Journal of Business & Economic Statistics, January 1986, Vol. 4, No. 1

Combining Economic Forecasts

Robert T. Clemen
College of Business Administration, University of Oregon, Eugene, OR 97403

Robert L. Winkler
Fuqua School of Business, Duke University, Durham, NC 27706

A method for combining forecasts may or may not account for dependence and differing precision among forecasts. In this article we test a variety of such methods in the context of combining forecasts of GNP from four major econometric models. The methods include one in which forecasting errors are jointly normally distributed and several variants of this model as well as some simpler procedures and a Bayesian approach with a prior distribution based on exchangeability of forecasters. The results indicate that a simple average, the normal model with an independence assumption, and the Bayesian model perform better than the other approaches that are studied here.

KEY WORDS: Bayesian combination of forecasts; Average of forecasts; Weighted average of forecasts; Exchangeable forecasters; Econometric models; Forecast evaluation.

1. INTRODUCTION

Decision makers who have to make an assessment about some uncertain future event often seek information in the form of forecasts concerning the event. For example, many individual economists and various econometric models produce forecasts of GNP and other economic variables. When the forecasts diverge, the decision maker who wants a single forecast faces the thorny problem of combining the forecasts. In this article we focus on the combination of GNP forecasts from four major econometric models.

A simple procedure for combining forecasts is to take an arithmetic average of the forecasts. This procedure serves as a useful benchmark and has been shown to perform better than some schemes that are more complicated (Makridakis and Winkler 1983).

A model in which forecasting errors are jointly normally distributed takes into account dependence and differing precision among forecasts (Newbold and Granger 1974; Winkler 1981; Winkler and Makridakis 1983). We take this model as our point of departure from the simple average, and we consider several variants on the model as well as a Bayesian approach with a prior distribution based on exchangeability. The results indicate that the simple average, the normal model with an independence assumption, and the Bayesian model perform better than the other approaches that are studied here.

Methods for combining forecasts are discussed in Section 2, and the empirical analysis involving GNP forecasts is presented in Section 3. Section 4 summarizes the article and discusses some implications of our results.

2. METHODS FOR COMBINING FORECASTS

Let x represent the variable being forecast, and suppose we have k forecasts $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k$ of x. Then the simple average yields a forecast y_1 of the form

$$y_1 = \sum_{i=1}^{k} \hat{x}_i/k. \tag{1}$$

No information about the precision of the forecasts or about dependence among the forecasts is needed to generate a simple average. The fact that each forecast receives the same weight $1/k$, however, does imply that the forecasts are being treated as if they were exchangeable.

An alternative to an ad hoc method such as a simple average is an approach that models the precision of the forecasting methods as well as their statistical interaction. This is in the spirit of a Bayesian approach to modeling information from experts or other sources (e.g., Morris 1977). One such model (Newbold and Granger 1974; Winkler 1981) treats the vector of forecast errors $e = (\hat{x}_1 - x, \hat{x}_2 - x, \ldots, \hat{x}_k - x)'$ as normally distributed with zero mean vector and positive definite covariance matrix Σ, where a prime denotes transposition. The motivation for the normal model is that there are many sources of error in forecasting and that as a result the "normal theory of errors" seems applicable. In addition, the normal distribution provides a natural model for positive dependence among experts as a result of shared information (Winkler 1981; Clemen 1984).

The combined forecast generated by the normal model is

$$y_2 = u'\hat{\Sigma}^{-1}\hat{x}/u'\hat{\Sigma}^{-1}u, \tag{2}$$

where $u = (1, 1, \ldots, 1)'$, $\hat{x} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k)'$, and $\hat{\Sigma}$ is an estimate of Σ. The combined forecast y_2 is a weighted average of the individual forecasts x_i, with the vector of weights

$$w' = (w_1, \ldots, w_k) = u'\hat{\Sigma}^{-1}/u'\hat{\Sigma}^{-1}u \tag{3}$$

depending on $\hat{\Sigma}$. Thus, the estimate $\hat{\Sigma}$, which can be determined from past data on estimation errors, prior information, or some combination thereof, plays an important role in the determination of the combined forecast.

A common choice for $\hat{\Sigma}$ is a sample covariance matrix that can be justified from the standpoint of obtaining a variance-minimizing combination of forecasts (Halperin 1961) or from a Bayesian approach using a diffuse prior distribution (Geisser 1965). A potential problem in using the sample covariance matrix to estimate Σ is that the estimate can be quite unstable unless large data sets are available for estimation. This problem is particularly pronounced when the pairwise correlations are high, as often seems to be the case with economic forecasts (e.g., see Section 3 of this article or Figlewski and Urich 1983).

We take the normal model with the sample covariance matrix as our point of departure. The weights assigned to the forecasts and thus the forecasts themselves can be modified by changing the estimate $\hat{\Sigma}$ or some other aspect of the procedure. We consider several alternative schemes, including some suggested by Newbold and Granger (1974), some considered by Granger and Ramanathan (1984), and some proposed here and not previously studied.

The simplest modification, if it might be called that, is to vary n, the number of data points used in the estimation of Σ and therefore in the determination of the weights. Using more data points should lead to better, more stable estimates and weights, provided that the process generating vectors of forecast errors is stationary. In the presence of nonstationarity, however, a smaller n might be preferable because "older" observations may have been generated by a process with different parameter values.

Another way to deal with potential nonstationarity is to weight recent experience more heavily. For example, if the last n data points are used to estimate Σ, then we let

$$(\hat{\Sigma})_{ij} = \sum_{t=1}^{n} \beta^t e_{it}e_{jt} / \sum_{t=1}^{n} \beta^t, \tag{4}$$

where $\beta \geq 1$ is a smoothing parameter, e_{it} and e_{jt} are the errors for forecasts i and j for observation t, and observation n is the most recent observation. Here each successively "older" observation receives less weight, and $\hat{\Sigma}$ can adapt more quickly over time if the process is nonstationary. The case of $\beta = 1$ corresponds to our basic normal model with equal weight given to all observations.

As noted above, estimates of Σ are often quite unstable when the correlations among forecast errors are high. The weights w_i and hence the combined forecast are extremely sensitive to small changes in the correlations, which may be poorly estimated because of a relatively small sample or nonstationarity. One way to avoid this instability is to treat the forecast errors as independent. To avoid complications caused by dependence, we can assume that Σ is diagonal and set all off-diagonal elements of $\hat{\Sigma}$ to zero. This approach assigns weights to forecasts only on the basis of precision and ignores correlations.

If the correlations among forecast errors are not ignored, forecasts may be assigned negative weights in some cases, and the combined forecast may lie outside the range of the k forecasts. This might be considered unreasonable by some decision makers, and it can be avoided by constraining the combined forecast to fall between the lowest and highest forecasts, inclusive. A combined forecast lower (higher) than the lowest (highest) forecast can be modified by setting it equal to the lowest (highest) forecast. This approach represents a compromise of sorts in the sense that it does not totally ignore dependence, but it does keep the effect of dependence from being too extreme. Essentially, it uses the basic normal model but monitors the "reasonableness" of the combined forecasts by imposing a convexity constraint (forcing the combined forecast to lie in the convex set generated by the k forecasts).

Estimating the weights in the normal model can be thought of as estimating regression coefficients that must sum to one. One obvious generalization is to relax the assumption that the weights must sum to one. For example, we can use an unconstrained application of ordinary least squares (OLS) to a regression model without an intercept term. If the forecasts are perfectly calibrated in the sense that the forecast errors have a zero mean vector, as assumed in the normal model, then the regression surface must pass through the origin. To allow for the possibility that this assumption is not appropriate, we can use OLS on a regression model with an intercept term. Estimation via OLS without an intercept term provides more flexibility than the basic normal model, and OLS with an intercept term yields even more flexibility.

A more extreme approach to the combination of forecasts is to assign all of the weight to a single forecast. For example, we might feel that good performance is likely to continue, which suggests that all of the weight should be given to the econometric model with the smallest error on the previous forecast. Or, taking a contrary view, we might give all of the weight to the model with the *largest* error on the previous forecast, hoping that the poor performance will motivate adjustments to improve the next forecast. When ties occur, the weight can be divided equally among those tied for best or worst.

Modifications such as constraining the combined forecasts to be within the range of the individual forecasts, totally ignoring correlations, applying OLS with or without an intercept term, and using the model that was best or worst on the previous forecast are ad hoc in nature, of course. A more formal approach can be developed in a Bayesian context, using prior information about Σ as well as past data. Suppose that the prior information concerning Σ can be represented by an inverted Wishart density with covariance matrix Σ_0 and α degrees of freedom, where $\alpha > k - 1$. Furthermore, the past data consist of n vectors of forecast errors $e_1, \ldots, e_n$, where $e_i = (e_{i1}, e_{i1}, \ldots, e_{ik})'$. Then the posterior distribution for Σ is inverted Wishart with covariance matrix

$$\Sigma^* = [(\alpha\Sigma_0^{-1} + n\hat{\Sigma}^{-1})/(\alpha + n)]^{-1} \tag{5}$$

and degrees of freedom

$$\alpha^* = \alpha + n, \tag{6}$$

where $\hat{\Sigma}$ is the sample covariance matrix (LaValle 1970).

Having revised the distribution for Σ, we now return to the combination of forecasts. Suppose that x and Σ are independent a priori, with an improper diffuse density for x and an inverted Wishart density for Σ having covariance matrix Σ^* and α^* degrees of freedom. Then the posterior mean for x, after the vector of forecasts $\hat{x}$ is seen, is

$$y_3 = u'\Sigma^{*-1}\hat{x}/u'\Sigma^{*-1}u \quad (7)$$

(Winkler 1981). Thus, the forecast from this Bayesian model is of the same form as (2) with Σ^* used in place of $\hat{\Sigma}$.

The values chosen for α and Σ_0 have implications for the combined forecast of the Bayesian model. It may be reasonable to specify a priori that we believe that forecasts within a given class (e.g., generated by econometric models or generated judgmentally by economists) are exchangeable. The prior estimate Σ_0 of Σ would thus be an intraclass correlation matrix with $\sigma^2_i = \sigma^2$ and $\sigma_{ij} = \sigma^2\rho$ for all $i \neq j$. The effect in the application of (5) would be to generate a covariance matrix Σ^* for which any differences among forecasters in variances or pairwise covariances are reduced. The degree of the shift depends on α, which can be thought of as an equivalent prior sample size. As we place more weight on the prior information by increasing α, the weights w_i become more nearly equal. In the limit, the prior information completely dominates the sample information, and the combined forecast is a simple average of the individual forecasts. As noted earlier, combining forecasts by means of a simple average implies a (usually tacit) assumption that the forecasters are exchangeable. In a different context, Lindley and Smith (1972) show that ridge regression is formally equivalent to a Bayesian model with a prior belief of exchangeability and note that such a belief might not be reasonable in many situations. Our Bayesian model is analogous to ridge regression, but exchangeability may often be a natural prior assumption in the setting of combining forecasts.

3. THE COMBINATION OF GNP FORECASTS FROM FOUR MAJOR ECONOMETRIC MODELS

Wharton Econometrics (Wharton), Chase Econometrics (Chase), Data Resources, Inc. (DRI), and the Bureau of Economic Analysis (BEA) make quarterly forecasts of many economic variables. We used their level forecasts of real and nominal GNP (1970–1982) (obtained directly from Wharton and BEA and from the *Statistical Bulletin* published by the Conference Board for Chase and DRI) to construct growth rate forecasts (in percentage terms), and we calculated the deviations from actual growth as determined from GNP reported in *Business Conditions Digest*. For both variables, forecasts with four different horizons (1, 2, 3, and 4 quarters) were analyzed. For example, the one-quarter nominal GNP forecast predicts the percentage change (annual rate) of nominal GNP over the next quarter, whereas the four-quarter forecast predicts the percentage change for the three-month period four quarters in the future.

BEA makes only one set of forecasts per quarter, usually early in the quarter, based in part on the first report of GNP for the previous quarter. On the other hand, Wharton, Chase, and DRI update their forecasts each month. We attempted to assemble forecasts from them that were comparable to BEA's in timing and data used in making the forecast.

The data covered the 1971–1982 period. The number of observations available for each of the two variables was 45 for four-quarter forecasts and 46 for the remaining forecasts. Each observation consisted of four forecasts and the actual value.

For the basic normal model, we used an adaptive method to calculate the sample covariance matrix $\hat{\Sigma}$. For any given forecasting situation, $\hat{\Sigma}$ was based on the preceding n observations and the weights assigned to the individual forecasts therefore changed in accordance with this moving window. We tried values of 5, 10, 15, and 20 for n. Unless otherwise stated, the reported results for the normal model and its variants are based on the case of $n = 20$. In the model with recent experience weighted more heavily, we considered $\beta = 1(.1)2$. For the Bayesian model with the prior distribution of Σ based on exchangeability, we used a pooled estimate of σ^2, chose $\rho = .7$, and tried several values of α. Finally, the Bayesian model was also tried with independence imposed on both the prior estimate Σ_0 and the sample estimate $\hat{\Sigma}$.

The calibration of the forecasts from the four econometric models is of interest because the normal model (as well as its variants) assumes that the forecasts are calibrated in the sense that they have expected errors of zero. The average errors presented in Table 1 reveal a slight tendency to overforecast real GNP for the longer time horizons. Of the 32 average errors (2 variables, 4 horizons, 4 econometric models), however, only four were larger than 1% in absolute value, and not one was significantly different from zero at the .05 level (although $t = 2.01$ for DRI for four-quarter forecasts

Table 1. Average Errors for Individual Forecasts

Variable	Horizon	Wharton	Chase	DRI	BEA
Nominal GNP	1	−.03 (−.05)	−1.05 (−1.87)	−.80 (−1.40)	−.63 (−1.21)
	2	.08 (.11)	−.70 (−.91)	−.36 (−.55)	.02 (.03)
	3	.03 (.04)	−.26 (−.29)	.02 (.03)	−.03 (−.04)
	4	−.20 (−.22)	−.15 (−.16)	.44 (.52)	.00 (.00)
Real GNP	1	.04 (.09)	−.77 (−1.65)	−.29 (−.53)	.10 (.22)
	2	.86 (1.51)	−.02 (−.03)	.44 (.75)	.66 (1.18)
	3	.98 (1.60)	.57 (.83)	.90 (1.40)	.97 (1.46)
	4	1.16 (1.70)	.97 (1.29)	1.39 (2.01)	1.22 (1.68)

NOTE: t statistics are in parentheses.

Table 2. Autocorrelations of Forecast Errors for Individual Forecasts

Variable	Horizon	Lag	Wharton	Chase	DRI	BEA
Nominal GNP	1	1	−.11	−.09	−.32*	−.17
		2	.08	.22	.10	−.26
		3	−.19	−.14	−.18	.06
	2	1	.05	.18	.10	.01
		2	.08	.22	.12	.15
		3	−.01	.05	−.04	−.04
	3	1	.13	.27	.30*	.25
		2	.14	.19	.17	.05
		3	−.06	.08	.00	−.04
	4	1	.10	.26	.27	.14
		2	.01	.17	.11	.11
		3	−.17	.06	−.06	−.02
Real GNP	1	1	−.19	−.03	−.16	−.11
		2	.06	.20	.01	−.16
		3	−.20	−.07	−.10	.05
	2	1	−.15	.06	.07	−.07
		2	−.05	.11	.06	.00
		3	.00	.05	.06	.00
	3	1	−.01	.26	.23	.18
		2	−.04	.03	.01	.01
		3	−.09	.03	−.03	−.11
	4	1	.14	.22	.22	.12
		2	−.12	.05	−.06	−.04
		3	−.21	−.01	−.17	−.16

* Significant at the .05 level.

of real GNP was barely nonsignificant). In addition, after completing the analysis described below, we calibrated the forecasts by adjusting for their average error and then repeated the analysis with the calibrated forecasts. The combined forecasts based on the calibrated forecasts performed slightly worse than those based on the raw forecasts. Thus, it appears that these four econometric models do a relatively good job in terms of calibration, which is what we would expect of premier econometric models.

The normal model also assumes that all forecasters have errors that are free of autocorrelation. From the autocorrelations given in Table 2 and visual inspections of residual patterns, this appears to be a reasonable assumption. Most of the autocorrelations were very close to zero. There was a slight tendency toward positive (but small) first-order autocorrelations for three- and four-quarter-ahead forecasts and negative (but even smaller) first-order autocorrelations for one-quarter-ahead forecasts.

Table 3. MAD and MSE for Individual Forecasts and Combined Forecasts

Variable	Horizon	Wharton	Chase	DRI	BEA	Simple Average	Normal Model	Independence Model	Convexity Constraint	OLS, No Intercept	OLS	Best	Worst
MAD													
Nominal GNP	1	3.68	3.77	3.84	3.75	3.55	4.23	3.57	3.98	4.38	4.51	3.95	3.58
	2	4.48	5.24	4.30	4.64	4.56	4.65	4.54	4.65	4.67	4.95	4.35	4.90
	3	5.27	6.30	5.28	5.10	5.32	5.32	5.26	4.98	5.20	5.30	5.20	5.74
	4	5.65	5.79	5.32	5.46	5.45	6.56	5.57	6.19	6.41	5.57	5.38	5.18
Real GNP	1	2.86	2.95	3.12	2.91	2.88	3.14	2.85	3.03	3.26	3.52	3.35	2.88
	2	3.45	3.89	3.52	3.62	3.54	3.66	3.54	3.71	3.34	3.38	3.36	3.58
	3	3.59	4.17	3.60	3.70	3.59	4.13	3.58	3.79	3.86	3.87	3.67	4.03
	4	3.53	3.96	3.63	4.06	3.59	4.19	3.63	3.86	4.38	4.41	3.98	3.64
MSE													
Nominal GNP	1	18.94	21.57	19.63	20.41	18.17	25.28	18.37	22.07	27.03	30.65	21.24	19.48
	2	29.79	42.26	29.06	34.78	31.94	34.64	31.77	34.87	35.84	42.54	30.49	38.73
	3	44.56	55.83	40.44	40.82	43.76	42.19	43.10	40.48	41.90	42.63	40.92	48.82
	4	63.89	53.55	44.67	45.48	47.91	76.54	50.94	69.57	55.51	45.04	49.86	44.75
Real GNP	1	11.63	12.53	14.29	12.36	11.25	13.74	11.15	13.08	14.97	18.73	16.38	12.17
	2	17.92	21.98	17.12	18.88	17.77	19.59	17.78	19.64	18.30	21.38	17.06	18.85
	3	22.01	25.99	20.09	24.50	21.85	27.09	21.88	24.94	25.06	26.54	23.98	25.50
	4	23.59	26.43	20.94	25.82	22.30	30.99	23.36	25.48	27.85	28.63	22.68	22.72

Results for the individual econometric models and many of the combining methods are summarized in Table 3. Mean absolute deviations (MAD's) and mean square errors (MSE's) are given for the eight variable/horizon combinations. In all cases where fitting was involved, the evaluation was based entirely on forecasts not used in the fitting process. To include the fitting data would give the more complex methods an unfair advantage. Facilitating comparisons further, we calculated all values of MAD and MSE for exactly the same set of forecast occasions. That is, when some occasions were used to fit a particular combining method, these occasions were excluded in the evaluation of *all* methods.

The intent of this article is to study combined forecasts rather than individual forecasts. We note, however, that Wharton and DRI tended to have lower values of MAD and MSE than the other two models and that Chase had the worst overall performance of the four. The four econometric models demonstrated greater accuracy when forecasting real GNP than when forecasting nominal GNP, with the discrepancy increasing for the longer horizons. As anticipated, the forecast accuracy generally decreased as the lead time increased.

The simple average performed quite well in this study, as can be seen from Table 3. If the "best" individual model could be identified on each forecast occasion, it would outperform the simple average. Since such identification may not always be feasible, the simple average provides an alternative that yields good performance and is very robust (Makridakis and Winkler 1983). The values of MAD and MSE for the simple average were roughly comparable on an overall basis to those for the better individual models.

In contrast to the simple average, our basic normal model fared poorly. For every other variable/horizon combination except three-quarter-ahead forecasts of nominal GNP, the normal model had a higher MAD and a higher MSE than the simple average, and the difference was often substantial. It is instructive to explore the reasons for this poor performance. In our data set, the pairwise correlations of forecast errors were very high. In all, we can estimate 48 correlations (6 pairwise combinations and 8 variable/horizon combinations). These correlations, which are given in Table 4, were uniformly high, ranging from .82 to .96. The high correlations and differences among error variances resulted in the normal model placing negative weights on some of the forecasts. The weights and the combined forecasts are highly sensitive to small changes in the estimated correlations, as in multicollinearity problems in general. To illustrate this notion, some extreme cases with $n = 5$ are shown in Table 5. In Case 4, for example, the forecasts ranged from 3.258 to 8.456 and the actual value was 6.804, but the weights ranged from -2.364 to 3.355 and the combined forecast from the normal model was -9.012.

The poor performance of the basic normal model is not surprising. The potentially serious multicollinearity problem and previous empirical results (e.g., Winkler and Makridakis 1983) led us to expect difficulties. The next question is the extent to which the variations on the normal model discussed in Section 2 improved matters. The convexity constraint that forces the combined forecast to lie within the range of the individual forecasts resulted in some improvement in performance in terms of either MAD or MSE, as can be seen from Table 3. Over the eight variable/horizon combinations, the average reduction in MAD from the MAD for the normal model was 4.5%, and the average reduction in MSE was 6.9%. The simple average, however, yielded average reductions of 9.2% in MAD and 17.6% in MSE.

Imposing an independence assumption by setting the off-diagonal elements of $\hat{\Sigma}$ equal to zero yielded very good results. The values of MAD and MSE shown in Table 3 for the independence model are comparable to those for the simple average. The independence model provided average reductions of 9.2% in MAD and 17.4% in MSE when compared with the basic normal model. The good performance of this approach is consistent with previous studies (e.g., Newbold and Granger 1974 and Winkler and Makridakis 1983).

Using (4) to place more weight on recent observations did not lead to large improvements over the basic normal model. Some values of MAD and MSE decreased, while others increased; to conserve space, we will not present detailed results. Slight improvements occurred for β near 1 (e.g., a 2.2% reduction in MAD when $\beta = 1.1$). Larger values of β *reduced* the performance drastically (e.g., a 23.4%

Table 4. Pairwise Correlations of Forecast Errors From Different Econometric Models

		Nominal GNP			Real GNP		
Horizon		Wharton	Chase	DRI	Wharton	Chase	DRI
1	Chase	.90			.90		
	DRI	.91	.88		.86	.84	
	BEA	.83	.82	.82	.90	.85	.82
2	Chase	.88			.90		
	DRI	.90	.92		.93	.93	
	BEA	.87	.89	.93	.90	.90	.91
3	Chase	.92			.90		
	DRI	.95	.96		.92	.95	
	BEA	.94	.94	.96	.94	.93	.93
4	Chase	.88			.91		
	DRI	.88	.95		.92	.93	
	BEA	.87	.94	.94	.92	.93	.90

Table 5. Examples of Cases With Negative Weights and Extreme Forecasts

Case	Forecasts (Weights) Wharton	Chase	DRI	BEA	Normal Model Combined Forecast	Actual Value
1	11.470 (−.591)	8.860 (1.431)	8.171 (−1.997)	12.422 (2.157)	16.375	7.725
2	12.488 (−1.616)	8.908 (5.980)	8.954 (−.909)	12.278 (−2.455)	−5.193	13.116
3	7.648 (−3.002)	5.753 (−.451)	6.164 (6.720)	−1.108 (−2.267)	18.381	3.026
4	8.456 (−2.364)	7.395 (.116)	3.258 (3.355)	7.564 (−.107)	−9.012	6.804

increase in MAD when $\beta = 1.8$). It may be that the process was relatively stable. Moreover, higher values of β resulted in less weight being attached to "older" observations, thereby reducing the effective sample size and exacerbating the multicollinearity problem. Imposing the convexity constraint on the forecasts generated using (4) eliminated the drastic reductions in performance with large β but did not lead to large improvements over the basic normal model. Finally, using (4) and also imposing an independence assumption had very little effect as compared with using the independence assumption without (4). On balance, placing more weight on recent observations did not prove helpful in this study.

For tne above results, $\hat{\Sigma}$ and the resulting forecasts were based on the preceding $n = 20$ observations in any forecasting situation. We also tested all of the combined forecasting approaches with $n = 5$, 10, and 15. The results for the smaller values of n were similar to those for placing more weight on recent observations. With an independence assumption, changing n appeared to have very little effect. Without independence, reducing n led to some reductions and some increases in values of MAD and MSE. Any improvements were relatively minor, however. The results do not justify reducing n below the initial value of 20.

Relaxing the constraint that the weights sum to one and adding an intercept term yielded the results shown in the columns headed "OLS, No Intercept" and "OLS" in Table 3. Neither approach performed very well. The most general model, with unconstrained weights and an intercept term, was roughly comparable to the basic normal model. The model without the intercept term was slightly, but only slightly, better. Neither of the OLS approaches came close, on an overall basis, to the simple average or the normal model with the independence assumption. This evidence runs counter to the claim of Granger and Ramanathan (1984) that "the common practice of obtaining a weighted average of alternative forecasts should . . . be abandoned in favour of an unrestricted linear combination *including a constant term*" (p. 201).

The last two columns in Table 3 correspond to the use of the forecast from the econometric models with the best and worst forecasts, respectively, on the previous forecasting

Table 6. MAD and MSE for the Bayesian Model, With and Without Independence Assumption

Variable	Horizon	a = 4	10	20	50	100	20 (with independence)
MAD							
Nominal GNP	1	3.92	3.77	3.68	3.60	3.57	3.56
	2	4.43	4.45	4.48	4.52	4.54	4.55
	3	4.78	4.95	5.07	5.19	5.25	5.29
	4	6.04	5.83	5.69	5.60	5.56	5.53
Real GNP	1	2.93	2.84	2.84	2.85	2.86	2.87
	2	3.58	3.56	3.55	3.54	3.54	3.54
	3	3.62	3.55	3.54	3.56	3.58	3.59
	4	3.69	3.62	3.63	3.63	3.63	3.63
MSE							
Nominal GNP	1	21.62	20.07	19.24	18.62	18.40	18.27
	2	31.79	31.39	31.39	31.58	31.72	31.83
	3	38.66	40.11	41.33	42.55	43.10	43.44
	4	60.61	55.98	53.57	51.64	50.78	50.22
Real GNP	1	12.12	11.57	11.34	11.23	11.22	11.22
	2	18.46	18.12	17.96	17.84	17.80	17.77
	3	22.86	22.16	21.92	21.84	21.83	21.85
	4	25.82	24.53	23.99	23.64	23.48	23.26

occasion. These relatively simple approaches did reasonably well. "Best" improved on the basic normal model by an average of 6.4% in MAD and 12.2% in MSE. Somewhat surprisingly, "Worst" did almost as well, with average gains of 6.1% in MAD and 10.6% in MSE. Neither of these schemes performed quite as well as the simple average or the normal model with the independence assumption, but they were better than the normal model with the convexity constraint and much better than the OLS schemes.

The final approach used was the Bayesian model with a priori exchangeability, which generated combined forecasts from (7). This model was run for several values of α (the equivalent prior sample size) with and without the independence assumption. Some results are shown in Table 6. The values of MAD and MSE were somewhat but not highly sensitive to the choice of α for the nonindependent case, with any slight edge going to the higher values of α. With the independence assumption, the results were extremely insensitive to the choice of α, and only the case of $\alpha = 20$ is shown in Table 6. Note that in contrast to the results in Table 3, there was not much difference between the independent and nonindependent cases. The prior information appeared to have a strong stabilizing effect that produced great improvements in the nonindependent case. In general, the Bayesian model performed much better than the basic normal model and some of its variants. Overall, the Bayesian model was roughly comparable to the simple average and the normal model with the independence assumption.

4. SUMMARY AND DISCUSSION

In this article we have compared the performance of several methods for combining GNP forecasts from four major econometric models. These methods range from a simple average through a normal model with several ad hoc variants to a Bayesian model. In an overall comparison of the methods, three approaches performed better than the others. These three were the simple average, the normal model with an independence assumption, and the Bayesian model (with or without the independence assumption). The basic normal model performed poorly, and adjustments such as varying the number of data points used in the estimation process, weighting recent observations more heavily, and imposing a convexity constraint on the combined forecasts had relatively minor, if any, impact on the performance.

The three "successful" techniques in this study share some common characteristics. The simple average, of course, forces the weights assigned to the forecasts to be equal. The Bayesian model, with its a priori exchangeability assumption, takes the weights that would be used without the prior information and "shrinks" them toward equality. The larger a is, the more the weights move toward equality. The independence assumption does not lead to equal weights per se, but as long as the k forecast error variances are relatively similar, as was the case here, the weights will be reasonably close to each other. All three of these approaches avoid the extreme weights that can occur with the basic normal model and some of its variants. As a result, all three approaches are quite robust.

Our results concerning the performance of the different combination schemes are similar to results obtained by Makridakis and Winkler (1983) and Winkler and Makridakis (1983). Of particular interest is the fact that we have considered forecasts from econometric models rather than from time-series extrapolation methods. The relatively strong performance of the simple average for combining forecasts of GNP should be of comfort to decision makers who regularly combine individual econometric forecasts by simple averaging. Without sufficient data to estimate variances precisely (or in the event of nonstationarity), the assumption that forecasters are perfectly exchangeable (same variance, same correlations) would appear to be a reasonable assumption as long as the forecasters are all in the same "league." Under this exchangeability, the appropriate combining rule is the simple average.

The Bayesian model presented here also utilizes an exchangeability assumption for the prior distribution but then combines this prior distribution with sample information. To the extent that the sample information suggests that the forecasters should not be viewed as exchangeable, the weights based on the posterior distribution may move away from the prior starting point of equal weights. The decision maker can control how quickly the sample information can change matters by the choice of α, the equivalent prior sample size. As $a \to \infty$, the Bayesian combined forecast approaches a simple average. Of course, the prior distribution need not rest on an exchangeability assumption, but in many forecasting situations the forecasters will be viewed as similar enough to make such an assumption reasonable. Another option might be to assume exchangeability only within subsets of the forecasters, particularly when the forecasters are of different "types" (e.g., a subset of forecasts from econometric models, a subset of judgmental forecasts, or a subset of forecasts from extrapolation methods). In any event, the Bayesian model provides a formal procedure for including prior information, and the simple average is one possible outcome of this procedure.

In contrast to the simple average and the Bayesian model, which have a solid rationale when exchangeability of forecasters seems reasonable (either exclusively or suitably modified by sample information), the normal model with the independence assumption is more difficult to justify. The forecast errors for the four econometric models considered here are clearly not independent, nor did we expect them to be independent. The degree of dependence is very high. The independence assumption just happens to be an ad hoc assumption that works well in this case (and probably in many other cases also). Independence forces the weights to lie in the unit interval, and barring very large differences in forecast accuracy among the forecasters, the weights will be quite close. Thus, although exchangeability appears to be more reasonable than independence in the economic forecasting situation, the two assumptions tend to lead to similar combined forecasts.

The results presented here for methods of combining forecasts are consistent with the message given in Makridakis et al. (1984) for extrapolation methods: simpler methods perform well in comparison to more complex methods. A promising research topic concerns the robustness of these results on combining forecasts in situations where performance differs more substantially and consistently among individual forecasters and a simple average therefore seems less appealing. Greater differences in individual performance would lead to greater differences in the weights with an independence assumption and with the Bayesian model. The individual differences might enable the scheme of using the best forecaster from the previous forecast to perform well. Moreover, if the differences are relatively consistent, the normal model variants such as the OLS approaches might do better, although the presence of high dependence will always create difficulties for these methods.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation Grants IST8018578 and PRA8413106 and by Indiana University. We are grateful to George Jaszi and Donald Straszheim for providing the forecasts from the BEA and Wharton models, respectively, and to the referees for helpful suggestions.

[Received July 1984. Revised May 1985.]

REFERENCES

Clemen, R. T. (1984), "Modeling Dependent Information: A Bayesian Approach," unpublished Ph.D. dissertation, Indiana University.

Figlewski, S., and Urich, T. (1983), "Optimal Aggregation of Money Supply Forecasts: Accuracy, Profitability and Market Efficiency," *Journal of Finance,* 28, 695–710.

Geisser, S. (1965), "A Bayes Approach for Combining Correlated Estimates," *Journal of the American Statistical Association,* 60, 602–607.

Granger, C. W. J., and Ramanathan, R. (1984), "Improved Methods of Combining Forecasts," *Journal of Forecasting,* 3, 197–204.

Halperin, M. (1961), "Almost Linearly Optimum Combination of Unbiased Estimates," *Journal of the American Statistical Association,* 56, 36–43.

LaValle, I. H. (1970), *An Introduction to Probability, Decision, and Inference,* New York: Holt, Rinehart and Winston.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society,* Ser. B, 34, 1–41.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1984), *The Forecasting Accuracy of Major Time Series Methods,* Chichester, England: John Wiley.

Makridakis, S., and Winkler, R. L. (1983), "Averages of Forecasts: Some Empirical Results," *Management Science,* 29, 987–996.

Morris, P. A. (1977), "Combining Expert Judgments: A Bayesian Approach," *Management Science,* 23, 679–693.

Newbold, P., and Granger, C. W. J. (1974), "Experience With Forecasting Univariate Time Series and the Combination of Forecasts," *Journal of the Royal Statistical Society,* Ser. A, 137, 131–164.

Winkler, R. L (1981), "Combining Probability Distributions From Dependent Information Sources," *Management Science,* 27, 479–488.

Winkler, R. L., and Makridakis, S. (1983), "The Combination of Forecasts," *Journal of the Royal Statistical Society,* Ser. A, 146, 150–157.