The Combination of Forecasts

Article *in* Journal of the Royal Statistical Society Series A (General) · January 1983

DOI: 10.2307/2982011

J. R. Statist. Soc. A (1983),
146, *Part* 2, *pp.* 150–157

The Combination of Forecasts

By ROBERT L. WINKLER and SPYROS MAKRIDAKIS

Indiana University, USA — *INSEAD, Fontainebleau, France*

SUMMARY

Aggregating information by combining forecasts from two or more forecasting methods is an alternative to using just a single method. In this paper we provide extensive empirical results showing that combined forecasts obtained through weighted averages can be quite accurate. Five procedures for estimating weights are investigated, and two appear to be superior to the others. These two procedures provide forecasts that are more accurate overall than forecasts from individual methods. Furthermore, they are superior to forecasts found from a simple unweighted average of the same methods.

Keywords: FORECASTING; COMBINED FORECASTS; WEIGHTED AVERAGES; ESTIMATION OF WEIGHTS; ACCURACY OF FORECASTS

1. INTRODUCTION

The traditional approach to forecasting involves choosing the forecasting method judged most appropriate of the available methods (Chambers *et al.*, 1971) and applying it to some specific situations. The choice of a method depends upon the characteristics of the series and the type of application (Makridakis and Wheelwright, 1978). The rationale behind such an approach is the notion that a "best" method exists and can be identified. An alternative to the traditional approach is to aggregate information from different forecasting methods by aggregating forecasts. This eliminates the problem of having to select a single method and rely exclusively on its forecasts.

The combination of forecasts has been tried before on a somewhat limited basis with considerable success. Methods for combining forecasts in terms of weighted averages were discussed in Bates and Granger (1969) and Newbold and Granger (1974) (see also Dickinson, 1973, 1975; Bunn, 1978). In Newbold and Granger (1974), the authors concluded (p. 143):

> It does appear . . . that Box–Jenkins forecasts can frequently be improved upon by combination with either Holt–Winters or stepwise autoregressive forecasts, and we feel that our results indicate that in any particular forecasting situation combining is well worth trying, as it requires very little effort. Further improvement is frequently obtained by considering a combination of all three types of forecast.

In an extensive study of the accuracy of forecasting methods, Makridakis *et al.*, (1982, 1983) found that a simple average and a weighted average of forecasts from six methods outperformed virtually all or perhaps even all of the individual methods, including the six methods being combined as well as the other sixteen individual methods included in the study.

In a recent study, Makridakis and Winkler (1983) found that the accuracy of combined forecasts was little influenced by the specific methods included in the combination. Furthermore, it was shown that accuracy increased with increases in the number of methods being combined, although a degree of saturation was reached after about four or five methods. Finally, it was found that the variability of accuracy among different combinations decreased as the number of methods included in the combination increased. In Makridakis and Winkler (1983), a simple

 0035–9238/83/146150 $2.00

average was used to combine the forecasts of the various methods considered in the study. However, this might not be the best combination rule; weighted averages might perform better.

The use of a weighted average to combine forecasts allows the consideration of the relative accuracy of the individual methods and of the covariances of forecast errors among the methods. Perhaps one method tends to be more accurate than another, and thus it might seem reasonable to give the first method more emphasis than the second. Also, potential dependence among the methods may influence the relative emphasis that should be given to different methods. For an extreme example, suppose that we have three methods and that the correlations among their forecast errors are zero for methods 1 and 2, zero for methods 1 and 3, and one for methods 2 and 3. In this case, the forecasts provided by the second and third methods are redundant and should not each be given the same weight as that given to the first method.

The weights assigned to different forecasting methods, then, should be related to the covariance matrix of forecast errors. If this covariance matrix is known, the optimal weights can be determined (see Section 2). Of course, the covariance matrix will not be known in practice. Bates and Granger (1969) suggested five procedures for estimating the weights and tried these procedures for a combination of two forecasting methods. In Newbold and Granger (1974), a combination of three methods was considered with all five procedures. The combined forecasts in Makridakis *et al*., (1982, 1983) involved six methods, but alternative procedures for estimating the weights were not investigated.

In this paper we present empirical results involving the combination of ten forecasting methods for a large sample of series (1001 series). The five procedures suggested in Bates and Granger (1969) and Newbold and Granger (1974) to estimate the weights are used and compared. Also, the accuracy of the combined forecasts using various weighting schemes is compared with the accuracy of forecasts using a simple unweighted average of the methods and with the accuracy of the individual methods themselves.

The paper is organized as follows. The combination of forecasts via weighted averages is discussed in Section 2. The design of the empirical study is presented in Section 3, and the results of the study are given in Section 4. Section 5 consists of a brief summary and discussion of the implications of the results.

2. COMBINING FORECASTS VIA WEIGHTED AVERAGES

A weighted average of forecasts that minimizes the variance of the combined forecast error can be found (Newbold and Granger, 1974), and this weighted average can, alternatively, be developed from a Bayesian model and interpreted as a posterior mean (Winkler, 1981). The combined forecast for period t is of the form

$$\hat{x}_t = \sum_{i=1}^{p} w_i \hat{x}_t^{(i)},$$

where $\hat{x}_t^{(i)}$ is the forecast for period t from forecasting method i, w_i is the weight assigned to method i and p is the number of methods.

If Σ, the covariance matrix of forecast errors from the p methods, is known, then the optimal weights are

$$w_i = \sum_{j=1}^{p} \alpha_{ij} \Big/ \sum_{h=1}^{p} \sum_{j=1}^{p} \alpha_{hj},$$

where the α_{ij} terms are the elements of Σ^{-1}. We work in terms of percentage errors, so that the elements of Σ are of the form

$$(\Sigma)_{ij} = \text{Cov}\,[e_t^{(i)}, e_t^{(j)}] \quad \text{for } i \neq j$$

and

$$(\Sigma)_{ii} = V[e_t^{(i)}],$$

where

$$e_t^{(i)} = [x_t - \hat{x}_t^{(i)}]/x_t$$

for $i = 1, \ldots, p$.

The five procedures given in Newbold and Granger (1974) for estimating the weights when Σ is not known are as follows:

$$1.\quad w_i = \left(\sum_{s=t-\nu}^{t-1} e_s^{(i)2}\right)^{-1} \Big/ \sum_{j=1}^{p} \left(\sum_{s=t-\nu}^{t-1} e_s^{(j)2}\right)^{-1}. \tag{1}$$

$$2.\quad w_i = \sum_{j=1}^{p} \hat{\alpha}_{ij} \Big/ \sum_{h=1}^{p} \sum_{j=1}^{p} \hat{\alpha}_{hj}, \tag{2}$$

where the $\hat{\alpha}_{ij}$ terms are the elements of $\hat{\Sigma}^{-1}$, with

$$(\hat{\Sigma})_{ij} = \nu^{-1} \sum_{s=t-\nu}^{t-1} e_s^{(i)} e_s^{(j)}.$$

$$3.\quad w_i = \beta w_{i,t-1} + (1-\beta)\left[\left(\sum_{s=t-\nu}^{t-1} e_s^{(i)2}\right)^{-1} \Big/ \sum_{j=1}^{p} \left(\sum_{s=t-\nu}^{t-1} e_s^{(j)2}\right)^{-1}\right], \tag{3}$$

where $0 < \beta < 1$ and $w_{i,t-1}$ is the weight assigned to method i based on the data preceding period $t-1$.

$$4.\quad w_i = \left(\sum_{s=1}^{t-1} \gamma^s e_s^{(i)2}\right)^{-1} \Big/ \sum_{j=1}^{p} \left(\sum_{s=1}^{t-1} \gamma^s e_s^{(j)2}\right)^{-1}, \tag{4}$$

where $\gamma \geqslant 1$.

$$5.\quad w_i = \sum_{j=1}^{p} \hat{\alpha}_{ij} \Big/ \sum_{h=1}^{p} \sum_{j=1}^{p} \hat{\alpha}_{hj}, \tag{5}$$

where the $\hat{\alpha}_{ij}$ terms are elements of $\hat{\Sigma}^{-1}$, with

$$(\hat{\Sigma})_{ij} = \sum_{s=1}^{t-1} \gamma^s e_s^{(i)} e_s^{(j)} \Big/ \sum_{s=1}^{t-1} \gamma^s,$$

where $\gamma \geqslant 1$.

In the empirical analysis, we used the values 3, 6, 9 and 12 for ν; 0.5, 0.7 and 0.9 for β; and 1.0, 1.5 and 2.0 for γ. The smaller values of ν restrict the estimation to the most recent observations; smaller values of β and larger values of γ imply that more weight is given to recent observations.

3. DESIGN OF THE STUDY

The 1001 time series used in the Makridakis *et al.*, (1982, 1983) forecasting accuracy study were used here as well. As can be seen from Table 1, the 1001 series included different sources and types of data.

TABLE 1
Number of series of each type

	Micro-data				*Macro-data*			
	Total firm	*Major divisions*	*Below major divisions*	*Industry*	*GNP or its major components*	*Below GNP and its major components*	*Demographic*	
Yearly	16	29	12	35	30	29	30	181
Quarterly	5	21	16	18	45	59	39	203
Monthly	10	89	104	183	64	92	75	617
	31	139	132	236	139	180	144	1001

The following ten forecasting methods were considered:

1. Naive (Persistence).
2. Simple moving average.
3. Single exponential smoothing.
4. Adaptive response rate exponential smoothing.
5. Holt's linear exponential smoothing.
6. Brown's linear exponential smoothing.
7. Brown's quadratic exponential smoothing.
8. Linear regression.
9. Holt–Winters.
10. Automatic AEP.

There was not any specific reason for selecting these particular methods as opposed to other alternatives, except that their forecasts were available and that they can all be used in a completely automated mode. For the first eight methods, the data were deseasonalized before the method was applied. (Using the same methods without deseasonalizing resulted in much less accurate forecasts for many of the series.) For more details concerning the individual methods or the series, see Makridakis *et al.*, (1982, 1983).

The ten forecasting methods were used to generate forecasts for several periods, or horizons, into the future (6 periods for yearly data, 8 periods for quarterly data and 18 periods for monthly data) for each of the 1001 series. The total number of forecasts was 13 816. The combination rule given in Section 2 was used to find combined forecasts for all of these cases and for all of the procedures discussed in Section 2 for estimating weights. For forecasts with a horizon of more than one period, the weights for the one-period-ahead forecasts were used.

4. RESULTS

The average mean absolute percentage error (MAPE) and the average mean square error (MSE) were calculated for individual forecasting methods and for combined forecasts. Because of the variety of series, average MSE does not provide a fair overall evaluation because it tends to be dominated by those series involving large forecasts and values. Thus, the results are presented primarily in terms of MAPE and in terms of some pairwise comparisons between combined forecasts and individual methods.

The average MAPE is given for the ten individual forecasting methods in Table 2. Average MAPE has been computed for different time horizons, ranging from one-period-ahead forecasts to 18-period-ahead forecasts (as well as for the model fitting). Averages over various combinations

TABLE 2

Average MAPE over 1001 time series for individual forecasting methods and various time horizons

Method	Model Fitting	Horizon 1	2	3	4	5	6	8	12	15	18	Average of Horizons 1–4	1–6	1–8	1–12	1–15	1–18
1	9.6	9.1	11.3	13.3	14.6	18.4	19.9	19.1	17.1	21.9	26.3	12.1	14.4	15.2	15.7	16.4	17.4
2	8.4	11.5	14.9	17.0	17.8	21.5	22.3	20.6	17.8	23.2	29.4	15.3	17.5	18.1	18.1	18.6	19.6
3	9.5	8.6	11.6	13.2	14.1	17.7	19.5	17.9	16.9	21.1	26.1	11.9	14.1	14.8	15.3	16.0	16.9
4	10.6	9.4	13.5	14.0	15.3	18.1	20.2	18.0	17.1	21.4	26.0	13.1	15.1	15.6	15.9	16.5	17.4
5	8.8	8.7	11.0	13.3	15.2	19.1	21.6	24.8	23.9	33.7	48.3	12.1	14.8	16.7	18.4	20.2	22.9
6	9.0	8.7	10.9	13.8	15.0	18.7	21.1	24.5	23.1	30.8	43.7	12.1	14.7	16.6	18.0	19.6	21.9
7	9.3	9.8	12.7	16.6	18.8	25.7	31.0	45.1	40.7	64.0	108.3	14.5	19.1	23.7	26.9	31.2	38.5
8	15.6	15.5	16.9	19.1	18.3	21.9	23.0	24.2	29.7	49.1	70.7	17.4	19.1	20.0	22.6	25.5	29.8
9	9.3	8.7	10.9	13.2	14.9	19.0	21.5	24.3	23.0	32.8	47.0	11.9	14.7	16.5	18.1	19.8	22.4
10	9.9	9.1	11.9	13.4	13.7	17.9	20.3	20.3	19.3	24.8	28.8	12.0	14.4	15.5	16.3	17.5	18.8

TABLE 3

Average MAPE for combined forecasts

		Model Fitting	Horizon 1	2	3	4	5	6	8	12	15	18	1–4	1–6	1–8	1–12	1–15	1–18
Procedure 1	ν= 3	7.8	8.2	10.2	12.1	13.0	16.6	18.9	19.8	19.1	24.2	31.5	10.9	13.2	14.4	15.5	16.6	18.0
	ν= 6	7.5	8.2	10.1	11.8	12.9	16.4	18.4	18.8	18.5	23.4	30.0	10.8	13.0	14.1	15.1	16.2	17.5
	ν= 9	7.8	8.3	10.3	11.7	12.6	15.7	17.1	18.5	18.1	23.4	30.2	10.7	12.6	13.7	14.7	15.9	17.3
	ν=12	7.9	8.4	10.3	11.8	12.6	15.6	17.2	18.5	18.3	23.0	28.9	10.8	12.7	13.7	14.8	15.9	17.2
Procedure 2	ν= 3	14.1	12.5	16.8	22.6	25.3	37.6	46.1	57.4	38.0	57.7	86.4	19.3	26.8	31.3	32.0	27.3	30.0
	ν= 6	8.5	9.3	12.0	14.2	15.4	20.2	25.3	35.3	28.5	35.4	48.5	12.6	16.0	18.9	20.9	22.9	25.5
	ν= 9	8.3	9.6	10.9	13.5	14.2	18.1	20.4	23.6	24.0	31.8	41.5	12.1	14.4	16.1	17.9	19.7	22.0
	ν=12	8.4	9.2	11.0	13.8	14.0	17.8	20.0	22.7	23.1	31.7	39.9	12.0	14.3	15.8	17.6	19.5	21.6
Procedure 3	ν=3 β=0.5	7.8	8.2	10.1	11.9	12.8	16.3	18.3	18.8	18.5	23.4	29.5	10.8	12.9	14.0	15.1	16.1	17.4
	ν=3 β=0.7	7.8	8.2	10.1	11.8	12.7	16.2	18.1	18.6	18.2	23.1	29.2	10.7	12.8	13.9	14.9	16.0	17.2
	ν=3 β=0.9	7.8	8.2	10.2	11.9	12.9	16.1	18.3	18.6	18.2	23.0	29.0	10.8	12.9	14.0	15.0	16.0	17.3
	ν=6 β=0.5	7.5	8.2	10.1	11.8	12.8	16.3	18.3	18.7	18.3	23.2	29.7	10.7	12.9	14.0	15.0	16.0	17.3
	ν=6 β=0.7	7.5	8.1	10.1	11.7	12.7	16.2	18.2	18.6	18.2	23.0	29.4	10.7	12.8	13.9	14.9	15.9	17.2
	ν=6 β=0.9	7.5	8.1	10.0	11.8	12.8	16.1	18.2	18.6	18.3	23.2	29.3	10.7	12.8	13.9	14.9	16.0	17.3
	ν=9 β=0.5	7.8	8.3	10.2	11.7	12.5	15.6	17.0	18.4	18.1	23.1	29.3	10.7	12.6	13.7	14.7	15.8	17.2
	ν=9 β=0.7	7.8	8.3	10.2	11.6	12.5	15.5	17.0	18.3	18.1	22.9	29.0	10.6	12.5	13.6	14.7	15.8	17.1
	ν=9 β=0.9	7.8	8.2	10.2	11.7	12.5	15.5	17.2	18.4	18.3	23.4	29.3	10.6	12.5	13.7	14.8	15.9	17.3
	ν=12 β=0.5	7.9	8.4	10.3	11.7	12.5	15.5	17.1	18.4	18.3	23.0	28.9	10.7	12.6	13.8	14.8	15.9	17.3
	ν=12 β=0.7	7.9	8.4	10.3	11.7	12.5	15.5	17.1	18.4	18.3	23.1	29.0	10.7	12.6	13.7	14.8	15.9	17.3
	ν=12 β=0.9	7.9	8.3	10.3	11.8	12.6	15.6	17.2	18.5	18.4	23.6	29.6	10.7	12.6	13.8	14.9	16.1	17.5
Procedure 4	γ=1.0	22.0	8.4	10.5	12.3	13.3	15.8	18.2	19.3	18.5	23.8	29.9	11.1	13.1	14.2	15.2	16.3	17.6
	γ=1.5	22.0	8.5	10.5	12.4	13.2	16.0	18.3	19.3	18.5	23.5	30.3	11.2	13.2	14.3	15.2	16.3	17.6
	γ=2.0	22.0	8.6	10.5	12.5	13.2	16.1	18.6	19.8	18.8	24.0	31.4	11.2	13.3	14.5	15.4	16.5	17.9
Procedure 5	γ=1.0	78.7	9.0	11.4	12.6	13.5	15.3	17.2	19.6	19.5	24.5	28.4	11.6	13.2	14.3	15.4	16.5	17.8
	γ=1.5	85.9	9.4	12.1	13.9	15.0	17.5	20.0	23.7	23.6	32.1	35.3	12.6	14.6	16.4	17.7	19.4	21.2
	γ=2.0	76.8	9.1	11.0	13.8	14.9	18.8	21.2	25.1	25.6	33.6	45.0	12.2	14.8	16.8	18.3	20.2	22.9
Simple Average			8.5	10.8	12.3	13.4	16.6	18.8	18.8	18.5	23.5	30.5	11.3	13.4	14.4	15.4	16.5	17.8

of horizons (e.g. 1–4) are shown. The last column, headed "1–18", represents the overall average MAPE across all 13 816 forecasts for each method. This overall MAPE ranges from 16.9 (for single exponential smoothing) to 38.5 (for Brown's quadratic exponential smoothing).

In Table 3, the average MAPE is presented for the combined forecasts. The five procedures from Section 2 for estimating weights were tried with various values for the parameters. The overall MAPE ranges from 17.1 (for Procedure 3 with $\nu = 9$ and $\beta = 0.7$) to 30.0 (for Procedure 2 with $\nu = 3$). In general, the most accurate forecasts are provided by Procedure 3 (which is relatively insensitive to the choice of ν and β) and Procedure 1 (with the larger values of ν). Procedures 4 and 5 have larger values of MAPE, with the former performing better than the latter and both doing better for small γ than for large γ. Procedure 2 appears to be the weakest of the lot.

The relative performance of the different procedures does not vary greatly by horizon. Smaller values of ν and β for Procedures 1 and 3 fare slightly better for short time horizons, while larger values seem better for long time horizons. As can be seen from the results given in Table 4 for

TABLE 4

Average MAPE for combined forecasts for yearly, quarterly and monthly series

			Yearly	Quarterly	Monthly	All
Procedure 1		ν=3	16.7	18.0	18.1	18.0
		ν=6	17.0	17.4	17.6	17.5
		ν=9	17.3	17.0	17.4	17.3
		ν=12	19.7	16.8	17.3	17.2
Procedure 2		ν=3	35.3	53.1	37.5	30.0
		ν=6	16.0	33.0	25.5	25.5
		ν=9	19.8	21.6	22.1	22.0
		ν=12	22.3	19.6	21.9	21.6
Procedure 3	ν=3	β=0.5	16.7	16.9	17.6	17.4
		β=0.7	16.8	16.8	17.4	17.2
		β=0.9	17.3	16.8	17.3	17.3
	ν=6	β=0.5	17.0	17.2	17.4	17.3
		β=0.7	17.0	17.1	17.3	17.2
		β=0.9	17.2	16.9	17.4	17.3
	ν=9	β=0.5	17.3	16.7	17.2	17.2
		β=0.7	17.3	16.6	17.2	17.1
		β=0.9	17.4	16.5	17.4	17.3
	ν=12	β=0.5	19.8	16.6	17.2	17.3
		β=0.7	19.9	16.5	17.3	17.3
		β=0.9	20.0	16.5	17.5	17.5
Procedure 4		γ=1.0	16.9	18.2	17.6	17.6
		γ=1.5	16.7	18.5	17.6	17.6
		γ=2.0	16.6	18.9	17.9	17.9
Procedure 5		γ=1.0	15.8	20.8	17.6	17.8
		γ=1.5	16.1	25.7	21.0	21.2
		γ=2.0	17.5	23.8	23.2	22.9
Simple Average			18.0	16.7	18.0	17.8

TABLE 5

Percentage of time that combined forecast is better than individual forecasting methods

		Forecasting Method										
		1	2	3	4	5	6	7	8	9	10	All
Procedure 1	ν=3	55.3	59.7	53.8	56.8	54.9	55.9	66.2	58.1	53.0	53.9	6.7
	ν=6	53.1	56.9	50.7	53.7	51.5	52.2	63.1	57.5	51.0	53.2	9.3
	ν=9	55.9	59.9	53.8	57.0	56.0	57.7	66.5	58.5	54.0	54.9	7.0
	ν=12	55.0	58.8	52.8	55.9	56.7	58.5	67.1	58.3	54.8	55.7	7.2
Procedure 2	ν=3	54.3	56.7	52.6	55.4	48.7	50.6	61.3	58.3	48.9	52.8	8.6
	ν=6	52.7	56.2	50.1	52.9	49.3	50.0	61.3	57.4	48.6	51.0	8.5
	ν=9	53.1	56.4	50.6	53.3	49.7	51.7	61.8	56.6	49.7	52.3	8.9
	ν=12	51.9	55.5	48.9	52.8	52.7	53.4	64.0	56.9	51.6	53.3	8.1
Procedure 3	ν=3 β=0.5	55.8	60.3	54.3	57.2	55.8	57.0	66.3	58.5	53.5	54.4	7.1
	ν=3 β=0.7	56.1	60.4	54.4	57.3	56.1	57.2	66.2	58.7	54.0	54.8	7.0
	ν=3 β=0.9	56.4	60.7	54.4	57.3	56.1	57.0	65.6	58.8	54.4	55.1	7.3
	ν=6 β=0.5	56.3	60.3	54.2	57.1	56.2	57.4	66.3	58.6	54.3	54.5	7.1
	ν=6 β=0.7	56.5	60.4	54.4	57.4	56.1	57.7	66.2	58.7	54.3	55.0	7.3
	ν=6 β=0.9	56.4	60.3	54.4	57.2	56.3	57.4	65.8	58.7	54.5	55.2	7.3
	ν=9 β=0.5	56.2	60.0	54.0	57.0	56.1	57.8	66.5	58.6	54.2	55.2	7.1
	ν=9 β=0.7	56.2	59.9	54.0	57.2	56.2	57.9	66.4	58.6	54.4	55.2	7.2
	ν=9 β=0.9	56.0	59.8	54.1	56.9	56.3	57.7	66.1	58.5	54.7	55.3	7.2
	ν=12 β=0.5	55.1	59.0	53.2	56.1	56.9	58.7	66.9	58.4	54.9	55.7	7.2
	ν=12 β=0.7	55.1	59.1	53.0	56.2	56.9	58.7	66.9	58.3	54.9	55.7	7.1
	ν=12 β=0.9	54.9	58.9	53.2	55.9	56.9	58.4	66.7	58.1	55.2	55.5	7.2
Procedure 4	γ=1.0	55.5	60.6	53.9	56.5	55.2	56.8	65.6	58.2	53.4	54.1	7.0
	γ=1.5	55.5	60.2	53.8	56.3	55.4	57.0	66.1	57.9	53.2	54.0	6.9
	γ=2.0	55.5	60.0	53.6	56.4	54.9	56.6	66.3	57.4	52.7	53.6	6.7
Procedure 5	γ=1.0	54.4	59.0	52.3	55.8	53.8	54.6	64.6	57.1	52.1	53.0	8.0
	γ=1.5	52.0	55.1	49.6	51.8	49.9	50.9	62.9	57.4	48.2	50.9	8.0
	γ=2.0	53.0	55.7	50.7	52.9	48.9	50.2	60.8	56.0	46.8	51.6	8.1

yearly, quarterly and monthly series, the smaller values of ν and β perform slightly better for yearly series.

How do the combined forecasts perform *vis-à-vis* the individual forecasting methods? From Tables 2 and 3, it seems clear that the better combined forecasts have an advantage over all of the individual methods for shorter time horizons and over most of the individual methods for longer time horizons. On a forecast-by-forecast basis, Procedures 1, 3 and 4 have smaller errors over half of the time when compared with *any* of the individual methods, as can be seen from Table 5. A combined forecast is seldom better than *all* of the individual methods simultaneously; this happens only about 6–9 per cent of the time (see the last column in Table 5). However, on an overall basis Procedures 1, 3 and 4 beat each individual method. Finally, it should be noted from Table 5 that even the worst of the combining procedures does better than seven of the ten individual methods.

A weighted average, of course, is not the only type of combined forecast. An alternative is to use a simple average. A simple average of the ten methods has an overall MAPE of 17.8. The values of MAPE for a simple average for different horizons and series are given in the bottom rows of Tables 3 and 4. In general, they are uniformly higher than the comparable MAPE values for the better procedures in Tables 3 and 4. Weighted averages therefore outperform the simple average where combinations of forecasts are involved, even though the differences in accuracy are not large.

5. SUMMARY AND DISCUSSION

In this paper we have investigated the accuracy of combined forecasts consisting of weighted averages of forecasts from individual methods. Five different procedures were used to estimate weights, and two of the procedures (Procedures 3 and 1) outperformed the others. Both of these procedures relate the weights to reciprocals of sums of squared errors as opposed to basing the weights directly on an estimated covariance matrix of forecast errors. In Procedure 1, each weight is proportional to the reciprocal of a sum of squared errors. The weights in Procedure 3 are obtained by exponentially smoothing the weights from Procedure 1. Procedure 3 has the advantage of being relatively insensitive to the choice of values for its parameters.

The results concerning the relative merits of the different weighting schemes are consistent with previous results of Newbold and Granger (1974). Our results generalize their results in the sense that we have considered many more series, many more forecasting methods, and several time horizons instead of just one-step-ahead forecasts. In the remainder of this section, any mention of "combined forecasts" will refer to the forecasts generated by the better weighting procedures.

An argument for combining forecasts is that by aggregating information from different forecasting methods we may be able to generate forecasts that are more accurate than those from the individual methods. In this study, the combined forecasts were indeed more accurate under most conditions, with large time horizons providing some exceptions. On an overall basis, the combined forecasts were better than the individual methods.

The accuracy of weighted averages was also compared with the accuracy of a simple average, and the former outperformed the latter. In Makridakis *et al.*, (1982, 1983), a simple average of six methods tended to be slightly more accurate than a weighted average of the same methods. However, the weights were estimated by Procedure 2, which turned out to be the worst of the weighting schemes we investigated here. With schemes such as Procedures 1 and 3, it appears that differential weighting can lead to improved forecasts. Furthermore, these schemes are not difficult to implement, so it may be worthwhile to use differential weighting even though a simple average is the easiest way to combine forecasts and provides reasonable results.

The findings presented in this paper indicate that the combination of forecasts improves forecasting accuracy and should, therefore, be used more frequently in practice. Furthermore, it has been shown in Makridakis and Winkler (1983), using simple averages, that combining forecasts also reduces the variability of forecasting errors and hence the risk associated with the choice of forecasting methods. Aggregating information by combining forecasts has intuitive appeal, which is confirmed empirically by this study. We speculate that combining forecasts from more diverse sources (e.g. econometric models and expert judgements as well as time series methods) could lead to even greater gains in accuracy. However, this is an area that requires further research.

ACKNOWLEDGEMENTS

We would like to thank Eric Camelin for computational assistance. This research was supported in part by INSEAD and in part by the US National Science Foundation under Grant IST8018578.

REFERENCES

Bates, J. M. and Granger, C. W. J. (1969) The combination of forecasts. *Op. Res. Quart.*, **20**, 451–468.

Bunn, D. W. (1978) *The Synthesis of Forecasting Models in Decision Analysis*. Basel: Birkhäuser Verlag.

Chambers, J. C., Mullick, S. K. and Smith, D. D. (1971) How to choose the right forecasting technique. *Harvard Business Rev.*, **49**, 45–74.

Dickinson, J. P. (1973) Some statistical results in the combination of forecasts. *Op. Res. Quart.*, **24**, 253–260.

——— (1975) Some comments on the combination of forecasts. *Op. Res. Quart.*, **26**, 205–210.

Makridakis, S. *et al.*, (1982) The accuracy of extrapolation (Time Series) methods: results of a forecasting competition. *J. Forecasting*, **1**, 111–153.

Makridakis, S. *et al.*, (1983) *The Accuracy of Major Extrapolation (Time Series) Methods*. London: Wiley (in press).

Makridakis, S. and Wheelwright, S. (1978) *Forecasting: Methods and Applications*. New York: Wiley.

Makridakis, S. and Winkler, R. L. (1983) Averages of forecasts: some empirical results. *Manag. Sci.*, **29**, in press.

Newbold, P. and Granger, C. W. J. (1974) Experience with forecasting univariate time series and the combination of forecasts. *J. R. Statist. Soc.* A, **137**, 131–165.

Winkler, R. L. (1981) Combining probability distributions from dependent information sources. *Manag. Sci.*, **27**, 479–488.