Working Paper 84-6

DEFINING AND IMPROVING THE ACCURACY OF MACROECONOMIC FORECASTS: CONTRIBUTIONS FROM A VAR MODEL

by

William M. Lupoletti and Roy H. Webb

October 1984

This is a preprint of an article in the Journal of Business ©, April 1986, v, 59, iss. 2, pp. 263-85.

The authors are indebted to Bennett McCallum 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 nor to the Federal Reserve Bank of Richmond or the Federal Reserve System.

DEFINING AND IMPROVING THE ACCURACY OF MACROECONOMIC FORECASTS: CONTRIBUTIONS FROM A VAR MODEL

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 an optimizing individual in a stochastic, dynamic environment. They instead believe that conventional models contain ad hoc representations that are manipulated until they become consistent with historic time series.

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 comprises 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. 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 inadequately 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 tradeoff between over-simplification and overparameterization: omitting a relevant variable closes a channel that might contain more signal, while including a 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 advocate using empirical evidence to judge a model's ability to produce accurate forecasts.

The model presented in this paper consists of five variables: the monetary base, real GNP, the GNP implicit price deflator, the 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 [1984a] except for this paper's use of the Treasury bill rate in place of the commercial paper rate.

The model can be expressed as

$$X = C + B(L)X + E \qquad (1)$$

where X is the vector of endogenous variables; C is a vector of constant terms; B is a polynomial in the lag operator, L, in this case representing lags one through six; and 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.

THE DATA To derive a series of post-sample forecasts from the model, the equations described by (1) were estimated from 1952 Q2 to 1969 Q4, and the estimated coefficients were used to forecast each variable in 1970 Q1. Those values, in turn, were used to produce a forecast of 1970 Q2. The procedure was repeated until we had a set of forecasts from 1970 Q1 through

1971 Q2, all based on data comparable to that possessed by a real-time forecaster in 1969 Q4. We then constructed a similar set of forecasts for each quarter through 1983 Q4. We thus produced a series of forecasts with horizons ranging from one to six 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 Q3, for real GNP, the GNP deflator, and the Treasury bill rate. Thus we have one-quarter-ahead forecasts and realized values beginning in 1970 Q4, two-quarter forecasts (expressed as average compounded annual rates) beginning in 1971 Q1, and so forth up to six-quarter forecasts beginning in 1972 Q1. The only exception is Wharton's series of interest rate forecasts, which begins in 1973 Q2. All comparisons end in 1983 Q4. In addition, we compare the VAR model's forecasts with two sets of published time series forecasts: Charles Nelson's ARIMA forecasts of GNP and the deflator beginning in 1976 Q2 and Robert Litterman's forecasts from a VAR model beginning in 1980 Q3.

In most cases, the published forecasts were released at the end of a particular quarter.[2] Accordingly, the forecasters would have possessed most of the information needed to estimate quarterly average values of the current quarter's interest rate and monetary base, typically eleven of thirteen 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.

Thus we were concerned that our procedures might exaggerate the relative accuracy of the VAR model, since we employed the latest revisions of data when estimating the model and producing post-sample forecasts. 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 our informational advantage over real-time forecasters. We then used that less accurate data to estimate the VAR model's coefficients and produce post-sample forecasts. For example, the VAR forecasts based on data ending in 1976 Q1 took their first lagged values from data that were published in April 1976. We expected that this procedure would yield uniformly less accurate forecasts than 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 some cases the model produced substantially more accurate four-quarter-ahead forecasts based on the less accurate data. Table I shows representative forecasts from the base model and several variations, including the forecasts that were based on less accurate data. The largest change illustrated was for the Treasury bill rate four quarters ahead, for which the average forecast error fell from 2.8% to 1.9%. In addition, the error for the deflator fell from 2.1% to 1.4%. On the other hand, the corresponding real GNP forecast error rose from 2.8% to 3.0%. All in all, this experiment does not reveal any gross bias favoring the VAR model due to 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, due to our suspicion that recent financial deregulation temporarily distorted the demand for M1. 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 Q2 was chosen due to 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 I contains several comparisons which 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 post-sample forecasts, however, it generally decreased forecast accuracy. Increasing the lag length to eight 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 II through VIII 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 II 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 one quarter ahead.[3] At four and six quarter horizons, however, the VAR forecast lies between the best and worst commercial forecast for all variables except real GNP. The results thus contradict the opinion expressed by Klein [1984, p. 91], "[N]one 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."

Tables III-VI 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 twenty years simply did not provide enough observation to accurately estimate that many coefficients.

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 1980-83 period. It had the most accurate forecasts for real GNP four and six quarters ahead and for nominal GNP two, four, and six 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 two 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 four- and six-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--one, and possibly two 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 only considers 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. Based on the evidence discussed above, it appears that a simple five-variable VAR model provides a useful benchmark for the evaluation of macroeconomic forecasts, particularly at the longer horizons.

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 further asserted 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 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 one to four quarters. One-quarter-ahead forecasts start in 1976 Q2; two-and four-step forecasts begin in 1976 Q3 and 1977 Q1, respectively. All forecasts end in 1982 Q4.[4]

Table VII 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 four-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.[5]

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 four-quarter-ahead forecasts and probably for longer horizons as well, the VAR technique is superior. An ARIMA model uses data more parsimoniously; on the other hand, the production of ARIMA forecasts requires specialized computer software and judgment which can only be developed 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. For this comparison, one-, two-, four-, and six-step late-quarter forecasts are used; they begin respectively in 1980 Q3, 1980 Q4, 1981 Q2, and 1981 Q4, and they all end in 1983 Q4.

It is obviously difficult to draw conclusions based on such a small sample, but Table VIII shows that Litterman's VAR model has produced excellent forecasts of real GNP. (Recall that the unconstrained 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 unconstrained five-variable VAR two and four quarters ahead. Keeping in mind that the restrictions Litterman places on his VAR model require highly specialized software, the return to this additional complexity may not justify the costs.

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 thirty 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 and the National Bureau of Economic Research (ASA/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 should decrease the error of the combined forecast.

To derive weights,[6] suppose P_1 and P_2 are time series of unbiased one-step-ahead forecasts of 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 only be determined 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 squared errors:

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

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

It should be noted that it is quite possible the relative abilities of forecasters will change over time. It is therefore an empirical question as to 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

$$A = w_0 + w_1P_1 + w_2P_2 + u \qquad (5)$$

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 multi-step forecasts. And in 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. 271-8] describe results that indicate ratio techniques such as (4) outperform more complex forms that are similar to (5).[7] For this paper, a limited set of tests was examined for particular variables and forecasters. Post-sample 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 eight quarters, a moving range of the last sixteen 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 Q3; two- and four-quarter forecasts begin in 1970 Q4 and 1971 Q2, respectively.

Table IX presents the lowest average error for each technique, pair of forecasters, and variable, computed over the period 1976 Q1 to 1983 Q4. Results are only reported 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.[8]

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 post-sample prediction. Formula (4) may outperform formula (5) simply because in-sample bias has no predictive power.[9]

On a more mundane level, formula (5) may simply be overparamaterized. There is some evidence for this speculation: the shorter boundaries (eight and sixteen 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 may be a problem in a regression like (5).

THE COMBINATION OF STRUCTURAL AND VAR FORECASTS The results presented in Table IX 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 X compares forecasts produced by commercial forecasting services (for brevity, they are referred to below as structural models), a VAR model, and the ASA/NBER median survey[10] to forecasts combined by using the ratio technique. The result most obvious from Table X is that combining structural and VAR forecasts is a useful exercise. Of the sixteen examples in the table (four variables at four horizons each), the combination of the best structural forecaster and the VAR forecast is most accurate twelve times, including ties. In the four cases where the structural-VAR combination does not produce the best forecast, it is only .1 behind the winner each time.

Of the four types of combined forecasts presented in the table, the structural-VAR team is clearly superior. Structural-VAR combinations beat structural-structural combinations twelve of sixteen times (plus one tie); they are superior to the combination of two structural forecasts and the VAR forecast in eleven instances; and they produce lower errors than the ASA/NBER median in six of nine cases. Overall, some form of combined forecast beats the best individual forecast ten out of sixteen times, while an individual forecast is never 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. The combination of the best structural forecast and the VAR forecast produces synergism in nine of sixteen cases, while structural-structural combinations are synergistic only once. The latter result suggests that there is little return to the information in a second structural forecast, given knowledge of one such forecast. The structural-structural-VAR combination produces synergism three times; this is further evidence that the return to knowledge of a second structural forecast is low.

The incremental forecast improvement from a VAR model relative to a structural model depends on the variable under consideration. As shown by the asterisks in the fifth column of Table X, adding information from a VAR forecast to a structural forecast is useful for real GNP, nominal GNP, and the Treasury bill rate, but not 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 the two types of models employ in producing forecasts of real and nominal GNP and interest rates.

Since the return to the nth similar forecast is near zero, doubts are cast upon the survey median combination technique exemplified by ASA/NBER, in which a large number of forecasters are surveyed and all are assigned identical weights. 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 through 1983. 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.

Table I

VARIATIONS IN MODEL SPECIFICATION

	Real GNP		GNP Deflator		Nominal GNP		Interest Rate	
	1Q	4Q	1Q	4Q	1Q	4Q	1Q	4Q
Reported Model	5.2	2.8	2.3	2.1	5.9	3.3	1.2	2.8
Variation 1: Less accurate 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

Data are root-mean-square errors from post-sample forecasts, measured in the same units as the variables themselves.
1Q RMSEs were calculated over the range 1970 Q1 - 1983 Q4.
4Q RMSEs were calculated over the range 1970 Q4 - 1983 Q4.
All variables except the interest rate were expressed as per cent changes at annual rates.

Table II

AVERAGE FORECAST ERRORS FROM ALL MODELS, 1970 TO 1983

		Chase	DRI	Wharton	VAR
Real GNP:	1Q	4.1	4.0	4.2	5.3
	2Q	3.1	3.1	2.9	4.1
	4Q	2.5	2.5	2.2	2.8
	6Q	2.3	2.3	1.9	2.4
GNP Deflator:	1Q	1.8	2.0	1.9	2.3
	2Q	1.9	2.0	1.9	2.0
	4Q	2.2	2.1	2.0	2.1
	6Q	2.5	2.4	2.2	2.4
Nominal GNP:	1Q	5.1	4.6	4.9	6.0
	2Q	4.1	3.6	3.8	4.7
	4Q	3.5	3.0	3.0	3.3
	6Q	3.3	2.7	2.6	3.1
T-Bill Rate:	1Q	1.5	1.4	--	1.3
	2Q	2.2	2.1	--	2.1
	4Q	2.9	2.6	--	2.8
	6Q	3.5	3.2	--	3.5

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are:
1Q: 1970 Q4 - 1983 Q4
2Q: 1971 Q1 - 1983 Q4
4Q: 1971 Q3 - 1983 Q4
6Q: 1972 Q1 - 1983 Q4.

Table III

AVERAGE FORECAST ERRORS FROM THE VAR MODEL

		full period	start-75:1	75:2-80:3	80:4-83:4	around turning points
Real GNP:	1Q	5.3	6.1	4.9	4.8	5.8
	2Q	4.1	4.6	3.7	4.1	4.3
	4Q	2.8	3.7	2.3	2.4	2.8
	6Q	2.4	2.9	2.1	2.2	2.6
GNP Deflator:	1Q	2.3	1.8	2.3	2.9	2.6
	2Q	2.0	1.7	2.0	2.3	2.3
	4Q	2.1	2.1	2.3	2.0	2.4
	6Q	2.4	2.5	2.5	2.2	2.6
Nominal GNP:	1Q	6.0	6.6	5.6	5.9	6.4
	2Q	4.7	5.2	4.3	4.8	4.7
	4Q	3.3	4.1	2.9	3.1	2.7
	6Q	3.1	3.8	2.4	3.4	2.7
T-Bill Rate:	1Q	1.3	1.0	1.0	1.9	1.6
	2Q	2.1	1.7	1.6	3.1	2.6
	4Q	2.8	2.5	1.9	4.2	3.2
	6Q	3.5	3.3	2.5	5.0	3.6

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are as in Table II.

Table IV

AVERAGE FORECAST ERRORS FROM CHASE ECONOMETRICS

		full period	start-75:1	75:2-80:3	80:4-83:4	around turning points
Real GNP:	1Q	4.1	2.8	4.3	5.0	4.6
	2Q	3.1	2.2	2.8	4.4	3.5
	4Q	2.5	2.4	2.1	3.3	3.0
	6Q	2.3	2.3	2.1	2.4	2.7
GNP Deflator:	1Q	1.8	2.2	1.5	1.8	2.0
	2Q	1.9	2.3	1.5	1.8	2.0
	4Q	2.2	3.0	1.5	2.1	2.4
	6Q	2.5	3.0	2.2	2.4	2.7
Nominal GNP:	1Q	5.1	3.1	5.6	6.2	5.5
	2Q	4.1	2.6	4.0	5.6	4.4
	4Q	3.5	1.6	3.3	5.2	3.9
	6Q	3.3	1.8	2.9	4.8	3.6
T-Bill Rate:	1Q	1.5	1.0	1.6	1.8	1.9
	2Q	2.2	1.5	1.9	3.1	2.6
	4Q	2.9	1.6	2.8	3.9	3.0
	6Q	3.5	2.1	3.4	4.5	3.2

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are as in Table II.

Table V

AVERAGE FORECAST ERRORS FROM DATA RESOURCES, INC.

		full period	start-75:1	75:2-80:3	80:4-83:4	around turning points
Real GNP:	1Q	4.0	3.5	3.9	4.6	4.5
	2Q	3.1	3.1	2.3	4.2	3.7
	4Q	2.5	2.9	1.8	3.1	3.2
	6Q	2.3	2.5	2.0	2.5	2.8
GNP Deflator:	1Q	2.0	2.5	1.7	1.9	2.2
	2Q	2.0	2.6	1.5	1.5	2.1
	4Q	2.1	3.0	1.5	1.8	2.3
	6Q	2.4	3.3	1.8	2.2	2.5
Nominal GNP:	1Q	4.6	3.2	4.8	5.7	5.1
	2Q	3.6	2.6	3.1	5.1	4.0
	4Q	3.0	2.0	2.4	4.5	3.5
	6Q	2.7	1.5	1.7	4.5	3.4
T-Bill Rate:	1Q	1.4	0.9	1.4	1.9	1.8
	2Q	2.1	1.4	1.6	3.2	2.5
	4Q	2.6	1.7	2.1	4.0	2.9
	6Q	3.2	2.0	2.6	4.7	3.3

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are as in Table II.

Table VI

AVERAGE FORECAST ERRORS FROM WHARTON ECONOMETRIC FORECASTING ASSOCIATES

		full period	start-75:1	75:2-80:3	80:4-83:4	around turning points
Real GNP:	1Q	4.2	4.0	4.5	3.8	4.8
	2Q	2.9	2.6	2.4	3.8	3.4
	4Q	2.2	2.0	1.7	3.0	2.7
	6Q	1.9	2.1	1.4	2.4	2.3
GNP Deflator:	1Q	1.9	2.0	1.8	1.9	1.8
	2Q	1.9	2.2	1.7	1.8	1.9
	4Q	2.0	2.6	1.5	1.9	2.1
	6Q	2.2	3.0	1.6	2.1	2.3
Nominal GNP:	1Q	4.9	4.1	5.2	5.3	5.5
	2Q	3.8	2.8	3.4	5.2	4.3
	4Q	3.0	1.9	2.1	4.8	3.6
	6Q	2.6	2.0	1.2	4.4	3.3
T-Bill Rate:	1Q	--	--	1.8	1.5	--
	2Q	--	--	2.0	2.7	--
	4Q	--	--	2.2	3.6	--
	6Q	--	--	2.7	4.3	--

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are as in Table II.

Table VII

COMPARISON OF VAR AND NELSON'S ARIMA FORECASTS

		VAR	NELSON'S ARIMA
Real GNP:	1Q	4.8	4.5
	2Q	3.8	3.4
	4Q	1.9	2.4
GNP Deflator:	1Q	2.3	1.8
	2Q	1.9	1.7
	4Q	1.7	1.8
Nominal GNP:	1Q	6.0	5.9
	2Q	4.6	4.7
	4Q	3.3	3.8

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are:
1Q: 1976 Q2 - 1982 Q4
2Q: 1976 Q3 - 1982 Q4
4Q: 1977 Q1 - 1982 Q4.

Table VIII

COMPARISON OF UNCONSTRAINED VAR
AND LITTERMAN'S VAR FORECASTS

		Unconstrained VAR	Litterman's VAR
Real GNP:	1Q	4.8	4.0
	2Q	4.1	3.2
	4Q	2.5	1.9
	6Q	2.7	1.1
GNP Deflator:	1Q	2.8	3.4
	2Q	2.3	3.4
	4Q	2.1	3.5
	6Q	2.5	3.8
Nominal GNP:	1Q	6.0	6.0
	2Q	4.8	5.6
	4Q	3.2	4.7
	6Q	4.0	3.6
T-Bill Rate:	1Q	1.9	1.9
	2Q	3.1	2.9
	4Q	4.6	4.1
	6Q	4.7	4.3

Data are root-mean-square errors from post-sample forecasts.
Ranges for RMSEs are:
1Q: 1980 Q3 - 1983 Q4
2Q: 1980 Q4 - 1983 Q4
4Q: 1981 Q2 - 1983 Q4
6Q: 1981 Q4 - 1983 Q4.

Table IX

TWO TECHNIQUES FOR COMBINING FORECASTS

	Nominal GNP 1Q	GNP Deflator 2Q	Real GNP 4Q
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[4]	1.7[1]	2.5[1]
Wharton+VAR	5.1[1]	1.8[1]	2.4
DRI+Wharton	5.4[3]	1.9[2]	2.4[1]
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

Data are root-mean-square errors from post-sample forecasts and combinations.
RMSEs were calculated over the period 1976 Q1 to 1983 Q4.
Estimation was performed over all available data.
The best least squares combinations were generated with OLS, except as noted:

[1] MA(1) correction employed,
[2] MA(2) correction employed,
[3] AR(1) correction employed,
[4] AR(2) correction employed.

Table X

THE COMBINATION OF STRUCTURAL AND VAR FORECASTS

		individual forecasts			combined forecasts			
		(1) best structural	(2) 2nd best structural	(3) VAR	(1)+(2)	(1)+(3)	(1)+(2)+(3)	ASA/NBER
Real GNP:	1Q	*3.9^{w}	4.2^{d}	4.7	4.0	*3.9	4.0	4.1
	2Q	*2.9^{w}	3.1^{d}	3.7	3.0	*2.9	3.0	3.2
	4Q	2.2^{d}	2.2^{w}	2.3	2.1	*1.9	2.1	2.3
	6Q	1.7^{w}	1.8^{d}	2.2	1.7	*1.6	*1.6	--
GNP Deflator:	1Q	*1.6^{c}	1.8^{d}	2.4	*1.6	1.7	1.7	1.7
	2Q	1.6^{d}	1.6^{c}	2.1	1.6	1.6	*1.5	*1.5
	4Q	*1.6^{d}	1.7^{w}	2.1	*1.6	1.7	*1.6	*1.6
	6Q	*1.8^{w}	*1.8^{d}	2.4	*1.8	1.9	1.9	--
Nominal GNP:	1Q	4.9^{w}	5.2^{d}	5.7	5.0	*4.8	5.0	5.0
	2Q	4.0^{w}	4.0^{d}	4.5	4.0	*3.7	4.1	4.2
	4Q	3.4^{d}	3.4^{w}	3.1	3.4	*3.0	3.6	3.6
	6Q	2.9^{w}	3.1^{d}	3.0	3.0	*2.8	3.2	--
T-Bill Rate:	1Q	1.7^{d}	1.7^{c}	*1.4	1.7	*1.4	1.5	--
	2Q	2.4^{d}	2.5^{c}	2.3	2.4	*2.2	2.3	--
	4Q	3.1^{d}	3.4^{c}	3.1	3.2	*3.0	3.1	--
	6Q	3.7^{d}	4.0^{c}	3.7	3.7	*3.6	3.8	--

Data are root-mean-square errors from post-sample forecasts and combinations.
RMSEs were calculated over the period 1976 Q1 to 1983 Q4.
The ratio technique was used for combinations, with estimation performed over all available data.
Letters indicate commercial forecasters:

c = Chase
d = DRI
w = Wharton.

Asterisks indicates best forecast for each variable and horizon.

FOOTNOTES

1. See Sims [1980] for a lucid critique of the conventional strategy for constructing macro-models.

2. An exception is Wharton. Prior to 1976 their forecasts 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 one-quarter-ahead forecast. On the other hand, we define a one-quarter-ahead forecast as one made at time t for an outcome at time t+1. We find it potentially confusing, for example, that in his terminology a late-quarter forecast of that quarter's average interest rate could be produced after eleven weeks of data were observed. Our practice is analogous to that of Zarnowitz [1979] with respect to annual forecasts.

3. A striking feature of the tables is that four-quarter forecasts are often substantially more accurate than one-quarter forecasts for all variables except the interest rate. That is because all multi-step growth rates in this paper are expressed as average compound growth rates. This formulation allows errors of opposite sign in successive quarters to cancel out.

4. 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 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.

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

6. 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].

7. Granger may have reconsidered, however; see Granger and Ramanathan [1984].

8. Einhorn and Hogarth [1975] have suggested that mean weighting may be preferable to any more complex method. While we agree that mean weighting can be superior to least squares, we have found it to be either no better or worse than the ratio technique.

9. For this reason, in-sample results of forecast combinations, such as those presented by Nelson [1984], are unpersuasive. An entire literature has developed that purports to disprove the rational expectations hypothesis based on similar in-sample results. Webb [1984b] demonstrates that post-sample results can produce opposite conclusions.

10. 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 Q3, and it does not release six-quarter-ahead forecasts.

REFERENCES

Einhorn, H. J., and Hogarth, R. M. 1975. Unit Weighting Schemes for Decision Making. Organizational Behavior and Human Performance 13 (April): 171-192.

Goldfeld, S. M. 1976. The Case of the Missing Money. Brookings Papers on Economic Activity: 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: 197-204.

Klein, L. R. 1984. The Importance of the Forecast. The 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.

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.

Webb, R. H. 1984a. Vector Autoregressions as a Tool for Forecast Evaluation. Federal Reserve Bank of Richmond Economic Review 70 (January/February): 3-11.

Webb, R. H. 1984b. Inadequate Tests of the Rationality of Expectations. Presented at the Western Economic Association meetings, June.

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.