

# VECTOR AUTOREGRESSIONS AS A TOOL FOR FORECAST EVALUATION

Roy H. Webb

There are two important reasons for examining the historical accuracy of economic forecasts. For one, current users of economic forecasts need a guide to the probable accuracy of the projections they receive. Although the past record cannot perfectly predict future accuracy, it does provide valuable guidance. From another perspective, economists are interested in whether conventional model-building techniques provide a useful framework for economic research and policy analysis. One test of conventional large econometric models is whether or not they provide accurate forecasts. If not, one may then question other products of that framework as well.

Although one can compile a record of forecasts, compare them to actual results, and calculate descriptive statistics such as average errors, such summaries by themselves do not tell us whether a forecaster's record is especially good or bad. What is needed is a standard against which to judge a series of forecasts.

This article uses a relatively new statistical procedure, a vector autoregressive (VAR) model, as a standard of comparison for other forecasts. The article first explains how structural models are conventionally employed to generate forecasts. Conventional procedures for constructing and using large models are not endorsed by all economists, however, and a few objections are mentioned. Next, the article describes VAR models and explores their usefulness for generating forecasts. Also, it compares a particular VAR model's forecasts with a series of forecasts from a large structural model as well as with a composite forecast derived from a large number of individual forecasters. The final topic is the VAR model's estimate of the precision of its forecasts.

## Forecasts from Large Structural Models<sup>1</sup>

Economic theory can be used to impose structure on data sets by specifying exactly how variables may interact. One purpose of such restrictions is to produce superior forecasts. For example, a widely-used theoretical representation has the demand for real money balances depending on real GNP and an interest rate. This could be written

$$(1) \quad \frac{M}{P} = L(X, R)$$

where  $M$  represents the nominal money supply,  $P$  is the price level,  $L$  is a specific liquidity preference (or money demand) function,  $X$  is real GNP, and  $R$  is an interest rate. In order to generate forecasts of the left-hand variable, it is conventional to approximate equation (1) by

$$(2) \quad \log\left(\frac{M}{P}\right) = c + b_1 \log(X) + b_2 \log(R) + e$$

where  $\log$  indicates a logarithm;  $c$ ,  $b_1$ , and  $b_2$  are coefficients which can be statistically estimated from historic data; and  $e$  is an error term which is random noise if the theory embodied in equation (1) and its approximation, equation (2), are valid.

---

<sup>1</sup> Well-known large structural models include the Brookings Model, the Chase Econometrics Model, the Data Resources Model, the FMP Model, and the Wharton Model. Those models above are often referred to as Keynesian, due to their emphasis on the importance of aggregate demand and their analysis of demand by sectors (consumption, investment, etc.). The word Keynesian may be misleading, however, since Keynes himself [7] found fault with many statistical procedures used by today's model builders. Also, a large structural model could employ non-Keynesian theory and be vulnerable to all the objections mentioned in the text.

Once the coefficients in equation (2) are estimated, that expression can be used to predict real money balances, given values for real GNP and the interest rate. Such predictions of real money balances have not always been accurate,<sup>2</sup> and have typically led to modifications of equation (2).

One modification that is often made is to add the so-called lagged dependent variable term  $b_3 \log \left( \frac{M}{P} \right)_{-1}$  to the right-hand side of equation (2). Such a term is not rigorously derived from the theory underlying equation (1).<sup>3</sup> However, econometric investigators have found that including lagged values of the dependent variable often improves the statistical fit of an equation—that is, its average prediction errors are smaller within the time span over which the equation's coefficients are estimated. Another ad hoc technique might be to include additional lagged values of real GNP and the interest rate on the right side of equation (2).

As a result of those modifications, an equation for the demand for money might be (omitting the logs for notational convenience)

$$(3) \quad \frac{M}{P} = c + b_{11}X + b_{12}X_{-1} + b_{21}R + b_{22}R_{-1} + b_3 \left( \frac{M}{P} \right)_{-1} + e.$$

Although equation (1) can be derived from optimizing behavior of a representative individual, equation (3) specifies more complex behavior that is not derived from a dynamic model of an individual's optimizing decisions. Instead, it simply reflects statistical modifications that have been found to be consistent with the data.

Another objection to equation (3) is that real GNP and the interest rate are not truly exogenous—that is, they are not determined independently of real money balances. On the contrary, each variable influences the other as they are jointly determined. The main purpose of building large models is to take such interdependencies into account. In this example, there could be separate equations for the money

<sup>2</sup> See Judd and Scadding [6] for a thorough account.

<sup>3</sup> Investment in physical capital can be modeled as a "stock adjustment" process, which gives rise to a lagged dependent variable. Chow [1] used an analogy of money to consumer durables to justify the stock adjustment process. He did not, however, specify why adjustment of actual to desired money balances is so costly that it is not instantaneous. Since money can be easily exchanged for physical commodities or financial assets, the analogy of a stock of money to a stock of physical capital is unclear without a more complete model of transactions technologies.

supply, the price level, real GNP, and the interest rate. That approach, however, leaves two problems unresolved. First, although such simultaneous equation models require specialized econometric techniques, the complexity of many structural models may preclude the use of those techniques.<sup>4</sup> A second problem is that there are very few really exogenous variables (for example, a time trend, weather, and wars).

A final concern is the treatment of expectations. Since economic decisions of individuals are often based on what they expect to happen in the future, it might be more accurate to replace actual with expected real GNP in equation (1). In other words, an individual's demand for real balances would depend on his expected income rather than previously realized income.

Expectations raise a particular problem for model builders, however, since individuals' expectations are not observed directly. Rather than model the process of expectations formation, conventional practice is to substitute a series of lagged values for the expected future value of a variable. Such a practice is frequently observed in an equation such as

$$(4) \quad w = c + b_1U + \sum a_i p_{-i} + e$$

where  $w$  is the growth rate of wages,  $U$  is the unemployment rate,  $p_{-i}$  is the growth rate of prices  $i$  periods in the past,  $e$  is the error term, and the  $a_i$ 's,  $b_1$ , and  $c$  are coefficients that can be estimated. In equation (4) (often referred to as a Phillips Curve) the lagged inflation terms are meant to represent an individual's expectation of future inflation. Economic theory, however, does not support that representation as an individual's best effort to predict future inflation.

Thus the following areas of conventional model-building practice have been challenged: (1) many key structural equations are not actually derived from the theory they purport to represent, (2) many variables are inappropriately labeled as exogenous, and (3) while expectations of future events determine many actual economic decisions, they are typically entered into a large model in a crude, theoretically unjustified manner. Although by no means an exhaustive critique of large structural models,<sup>5</sup> those

<sup>4</sup> Los [10], for example, has criticized the use of ordinary least squares to estimate the FMP model, rather than using simultaneous equation methods.

<sup>5</sup> For a more complete critique, see Sims [12]; also, for a more thorough explanation of the construction of and philosophy behind large models, see Eckstein [3].

difficulties illustrate why many economists do not automatically accept the models' results. Yet if the models had a documented history of performing well, the force of those objections would be muted. Thus the relatively new statistical technique described below is of particular interest as a standard against which one product of the large structural models can be measured.

Other products of large models such as policy evaluation and hypothesis testing are at least as important as forecasting. Yet it is much harder to assess their performance in those areas than it is to measure predictive accuracy. Therefore, the forecasting performance of large models may be the only empirical evidence available to judge the success of modeling efforts.

### VAR Models

In sharp contrast to the structural approach described above, a VAR model uses little economic theory. Therefore, VAR models make no attempt to satisfy the objections made concerning the theoretical specification of conventional models. In this and in other areas, both VAR and conventional models are thus suspect a priori. It is an empirical question as to which model actually produces better forecasts.

An extremely simple VAR model is illustrated by equations (5) and (6) below:

$$(5) \quad R = b_1M_{-1} + c_1R_{-1} + e_1$$

$$(6) \quad M = b_2M_{-1} + c_2R_{-1} + e_2$$

where M and R represent the money supply and an interest rate, the b's and c's are coefficients, and the e's are error terms. Note that the money supply and the interest rate are treated symmetrically. Each is determined only by its own lagged value and the lagged value of the other variable. As a practical matter, much longer lags are necessary in order to generate adequate predictions. Accordingly, in the model which is described below, six lagged values are included for each variable. Also, most VAR models use more than two different variables, and in the model below, five variables are included. The two equations above, however, illustrate the essence of the VAR approach.

The VAR model thus provides a conceptually straightforward method of producing forecasts that do not assume particular values of exogenous variables. At any point in the past, it is possible to estimate a VAR model's coefficients based on data through that point in time and then produce forecasts as far ahead as desired. Those forecasts, in

turn, can be compared with actual results. Since the forecasts are mechanically generated and are based on data available at the time of the forecast, they provide a legitimate comparison for previously published forecasts from other sources.

VAR forecasts have a special appeal when used as a standard of comparison for forecasts from large structural models because the VAR models do not impose the controversial theoretical restrictions that those models contain. In particular, VAR models do not employ dubious exogeneity definitions. That is especially important for variables manipulated in the conduct of monetary policy. Although the large structural models often treat Federal Reserve actions as exogenous, some analysts believe that the Fed has usually responded in a predictable manner to the state of the economy, and therefore Federal Reserve actions are jointly determined with other macroeconomic variables.<sup>6</sup>

Thus on some points the VAR strategy avoids problems faced by conventional models. However, the VAR models' lack of theory and small number of variables lead many analysts to question their usefulness. It is therefore especially interesting to examine the actual performance of VAR and structural models. Although a model's performance has several dimensions, the easiest to measure is the accuracy of its forecasts. Accordingly, the following section contains some evidence on the forecasting ability of a particular VAR model.

### A Comparison of Forecasts

This section compares recent forecasts from three sources: a major consulting service, a survey of professional forecasters, and a VAR model. Forecasts began in the first quarter of 1976 and were taken through the third quarter of 1983. Details of the VAR model's construction are provided in the Appendix. The survey covers as many as seventy professional forecasters. Average values from the survey have been found to be more accurate than most individual forecasters.<sup>7</sup> The consulting service bases its forecasts on a large structural model, but modifies the model forecast with the judgment of its staff before its forecasts are published. A calendar quarter's last monthly forecast (usually issued during the last week of the quarter) was used.

<sup>6</sup> For a more detailed account of Federal Reserve response to economic conditions, see Hetzel [5].

<sup>7</sup> See Zarnowitz [13] for a description of the survey, and Zarnowitz [14] for an analysis of errors from forecasts derived from the survey.

As noted in the Appendix, in some respects the comparison favors the VAR model due to the procedures used to construct the model. Also, the VAR forecasts had access to the latest revisions of published data. Offsetting those advantages, however, are two important factors. While the VAR model only employs five variables, the structural model contains several hundred. That additional information should help improve the accuracy of its forecasts. In addition, unusual events such as the Carter credit controls of 1980 could have been incorporated into the published forecasts via judgmental adjustments. Therefore, after considering these factors, it is the author's judgment that the consulting service should have been able to provide forecasts with substantially greater accuracy than the VAR model if their model's theoretical restrictions were valid.

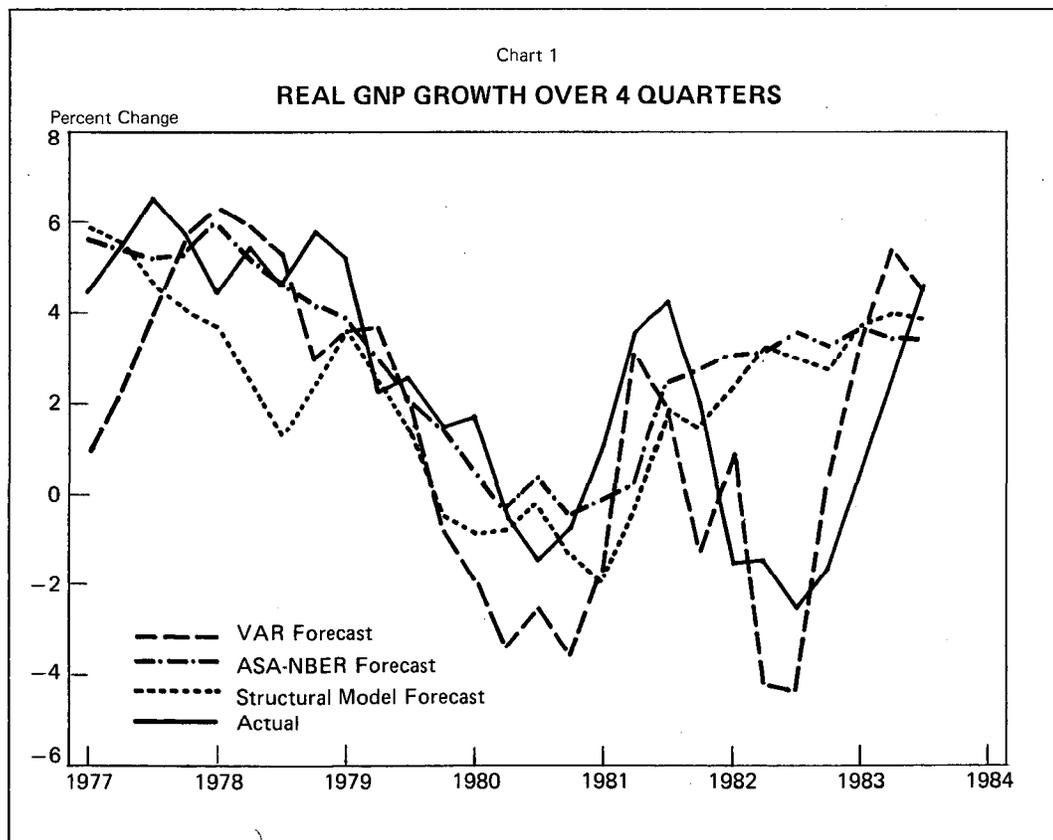
Charts 1-3 illustrate four-quarter-ahead forecasts and actual outcomes, with summary statistics given in table I for one-, four-, and eight-quarter forecasts. Some observers have questioned the accuracy of VAR predictions. Lawrence Klein, for example, is reported to have expressed the view that "VAR models are all right for predictions one quarter ahead,

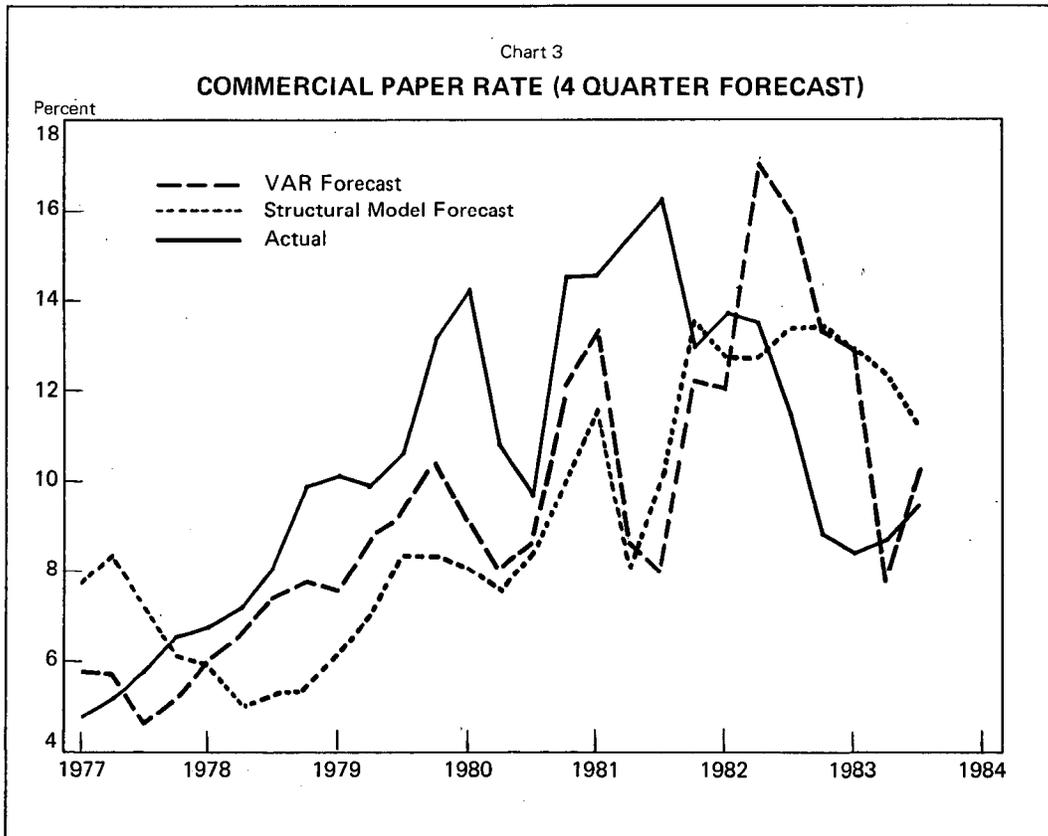
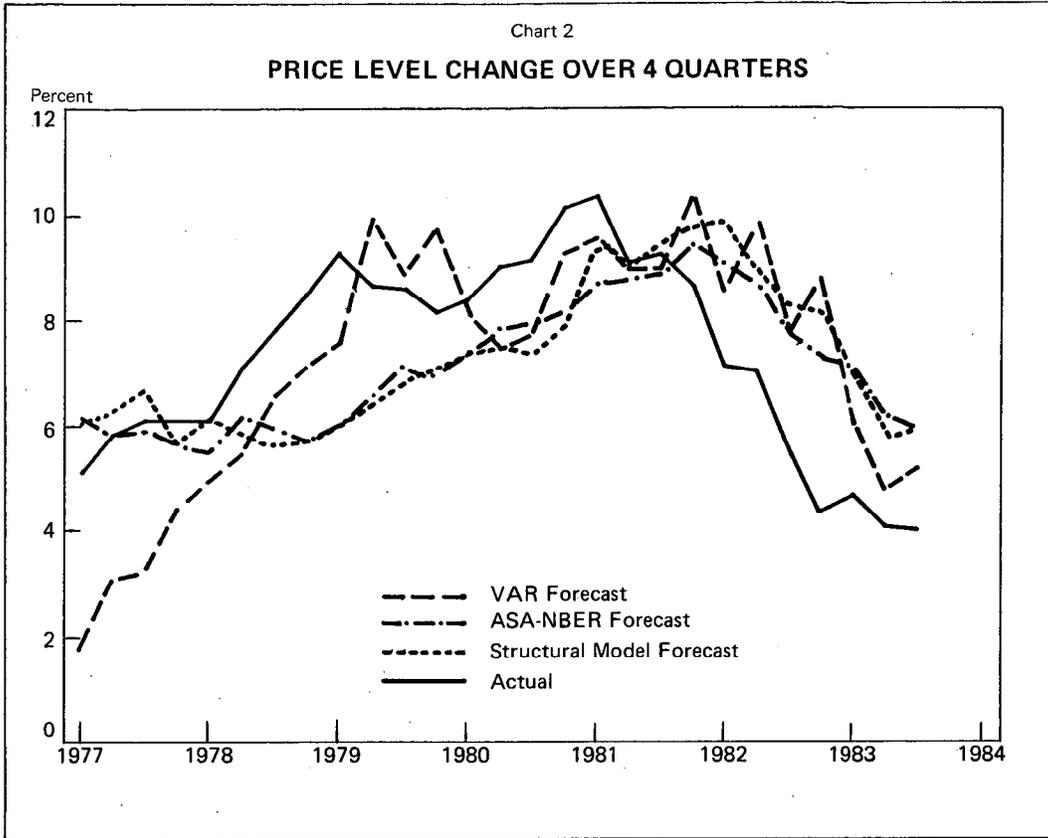
Table I  
**FORECAST ERRORS**

(Percent)

	Forecast Horizon		
	1 Quarter	4 Quarters	8 Quarters
<b>Real GNP Growth</b>			
VAR	4.49	2.36	1.56
Forecasting Service	4.66	2.65	1.89
ASA-NBER	4.23	2.36	
<b>Inflation Rate</b>			
VAR	2.62	1.85	2.46
Forecasting Service	1.65	1.87	2.21
ASA-NBER	1.80	1.70	
<b>Commercial Paper Rate</b>			
VAR	1.78	3.10	5.00
Forecasting Service	1.82	3.58	5.09

NOTE: Entries represent the root mean squared difference between actual and predicted values. Real GNP and inflation are percent changes expressed as annual rates. The commercial paper rate is the quarterly average value. Actual values range from 1976 Q2 to 1983 Q3 for one-quarter forecasts, from 1977 Q1 to 1983 Q3 for four-quarter forecasts, and 1978 Q1 to 1983 Q3 for eight-quarter forecasts. The ASA-NBER survey did not include an interest rate for the entire period, and also did not include eight-quarter forecasts.





but VAR predictions quickly deteriorate so that conventional models offer superior predictions further in the future." [2] The results shown here clearly contradict Klein's view. At a four-quarter horizon, the VAR model's predictions are more accurate than both published forecasts for real GNP, and more accurate than the forecasting service for inflation and the interest rate. And at an eight-quarter horizon, the VAR model's forecasts are more accurate than the forecasting service for real GNP and the interest rate. It is especially noteworthy in chart 1 that only the VAR model predicts the 1982 recession.

There is additional evidence from other models. Stephen McNees [11] has found that for real GNP and the unemployment rate, published forecasts from a VAR model constructed by Robert Litterman were more accurate than three large structural models at four- and eight-quarter horizons. (McNees, however, had only four observations at the longest horizon, making his comparisons tentative at this stage. Also, his results were less favorable for the Litterman model for several other variables.) Litterman [8] also compared a VAR model's performance with that of seven major forecasters from 1970-75, and found better performance from the VAR in many cases, especially at longer horizons.

### Uncertainty of Forecasts

Another use of VAR models is to estimate the uncertainty attached to a particular forecast. Since the VAR forecasts are not judgmentally adjusted, they yield objective estimates of uncertainty. In contrast, it is difficult to imagine an objective measure of the accuracy of judgmental adjustments that will be made to forecasts from large structural models.<sup>8</sup>

Forecast errors can be traced to several sources. One source is the error term included in statistical models. Taking equation (2) as an approximation to equation (1), for example, gives rise to such an error term. That modeled error can be expected to cause forecasts from both VAR and structural models to differ from actual outcomes. The variance of future errors from that source can be estimated using errors within the sample period. A second source of prediction errors for both types of models is that the

---

<sup>8</sup> The author is aware of only one large structural model that does not routinely modify the model forecasts.

coefficients are not known with perfect accuracy, but instead are statistically estimated and thus are to some extent erroneous. Another problem for structural models is the error in predicting future values of exogenous variables. Finally, the extent to which a model is incorrectly specified will add to forecast error. Some potential misspecifications are noted above for structural models. A misspecification that is particularly applicable to small VAR models is that relevant explanatory variables are omitted, thereby causing the in-sample error term to understate the true imprecision of forecasts.

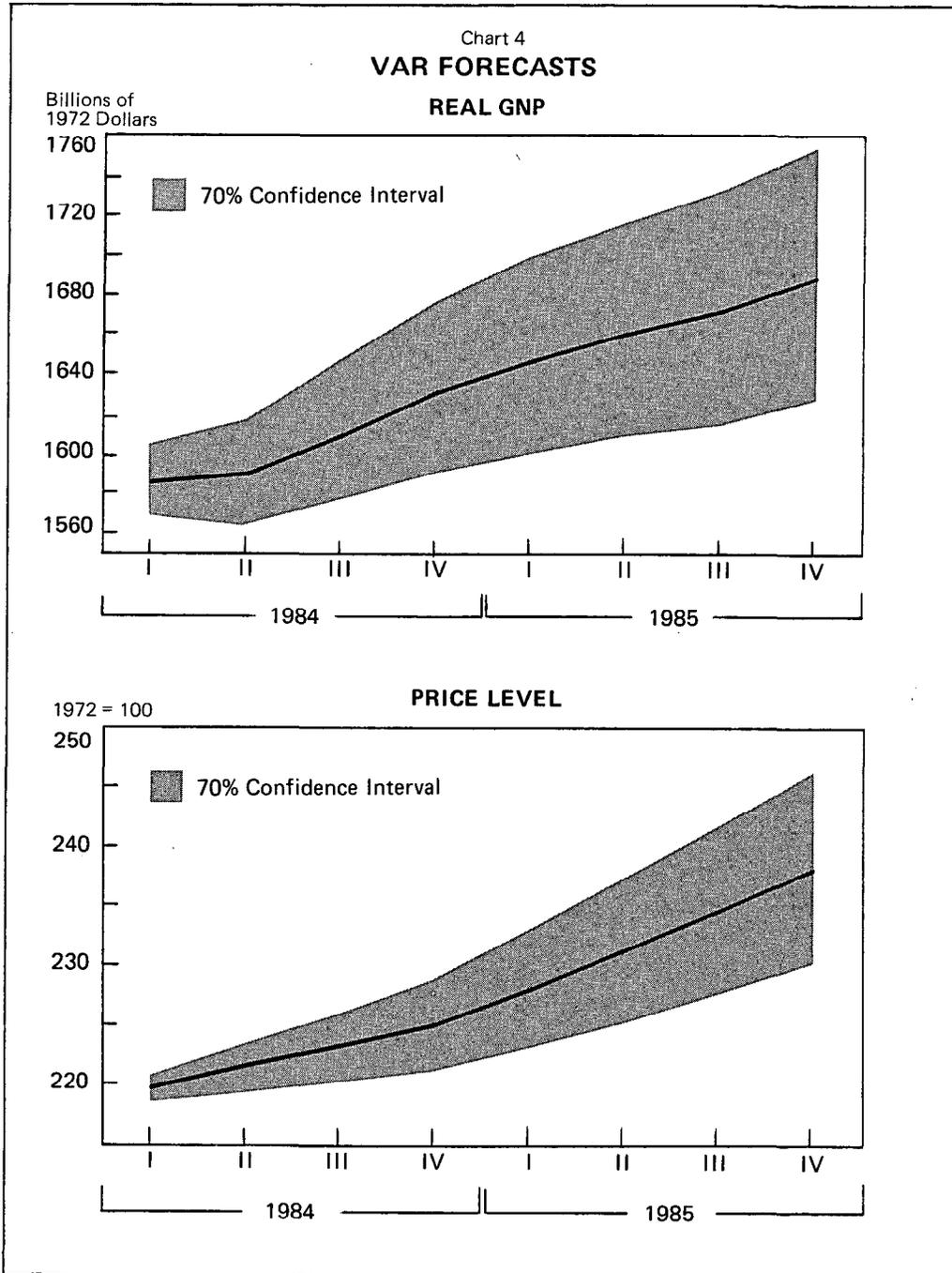
Analyzing probable forecast errors due to in-sample errors, errors in estimating coefficients, and errors in predicting exogenous variables is a conceptually straightforward task. Estimating probable forecast errors due to model misspecification, however, is much more difficult. Fair [3] has attempted this latter task for several models, and has found the probable error due to misspecification to be sizeable for both a VAR and a structural model.

The VAR model's probable forecast errors presented below account only for the first type of error, and thus are best interpreted as an upper bound on the probable accuracy of current forecasts. Even so, the illustrated imprecision is considerable. To illustrate, chart 4 contains the VAR forecast for real GNP and price level in 1984-85 and confidence intervals for that forecast. Taking account of the error mentioned above, the shaded areas indicate that there is a 70 percent likelihood that the actual value will fall within that range. The charts thus indicate a large degree of imprecision in forecasts which prospective users should take into account.

### Conclusion

There is a limited amount of information in our time series of economic data, and economists do not agree on the best strategy for extracting that information. Innumerable hours of labor have been devoted to building ever-larger models with continual ad hoc adjustments. Another strategy is to use relatively simple VAR techniques. This paper poses the question: which strategy actually produces more useful information?

In this article, the amount of useful information is measured by the accuracy of forecasts. If small, atheoretical VAR models can consistently match the



forecasting accuracy of large structural models, that could lead one to question the usefulness of the large models' theoretical restrictions for other purposes, such as policy evaluation and formally testing hypotheses concerning the structure of the economy. The results here are not conclusive. (Comparing a long series of published VAR forecasts with large models singled out for having the best forecasting

records would permit a more conclusive judgment to be made.) Nonetheless, the fact that in many comparisons, post-sample predictions from a simple VAR model did well vis-à-vis the published forecasts of a major consulting service as well as the median forecast from a survey of forecasters over a seven year period should encourage further research with this relatively new method.

## APPENDIX

This section describes the construction of a VAR model in sufficient detail so that the reader may (1) judge the extent to which experimentation in model construction qualifies the conclusions in the text, and (2) replicate the model and the results cited in the text.

Five variables are employed: the six-month commercial paper rate, the monetary base, the capacity utilization rate, the GNP implicit price deflator, and real GNP. The commercial paper and capacity utilization rates are levels (quarterly averages), and the other variables are percent changes from the previous quarter at annual rates. The data were taken from Citibank's on-line data base, updated through November 1983. All data were available starting in 1947, except for capacity utilization, which began in 1948. The model was estimated with six lagged values for each variable for every equation, in addition to five constant terms, yielding 155 estimated parameters.

One change that improved the inflation forecasts was substituting the monetary base for M1. Forecast statistics from the M1 specification are also shown in table II. Thus the form of the model shown in table III was based on some experimentation, namely: (1) the substitution of the monetary base for M1; and (2) the author's prior knowledge that these five variables moved together over recent years. Such experimentation, of course, was not available to the producers of the forecasts to which the VAR forecasts are compared in the text.

### Extensions and Improvements

The model as described above is unusually simple. Complications were deliberately avoided in order to make its workings easy to follow. There are several obvious changes which could improve the accuracy of its forecasts, however.

Although the variables were treated symmetrically in each equation, other approaches are possible. For example, restricting the lag lengths when the longest lags contribute little information could allow more accurate estimation of the remaining coefficients and thus more accurate forecasts. This could be accomplished by an ad hoc process, such as removing the last lagged value when the final t-statistic is near zero. Robert Litterman [8, 9] has used a more complicated procedure that allows a forecaster to

Table II

#### FORECAST ERRORS FROM VAR SIMULATIONS

(Percent)

Variables	Real GNP	GNP Deflator	Commercial Paper Rate
R B C P X	2.32	1.86	3.19
R M C P X	2.47	2.82	3.02

NOTE: Each entry is the root mean squared difference between actual and predicted values. Forecasts are average growth rates over four quarters for real GNP and the deflator, and the level four quarters ahead of the interest rate. Actual values ranged from 1977 Q1 to 1983 Q4.

Table III

#### A VECTOR AUTOREGRESSIVE MODEL

Each of five equations has the form

$$V_i = k_i + \sum_{l=1}^6 r_{il}R_{-l} + \sum_{l=1}^6 m_{il}M_{-l} + \sum_{l=1}^6 c_{il}C_{-l} + \sum_{l=1}^6 p_{il}P_{-l} + \sum_{l=1}^6 x_{il}X_{-l}$$

where

*i* is the equation number (1, 2, 3, 4, 5)

*l* is the lag number

*k<sub>il</sub>*, *r<sub>il</sub>*, *m<sub>il</sub>*, *c<sub>il</sub>*, and *x<sub>il</sub>* are estimated coefficients

*V<sub>i</sub>* is the dependent variable

*R* is the six-month commercial paper rate

*M* is the monetary base (St. Louis), percent change

*C* is the capacity utilization rate (Federal Reserve Board)

*P* is the GNP implicit price deflator, percent change

*X* is real GNP, percent change

introduce prior beliefs concerning the distribution of the coefficients on the lagged terms. He found that such restrictions did improve forecast accuracy in several VAR models.

The model was estimated over the entire period for which quarterly data were readily available. It is likely, however, that the structure of the economy has changed between 1947 and 1983. Thus it is possible that a later starting date would provide more accurate forecasts. An alternative strategy would be to allow

the estimated parameters to vary over time, thereby capturing any changes in the economic structure. Litterman [9] has reported positive results from such a procedure.

Therefore, the VAR model discussed above does not attempt to employ many statistical techniques that might improve its predictive accuracy. That it, nonetheless forecasts relatively well indicates the robustness of the VAR approach to economic forecasting.

### References

1. Chow, Gregory C. "On the Long-Run and Short-Run Demand for Money." *Journal of Political Economy* 74 (April 1966), 111-129.
2. "Comments and Discussion," following Christopher A. Sims, "Policy Analysis with Econometric Models." *Brookings Papers on Economic Activity* (1: 1982), 164.
3. Eckstein, Otto. *The DRI Model of the U. S. Economy*. New York: McGraw-Hill, 1983.
4. Fair, Ray C. "An Analysis of the Accuracy of Four Macroeconomic Models." *Journal of Political Economy*, 87 (August 1979), 701-718.
5. Hetzel, Robert L. "The Business Cycle and Federal Reserve Monetary Policy." Paper submitted to Federal Reserve System Research Committee on Financial Analysis, Richmond, 1979.
6. Judd, John P., and John L. Scadding. "The Search for a Stable Money Demand Function." *Journal of Economic Literature*, 20 (September 1982), 993-1023.
7. Keynes, J. M. "Professor Tinbergen's Method." *Economic Journal* (September 1939), pp. 558-568.
8. Litterman, Robert B. "Techniques of Forecasting Using Vector Autoregressions." Working Paper No. 115, Federal Reserve Bank of Minneapolis, 1979.
9. ————. "Specifying Vector Autoregressions for Macroeconomic Forecasting." Working Paper No. 208, Federal Reserve Bank of Minneapolis, 1982.
10. Los, Cornelis A. "Identifiability and Estimability of the FMP Model." Federal Reserve Bank of New York, Research Paper 8307, 1983.
11. McNees, Stephen K. "Forecast Accuracy Relative to Mechanical VAR Model." Presented to Federal Reserve System Research Committee on Business Analysis, Washington, 1983.
12. Sims, Christopher A. "Macroeconomics and Reality." *Econometrica* 48 (January 1980), 1-48.
13. Zarnowitz, Victor. "The New ASA-NBER Survey of Forecasts by Economic Statisticians." *The American Statistician* (February 1969), 12-16.
14. ————. "The Accuracy of Individual and Group Forecasts from Business Outlook Surveys." NBER Working Paper No. 1053 (December 1982).

### CORRIGENDUM

The note corrects two errors in the article, "Why Economic Data Should Be Handled with Care: The Case of the Suspiciously Slow Growth Statistic," published in the July/August 1983 issue of this *Review*. In the fourth paragraph, the penultimate sentence should read, "In order to estimate real GNP, the Department's analysts adjust the current dollar figure for inflation by dividing each detailed component of nominal GNP by a specific price deflator." (Also, the word "indices" should replace the word "index" in the next sentence.)

In addition, the fifth sentence in the sixth paragraph should read "Had that index been used to convert nominal GNP into an alternative estimate of real economic activity (an implicit quantity index rather than real GNP) then real growth in the first quarter would have been placed at 5.7 percent rather than 3.1 percent."

The author is indebted to Robert P. Parker of the Bureau of Economic Analysis for pointing out the errors in the original text. Views and opinions expressed in the text are solely those of the author and should not be attributed to any other person or institution.

Roy H. Webb