI-Chase combination 13 of 16 times; it is also superior to the combination of two structural forecasts and the VAR forecast in 13 instances; and it is either better than or equal to the accuracy of the ASA/NBER median in seven of nine cases.[14] Overall, some form of combined forecast beats the best individual forecast 11 out of 16 times, while only once is an individual forecast superior to the best combination.

Two forecasters' techniques can be compared by studying the performance of a combination of the two forecasts. If the techniques have a meaningful difference in that each adds information to the other, then the combination of the two should produce synergism: the combination should be better than its components. By this yardstick, structural forecasters appear to be alike, while the VAR technique is often different from the structural method. Combinations of structural and VAR forecasts produce synergism in 18 of 32 cases. In eight of the remaining 14 instances, the VAR forecast was superior to the structural forecast, so the absence of synergism means that, while the VAR forecast adds information to the structural forecast, the opposite is not true. In total, VAR forecasts add information to structural forecasts in 26 of 32 cases. However, DRI-Chase combinations are synergistic in only one of 16 cases, and the three-way combination results in synergy only twice. These results suggest that there is little return to the information in a second structural forecast, given knowledge of one such forecast.

The incremental forecast improvement from a VAR model relative to a structural model depends on the variable under consideration. The evidence summarized in table 10 indicates that adding information

13. The ASA/NBER survey questionnaire is distributed in mid-quarter, and results are reported at the end of the quarter. McNees and Ries (1983) classify it as a mid-quarter forecast. As such, it is at a disadvantage when compared to the late-quarter forecasts under consideration. ASA/NBER only began publishing interest rate forecasts in 1981:3, and it does not release 6-quarter-ahead forecasts.

14. We performed a limited test of the usefulness of combining the ASA/NBER and VAR forecasts, using data starting in 1976:1 to produce combinations from 1980:1 through 1983:4. The combination was as good as or better than the components in seven of nine cases.

from a VAR forecast to a structural forecast is often useful for real GNP, nominal GNP, and the Treasury bill rate but not very useful when forecasting the GNP deflator. Thus the strong pattern of synergism in the combined forecasts of the other three variables indicates a substantial difference in the information that the two types of models employ in producing forecasts of real and nominal GNP and interest rates.

Since the return to combining similar forecasts is near zero, there may be better combination techniques than the ASA/NBER method, in which a large number of forecasters, many of them similar, are surveyed and all given equal weight. An alternative approach that could prove more successful is to divide the universe of forecasters into groups, each group consisting of forecasters who are essentially alike. One would compare the members of a group, choose the best forecaster in each, and then combine the best from each group. For example, one might usefully include one structural, one VAR, one ARIMA, and one "informal" or judgmental forecaster with an established track record.

Conclusion

This paper has proposed the use of a small, mechanically operated VAR model in the evaluation of macroeconomic forecasts. It is shown that simplicity of construction, ease of operation, and relative accuracy make the VAR model under consideration a useful benchmark. It may come as a surprise to some that the VAR model produces forecasts that are competitive with those issued by three well-known commercial forecasters over the period 1970–83. Forecasts from the unrestricted VAR model are also competitive with those produced by ARIMA techniques and by a more complex Bayesian VAR method. It is also argued that VAR forecasts contain information that is systematically ignored by commercial forecasting models. Indeed, it appears that the consumer of macroeconomic forecasts can reduce the errors associated with structural econometric models in a simple, inexpensive way by building a small VAR model and combining its forecasts with those from a commercial forecasting service.

References

Einhorn, H. J., and Hogarth, R. M. 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* 13 (April): 171–92.

Goldfeld, S. M. 1976. The case of the missing money. *Brookings Papers on Economic Activity*, no. 3, pp. 683–730.

Granger, C. W. J. 1980. *Forecasting in Business and Economics*. New York: Academic Press.

Granger, C. W. J., and Newbold, P. 1977. *Forecasting Economic Time Series*. New York: Academic Press.

Granger, C. W. J., and Ramanathan, R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* 3, no. 2:197–204.

Hakkio, C. S., and Morris, C. S. 1984. Vector autoregressions: A user's guide. Research Working Paper no. 84-10. Kansas City, Mo.: Federal Reserve Bank of Kansas City, November.

Klein, L. R. 1984. The importance of the forecast. *Wharton Annual* 8:87–93.

Litterman, R. B. 1979. Techniques of forecasting using vector autoregressions. Working Paper no. 115. Minneapolis: Federal Reserve Bank of Minneapolis.

Litterman, R. B. 1984. Specifying vector autoregressions for macroeconomic forecasting. Research Department Staff Report no. 92. Minneapolis: Federal Reserve Bank of Minneapolis.

Litterman, R. B. 1985. Economic forecasts from a vector autoregression, 1980–83. (monthly release).

McNees, S. K. 1975. An evaluation of economic forecasts. *New England Economic Review* (November/December): 2–39.

McNees, S. K., and Ries, J. 1983. The track record of macroeconomic forecasts. *New England Economic Review* (November/December): 5–18.

Nelson, C. R. 1972. The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review* 62 (December): 902–17.

Nelson, C. R. 1984. A benchmark for the accuracy of econometric forecasts of GNP. *Business Economics* 19 (April): 52–58.

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

Todd, R. M. 1984. Improving economic forecasting with Bayesian vector autoregression. *Economic Review* (Federal Reserve Bank of Minneapolis) 8 (Fall): 18–29.

Webb, R. H. 1984*a*. Inadequate tests of the rationality of expectations. Working Paper no. 84-7. Richmond, Va.: Federal Reserve Bank of Richmond, October.

Webb, R. H. 1984*b*. Vector autoregressions as a tool for forecast evaluation. *Economic Review* (Federal Reserve Bank of Richmond) 70 (January/February): 3–11.

Zarnowitz, V. 1979. An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level. *Journal of Business* 52 (January): 1–33.