The Effectiveness of Combining Forecasts: Evidence Using Macroeconomic Variables

Authors	R. W. Hafer, and Gail Heyne Hafer
Working Paper Number	1984-027A
Creation Date	January 1984
Citable Link	https://doi.org/10.20955/wp.1984.027
Suggested Citation	Hafer, R.W., Hafer, G.H., 1984; The Effectiveness of Combining Forecasts: Evidence Using Macroeconomic Variables, Federal Reserve Bank of St. Louis Working Paper 1984-027. URL https://doi.org/10.20955/wp.1984.027

THE EFFECTIVENESS OF COMBINING FORECASTS:
EVIDENCE USING MACROECONOMIC VARIABLES

Gail Heyne Hafer* and R.W. Hafer**
Federal Reserve Bank of St. Louis
84-027

*Consultant, Drazen-Brubaker and Associates, Inc.
St. Louis, Missouri.

**Senior Economist, Federal Reserve Bank of St.
Louis, Missouri.

The authors would like to thank Larry DiMariano
for his research assistance. The views expressed
are the authors' and may not reflect those held by
their employers.

INTRODUCTION

The recent wave of plant postponements and cancellations has focused increased attention on the need of utilities to prepare reliable demand forecasts for use in construction planning and the calculation of revenue requirements. Estimates of economic variables play a major role in the load forecasts produced by many utilities. As Sawhill and Silverman demonstrate, overly optimistic estimates of gross national product (GNP) growth and inflation may justify plant expansion which would not be required for several years under more conservative forecasts of GNP growth and inflation.[1/]

This article discusses a simple and inexpensive means of improving the reliability of forecasts of GNP and inflation. This improvement is obtained by combining the forecasts produced by different models. In the tests presented below, we show that forecasts of GNP and inflation obtained from large econometric models for the past few years could, in general, have been improved upon by forming a composite forecast that combines the model prediction and that from a simple time-series model. Moreover, we demonstrate that pairwise combinations of the econometric model forecasts often yield lower forecast errors relative to each of the components.

The paper is divided into four sections. The first section discusses the data used for both the econometric forecasts and the time-series forecasts. The relative reliability of the forecasts produced by the econometric and time-series models is discussed in the second section. The third section discusses the combination forecasts using the econometric forecasts and the time-series forecasts, followed by the results of combining the econometric model forecasts. Concluding remarks close the paper in the final section.

DATA

The data used in this study are one-quarter ahead forecasts for GNP growth and inflation, where inflation is measured using the GNP deflator. The forecasts are made uniformly in the first month of the quarter. Therefore, the forecast made in the first quarter is based on information available into January; the second quarter forecast uses information through April, and so on. This procedure allows the forecasters to use preliminary data for the preceding quarter, but provides no information relevant to the current period being forecast.

Two types of forecasts are used. The first type includes several econometric forecasts reported in the Conference Board's Statistical Bulletin. The sample of professional econometric forecasts used includes the Conference Board, Wharton, Data Resources Inc., Chase Econometrics and Merrill-Lynch. Forecasters are included based solely on the fact that continuous series of forecasts are available over the time period studied.

The second type of forecast is derived from a simple time-series model. Recently, Nelson has demonstrated the usefulness of Box-Jenkins type of time series models in a comparison of econometric model forecasts to those derived from a time series model.[2/] Because of data limitations, however, we could not develop a sophisticated time series model. Consequently, our time series "model" is a simple two quarter moving average of the preliminary quarterly figures. This crude model generates its forecast by using the simple average of the preliminary figures released for the preceding two quarters. The moving average forecast for the third quarter, for example, uses the average of the preliminary data reported for the first and second quarters. This approach clearly incorporates less information than the econometric models, because we do not allow the moving average to incorporate new, updated data as it is revised for every new forecast.

The forecasts are compared to the actual values of GNP growth and inflation for the period 1979/I to 1984/II, the endpoint of our sample. Moreover, we use currently available figures for the actual series to ensure a comparable basis to evaluate the alternative forecasts.

COMPARING THE ACCURACY OF ALTERNATIVE FORECASTS

To gauge the accuracy of the different forecasts, two statistics are used. One is the simple mean error, defined as

$$(1)\quad ME = \frac{1}{n} \Sigma (F-A)$$

where n is the number of quarters, F is the forecasted value and A is the observed change. The other measure is the mean square error defined as

$$(2)\quad MSE = \frac{1}{n} \Sigma (F-A)^2.$$

The mean error measures the bias associated with the forecasts. The mean square error captures both the degree of bias and the variance of the forecasts error series.[3] When attempting to maximize the precision of forecasts, analysts may choose a method with a larger bias component but low variance.

The first set of comparisons are presented in Table 1. There we report the mean error and the mean square error for each econometric forecast and our moving average forecast for the period 1979/I to 1984/II. Columns (1) and (2) present the outcome from forecasting the growth of GNP. The negative mean errors in column (1) indicate that each forecast, on average, underpredicted the changes in GNP growth. The mean errors range from -0.39 percent for Wharton to -1.54 percent for Merrill-Lynch. Interestingly, the simple two-quarter moving average model ranked second best with an average error of -0.42 percent. When GNP forecasts are compared using a mean squared error criterion in column (2), the rankings change dramatically. We find that the DRI forecasts minimize the square errors, followed by those of Wharton.

The results of forecasting inflation during the past five years are reported in columns (3) and (4). On average, each forecasting method underpredicted the rate of inflation over the period. The Conference Board forecasts with a mean error of -0.34 percent were, on average, the closest to the actual. The usefulness of the simple moving-average model again is apparent: its average inflation forecast error of -0.39

percent is next to the lowest. Relative to the average error recorded by Wharton (-0.68 percent), the moving average model represents over a 40 percent reduction in the mean forecast error.

When we use the mean square error comparisons, shown in column (4), again the rankings change dramatically. Now Chase provides the lowest MSE of 2.13 percent and Wharton yields the highest mean square error of 2.96 percent. Based on this forecast criterion, the moving average model does better than that of Merrill-Lynch, reducing the mean square error by 25 percent (4.16 to 3.10).

COMPOSITE FORECASTS: ECONOMETRIC AND TIME SERIES

The comparison of different forecasters' track record of predicting GNP growth and inflation since 1979 reveals that there is a wide divergence in outcomes. Because the consumer of such forecasting services wants the best (i.e., the most consistently correct) forecast the most number of times, one is tempted to disregard the forecasts of all except the proven.4/ Because the "best" forecast may include a mix of several different predictions, each based on a different information base or outlook on economic activity, rather than limit oneself to a single prediction, a useful procedure is to combine forecasts.5/

Forecasts can be combined in a variety of ways. The simplest technique of course is to sum the forecasts and take the average. It has been shown, however, that a superior method of combining forecasts is to use a weighted average, where the weights are determined by estimating the regression:

$$(3) \quad A_t = \beta_1 F(1)_t + \beta_2 F(2)_t + \varepsilon_t$$

where A_t is the observed value being forecast, F(1) and F(2) are two different forecasts of A, and ε_t is a random error term. The least squares estimates of β_1 and β_2 provide the relative weights for the F(1) and F(2) predictions, respectively. This weighting procedure yields an optimal linear combination of the forecasts.

It is possible to estimate equation (3) for various combinations of forecasts over the 1979/I to 1984/II sample and to compare the mean square error derived from equation (3) to those reported in Table 1. A more instructive approach, however, is to estimate equation (3) for a subset of our total period and compare the respective out-of-sample predictions of the combined forecasts to those in Table 1.[6/] This procedure was done by using pairwise combinations of the econometric and the simple moving average model forecasts of GNP growth and inflation for the period 1979/I to 1982/IV, a total of twelve observations. Using the estimated β_i weights, post-sample combined forecasts were made for the 1983/I to 1984/II period. Although data limitations restrict the size of our estimation and forecast periods, the results are informative.

The results of combining the econometric forecasts and the moving average model for GNP growth are found in the upper panel of table 2. The reported estimates of β_1 and β_2 are interesting, because we find that none of the coefficients are significant at the five percent level of significance. This result warns of the possibility of multicollinearity in the forecasts. In other words, the finding is indicative of a situation in which the forecasts provide similar information. This also is suggested by the result that, for the combination of Chase/MA and Merrill-Lynch/MA, both forecasts receive almost equal weighting.

The out-of-sample GNP forecast statistics based on the least-squares weights are reported in columns (3) and (4) of the upper panel of Table 2. Because the forecast period encompasses only the 1983/I to 1984/II period, we report in parentheses the econometric model forecast statistics for this period to allow direct comparison with the outcome of the combination forecasts.

Looking first at the mean error results, we see that the linear combination forecasts have reduced the average errors relative to the econometric models in four out of five instances, ranging from a 22 percent reduction over the Merrill-Lynch predictions to 35 percent over those of the Conference Board. In only one instance, the Wharton model forecasts, do we find that the predictions from the combined model are less reliable as measured by the mean error than the econometric model.[7/]

Comparing the different GNP forecasts using the mean square error criterion shows an improvement in three out of five cases. Although the degree of improvement is not as dramatic, the mean square error is lowered using the weighted predictions by 6 percent to 16 percent. Recalling that all we have added to the econometric models' forecasts is information from a crude moving average model, this improvement clearly suggests that combining forecasts may prove valuable.

The lower panel of Table 2 presents the outcomes for the inflation forecasts. The reported weights indicate that in three out of five cases, the information contained in the moving average does not add greatly to that already captured by the econometric models. There are, however, two exceptions: One is the Wharton/MA combination. There we see that the two forecasts have similar information, as evidenced by the insignificance of both β's and the similarity of the estimated weights. The other exception is the Merrill-Lynch/MA combination. Surprisingly, the moving average forecast receives a much larger weight than the econometric prediction (0.77 vs. 0.[illegible]), and the estimated coefficient on the econometric forecast is not statistically significant at the 5 percent level. This result suggests that the simple moving average model out-performed the Merrill-Lynch forecasts.

Comparisons of the out-of-sample inflation forecasts again reveal the superiority of the combined forecasts: the mean error is lower in each of the five cases, ranging from a 2 percent reduction relative to the Conference Board to a 50 percent reduction compared to the Merrill-Lynch prediction. The mean square error results also indicate a general improvement in forecast accuracy. There we see that in four out of five comparisons, the mean square error of the combined forecast was lower than the mean square error of the econometric model. Again the range of improvement is wide: from a 2 percent improvement relative to the Conference Board's predictions to a 44 percent improvement over that of Merrill-Lynch.

Overall, the evidence presented in Table 2 indicates that combined forecasts provide an improvement over single model predictions. In our comparison, adding the information contained in a crude moving average time series model to that underlying large scale econometric models often yielded substantial reductions in forecast errors. Indeed, in terms of an mean error criterion, the combined forecasts improved upon the econometric forecasts in 90 percent of the cases reported. Using the mean square error forecasting criteria, addition of the moving average forecast to those derived solely from the econometric models improved forecast accuracy 70 percent of the time. By either standard of judging forecast performance, the results in Table 2 clearly indicate the usefulness of combining forecasts.

COMPOSITE FORECASTS: PAIRWISE ECONOMETRIC

The evidence in the preceding section showed that econometric forecasts of GNP growth and inflation during the past six quarters could have been improved by incorporating the information contained in a simple moving average model. The question addressed now concerns the relative improvement over individual econometric forecasts that may be obtained by using combinations of the different econometric models.

To investigate this question, ten pairwise combinations of the econometric forecasts were used. As before, the combinations were used in estimating equation (3) for the 1979/I - 1982/IV period to obtain the relative weights. Out-of-sample forecasts for the six quarters from 1983/I to 1984/II were generated using estimated linear regression weights. The combinations, their relative weights and forecast performances are reported in Table 3.

The upper panel in Table 3 reports the outcome of using the various combinations of econometric predictions to forecast GNP growth. The least squares estimates indicate that, in general, the econometric predictions contain similar information. In only three instances, the combinations of Wharton/Chase, DRI/Chase and DRI/Merrill-Lynch, does one of the predictions achieve statistical significance. It also is interesting to note that in several instances one of the weights actually is negative.

To see how the paired econometric forecasts do over the last six quarters in predicting GNP growth, the mean error and mean square error statistics are reported. In order to make comparisons with the single-model forecasts, a plus (+) is used to denote those composite forecasts that are better than both of the component models. Of the 10 pairwise combinations, we find that only two combinations yield a lower average forecast error than both of the relevant components. In fact, in two instances (Wharton/Chase and Wharton/Merrill-Lynch) the mean error statistics are larger than either component model's mean error for the same period. Looking at the mean square error outcome, four of the paired econometric forecasts improve upon the individual predictions. The lowest mean square error occurs with the Conference Board/DRI combination, followed by the Chase/Merrill-Lynch tandum. Thus, in terms of predicting GNP growth, the weighted combination of econometric forecasts yielded an improvement over both component forecasts in 40 percent of the cases.

The results of combining the econometric forecasts of inflation are found in the lower panel of Table 3. The relative weights derived from estimating equation (3) reveal a wide variety of relative information sets. In five of the pairs, one of the predictors achieves statistical significance. Unlike the post-sample performance of the composite GNP forecasts, the summary forecast error statistics indicate that, in six out of ten instances, the combined forecast would have lowered the average inflation forecast error relative to either of the component econometric predictions. The greatest reduction comes from the weighted combination of the Wharton and Merrill-Lynch predictions. The mean error of the composite forecast (-0.07 percent) is 110 percent lower than that produced by the Wharton forecast and 105 percent below that reported by Merrill-Lynch. Similarly, the composite DRI/Merrill-Lynch mean forecast error is about 90 percent lower than each of the individual econometric prediction.

In terms of the mean square error criterion, dramatic reductions in forecast errors also are observed. Over all, seven of the ten composite forecast models out-perform their component parts. For example, the mean square error of 0.18 percent reported for the Chase/Merrill-Lynch combination is 60 percent lower than that of the Chase forecast and 94 percent lower than the Merrill-Lynch performance. An almost equally dramatic improvement in forecast performance is obtained with the combination of DRI and Merrill-Lynch forecasts. There we find that the composite mean square error is 72 percent below the mean square error for DRI and 92 percent lower than that of Merrill-Lynch.

The evidence presented in Table 3 again demonstrates the general usefulness of combining alternative forecasts by a relatively simple least squares weighting scheme. With regard to predictions of GNP growth during the 1983/I to 1984/II sample, the composite forecast was shown to be superior to both the predictions 40 percent of the time. When the inflation forecasts are examined, this improvement increases to 70 percent.

CONCLUSION

We have demonstrated that substantial gains in reducing forecast errors can accrue through the use of linear combinations of forecasts. Two alternative combinations were examined: First, we combined forecasts produced by econometric models with a simple moving average forecast. Results indicate that the combination forecasts resulted in improved forecasts of GNP in at least 60% of the cases and improved forecasts of inflation in at least 80% of the cases.

In the second set of exercises, we used pairwise combination forecasts from the econometric models. The combined forecasts improved on each component's predictive power in 40% of the cases for GNP and 60% of the cases for inflation. Overall, the combination forecasts represent a simple and inexpensive means of reducing the bias associated with forecasts of economic variables.

FOOTNOTES

1/ John C. Sawhill and Lester P. Silverman, "Do Utilities Have Strategic Options? Ask the Customer," Public Utilities Fortnightly (March 31, 1983), 13-17.

2/ See Charles R. Nelson, "A Benchmark for the Accuracy of Econometric Forecasts of GNP," Business Economics (April 1984), 52-58.

3/ The mean square error criterion is often used, based on the assumption that the cost of making a forecast error is proportional to the squared error. In other words, the goal in improving forecast accuracy is based on the minimization of a squared error cost function. For a useful discussion, see C.W.J. Granger, Forecasting in Business and Economics (Academic Press, 1980), Chapter 1.

4/ An interesting retrospect on forecasting and the role of the forecaster is presented in Lawrence R. Klein, "The Importance of the Forecast," Journal of Forecasting (January-March, 1984), 1-9.

5/ Useful references on combining forecasts are J.M. Bates and C.W.J. Granger, "The Combination of Forecasts," Operations Research Quarterly 20 (1969), 451-468; Charles R. Nelson, "The Predictive Performance of the F.R.B.-M.I.T.-PENN Model of the U.S. Economy," American Economic Review (December 1972), 902-917; and Granger (1980), Chapter 8.

6/ Generally, such forecast combinations are compared using in-sample estimates instead of post-sample forecasts. The comparison used here is more useful, because it is the future that the forecaster must predict. Consequently, generating "optimal" forecasts of the past is moot. For another application of this procedure, see C.W.J. Granger and R. Ramanathan, "Improved Methods of Combining Forecasts," Journal of Forecasting (April-June, 1984), 197-204.

7/ The result that the combination forecast is worse than the components is surprising. It should be noted, however that we are comparing post-sample forecasts, as opposed to the usual in-sample comparison. What our result suggests is that the estimated weights may have changed over time. Thus, a useful extension to the procedures described here may be to continually update the weighting structure as data becomes available.

Table 1
Econometric Forecast Results: 1979/I-1984/II

Forecaster	GNP Mean Error (1)	GNP Mean Square Error (2)	INFLATION Mean Error (3)	INFLATION Mean Square Error (4)
Conference Board	-1.25%	24.21%	-0.34%	2.43%
Wharton	-0.39	22.47	-0.68	2.96
DRI	-1.00	20.79	-0.64	2.19
Chase	-1.93	29.05	-0.44	2.13
Merrill-Lynch	-1.54	28.09	-0.58	4.16
MA	-0.42	28.62	-0.39	3.10

Table 2
Composite Forecasts: Econometric and Moving Average
Forecast Period: 1983/I-1984/II

	Combination	Weight: Econometric (1)	Weight: MA (2)	Mean Error (3)	Mean Square Error (4)
Panel A: GNP	Conference Board/MA	0.662	0.423	-0.13%(-0.20)	7.06%(7.52)
	Wharton/MA	0.652	0.341	-1.56 (-1.10)	8.62 (6.72)
	DRI/MA	0.846	0.211	-1.34 (-1.77)	6.80 (7.43)
	Chase/MA	0.544	0.533	-1.56 (-2.25)	9.02(10.74)
	Merrill-Lynch/MA	0.536	0.546	-0.97 (-1.24)	6.51 (5.95)
Panel B: Inflation	Conference Board/MA	0.958*	0.040	0.91 (0.93)	1.98 (2.01)
	Wharton/MA	0.541	0.406	0.44 (0.67)	0.88 (1.06)
	DRI/MA	0.715*	0.234	0.55 (0.82)	0.65 (0.90)
	Chase/MA	0.726*	0.239	0.42 (0.53)	0.52 (0.45)
	Merrill-Lynch/MA	0.201	0.769*	0.69 (1.37)	1.82 (3.26)

Table 3
Composite Forecasts: Pairwise Econometric
Forecast Period: 1983/I-1984/II

	Combination	Weight Model 1 (1)	Weight Model 2 (2)	Mean Error (3)	Mean Square Error (4)
Panel A: GNP	Conference Board/Wharton	0.443	0.615	-0.16%	6.37%+
	Conference Board/DRI	0.168	0.916	-0.71	5.33+
	Conference Board/Chase	0.846	0.307	0.87	8.42
	Conference Board/Merrill-Lynch	0.869	0.276	1.12	8.54
	Wharton/DRI	-0.102	1.174	-1.16	5.74+
	Wharton/Chase	2.896*	-2.413	-3.57	20.96
	Wharton/Merrill-Lynch	1.379	-0.497	-2.23	11.31
	DRI/Chase	1.748*	-0.821	-2.06	8.14
	DRI/Merrill-Lynch	1.323*	-0.307	-1.78	8.00
	Chase/Merrill-Lynch	0.718	0.462	-0.16+	5.66+
Panel B: Inflation	Conference Board/Wharton	1.110	-0.104	0.99	2.27
	Conference Board/DRI	0.417	0.549	0.71	0.88
	Conference Board/Chase	0.519	0.464	0.67	0.96
	Conference Board/Merrill-Lynch	1.310*	-0.307	0.81+	2.12
	Wharton/DRI	0.080	0.857*	0.51+	0.51+
	Wharton/Chase	0.109	0.847	0.36	0.33+
	Wharton/Merrill-Lynch	1.365*	-0.459	-0.07+	0.81+
	DRI/Chase	0.599	0.348	0.47+	0.42+
	DRI/Merrill-Lynch	1.461*	-0.553	0.09+	0.25+
	Chase/Merrill-Lynch	1.125*	-0.171	0.19+	0.18+

NOTES: (*) denotes significance at five percent level. (+) denotes forecast that improves on each component's predictive performance.