MANAGEMENT SCIENCE
Vol. 29, No. 9, September 1983
Printed in U.S.A.

AVERAGES OF FORECASTS: SOME EMPIRICAL RESULTS*

SPYROS MAKRIDAKIS† AND ROBERT L. WINKLER‡

An alternative to using a single forecasting method is to average the forecasts obtained from several methods. In this paper we investigate empirically the impact of the number and choice of forecasting methods on the accuracy of simple averages. It is concluded that the forecasting accuracy improves, and that the variability of accuracy among different combinations decreases, as the number of methods in the average increases. Thus, combining forecasts seems to be a reasonable practical alternative when, as is often the case, a "true" model of the data-generating process or a single "best" forecasting method cannot be or is not, for whatever reasons, identified.

(FORECASTING; COMBINED FORECASTS; AVERAGES OF FORECASTS)

1. Introduction

The usual approach to forecasting involves choosing a forecasting method among several competing alternatives (such as exponential smoothing, ARIMA, or econometric models) and using that method to generate forecasts for the series of interest (e.g., see Markridakis and Wheelwright [5]). Different methods will generally provide different forecasts. However, forecasts from a given method may provide some useful information that is not conveyed in forecasts from other methods. Thus, instead of choosing a single forecasting method, it seems reasonable to consider aggregating information by generating forecasts from several methods and then combining these forecasts. In this manner, the ultimate forecasts should contain more information than is the case when only a single method is used. This could provide more accurate forecasts and improved decisions based on these forecasts.

Bates and Granger [1] and Newbold and Granger [6] discussed the combination of forecasts. It is interesting to note that very little empirical work has been done in this highly promising area. Some empirical results yielded the following conclusions (Newbold and Granger [6, 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 a more extensive study of the accuracy of forecasting methods, Makridakis *et al.* [3], [4] 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. This included the six methods being combined *and* the other 16 individual methods considered in the study. Moreover, although the weighted average takes into account the relative accuracy of the six methods and the covariances of forecast errors among the methods, the simple average of the forecasts from the six methods was more accurate than the weighted average. This result provided the motivation for further study of combinations that involve simple averages.

* Accepted by Ambar G. Rao; received May 1982. This paper has been with the authors 1 month for 1 revision.

† INSEAD, Fontainebleau, France.
‡ Indiana University.

0025-1909/83/2909/0987$01.25

In Makridakis *et al.* [3], [4], six methods were combined. There was no particular reason for the use of six methods as opposed to some other number of methods. Moreover, the specific methods were chosen because they were relatively easy to use (and would thus be feasible to combine in practice). No prior information about their relative accuracy or how they might complement each other when combined was considered. Varying the number and choice of methods being combined could affect the accuracy of the combined forecasts. An increase in the number of methods might lead to more accurate forecasts but would be more costly because more individual forecasts would have to be generated. Also, for a fixed number of methods, accuracy surely would vary somewhat as the specific methods were changed. A combination of any six methods would not yield the same forecasts as the six methods used in the Makridakis *et al.* study, for instance.

In this paper we report the results of a study designed to investigate the impact of the number and choice of forecasting methods on the accuracy of simple averages of forecasts from the methods. The design of the study is discussed in §2 and the results are given in §3. A brief summary and some discussion of the implications of the results are presented in §4.

2. Design of the Study

The 1001 time series used in the Makridakis *et al.* [3], [4] forecasting accuracy study were used here as well. The forecasting methods considered formed a subset of the methods used in Makridakis *et al.* As can be seen from Table 1, the 1001 series included different sources and types of data. Because of time constraints, some of the methods were run only on a systematic sample of 111 series chosen from the larger set of 1001 series.

The following fourteen 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
11. Box–Jenkins
12. Lewandowski's FORSYS system
13. Parzen's ARAMA methodology
14. Bayesian forecasting.

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

In 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.) The first ten methods were used for all 1001 series, and the last four methods were only used for the subsample of 111 series. For more details concerning the series, the individual methods, or their performance, see Makridakis *et al.* [3], [4].

In the forecasting competition reported in Makridakis *et al.*, the forecasting methods were used to generate forecasts for several periods into the future (6 periods for yearly data, 8 periods for quarterly data, and 18 periods for monthly data) for each series. The total number of forecasts was 13,816 for the 1001 series (and the first ten methods) and 1528 for the 111 series (and all 14 methods). For $p = 1, \ldots, 10$, we took all $\binom{10}{p}$ possible combinations of forecasting methods and found average forecasts for each combination for all 13,816 situations. That is,

$$\hat{x}_t = \frac{\hat{x}_t^{(1)} + \hat{x}_t^{(2)} + \cdots + \hat{x}_t^{(p)}}{p},$$

where $\hat{x}_t^{(i)}$ is the forecast of x_t from the ith method in the combination. Similarly, all $\binom{14}{p}$ possible combinations were used for $p = 1, \ldots, 14$ for the 1528 situations in the subsample of 111 series.

No model fitting is necessary for the simple combination rule studied here, although the information for model fitting was available. This means, for example, that information concerning the relative accuracy of the individual methods is not taken into consideration. On the other hand, a simple average is trivial to compute.

Various measures are available to evaluate forecasts. Because of the variety of series, mean square errors would tend to be dominated by those series involving large forecasts and values. To avoid this problem, we worked with percentage errors and evaluated the forecasts in terms of mean absolute percentage error (MAPE). The MAPE was computed for each possible combination for all forecasts as well as for particular subsets of forecasts (e.g., forecasts for yearly data, forecasts for different time horizons).

3. Results

For the entire set of 1001 series, ten individual forecasting methods were considered. In Figure 1, MAPE is shown as a function of p, the number of methods in the combination. The three curves in Figure 1 represent, for each value of p, the highest and lowest MAPE among all averages of p methods as well as the average MAPE over all combinations of p methods. The case with $p = 1$ corresponds to the use of individual methods, and at the other extreme, $p = 10$ corresponds to the average of all ten methods.

The average MAPE consistently declines as the number of methods increases, decreasing from 22.56 when $p = 1$ to 17.80 when $p = 10$. This provides an indication of the increases in accuracy that might be expected from an increase in the number of methods. The reductions in MAPE are themselves decreasing as p increases. That is, the marginal impact of including an additional method decreases as the number of methods increases.

Of course, the average MAPE has been averaged over all possible combinations for each p. If the best combination for each p could somehow be determined in advance, the "Low MAPE" curve in Figure 1 would be relevant. This curve declines slightly as p increases from 1 to 4 and then increases a bit, as might be expected since the number of combinations and the differences among combinations are smaller as the upper limit of $p = 10$ is approached. Nonetheless, it is interesting to note that it is possible to

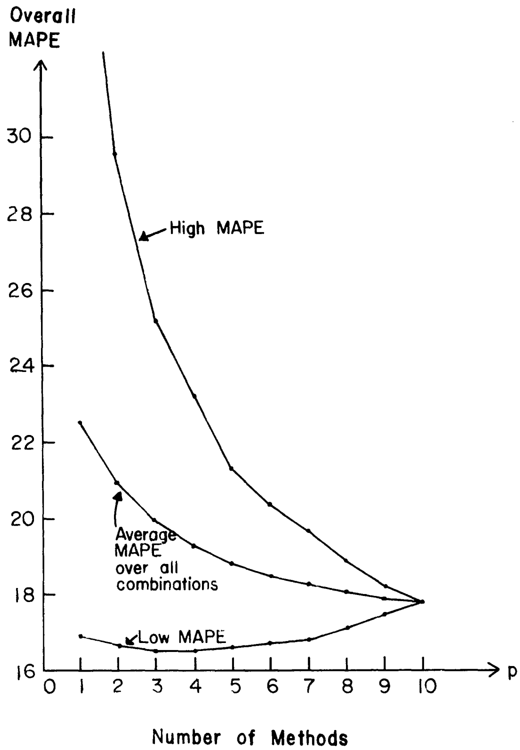

FIGURE 1. Overall MAPE as a Function of Number of Methods (1001 Series).

improve on the best individual method by combining it with some other method which is not as good on an individual basis.

The greatest improvements in MAPE come at the upper end, represented by the "High MAPE" curve in Figure 1. With an unfortunate choice of a single method, we could wind up with an MAPE of 38.5. But when $p = 5$, for instance, even the *worst* combination provides an MAPE of 21.3, which is better than the average MAPE for individual methods. From a practical point of view, improvements of this magnitude can be of considerable value. Furthermore, the reduction in risk associated with averages is even more important. When a single method is used, the risk of not choosing the best method can be very serious. The risk diminishes rapidly when more methods are considered and their forecasts are averaged. In other words, the choice of the best method or methods becomes less important when averaging.

The variability of MAPE for different combinations of p methods decreases as p increases, as indicated by the fact that the "High MAPE" and "Low MAPE" curves approach each other, meeting at $p = 10$. The variance of MAPE is shown as a function of p in Figure 2. The reduction in variability as p increases can be partially explained by the averaging effect well known in statistics and by the finite population factor associated with the limitation on the number of methods. However, the variance decreases much faster than these two factors would suggest. For instance, at $p = 5$, the variance of MAPE is 1.58, which is less than one-third of 5.13, the theoretical variance for the sample mean [that is, the initial variance (46.14) divided by 5 and multiplied by a finite population correction factor].

The average MAPE and variance of MAPE are given in Table 2 for yearly, quarterly, and monthly series and in Table 3 for different time horizons. For all three types of series and for all 18 time horizons, the average MAPE and the variance of MAPE decrease as the number of methods increases. The rate of decrease of average

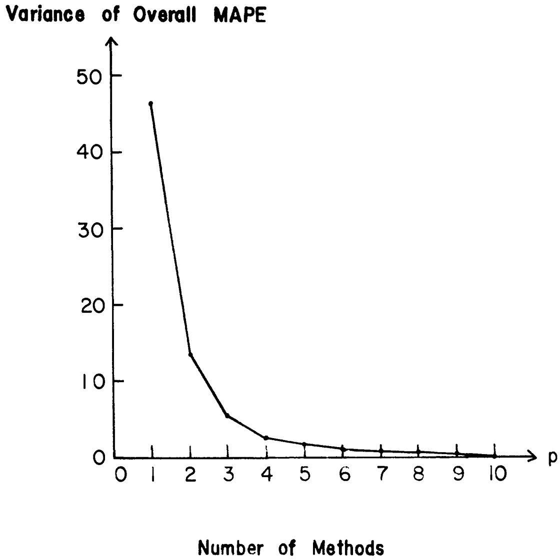

FIGURE 2. Variance of Overall MAPE as a Function of Number of Methods (1001 Series).

MAPE is greater for longer time horizons. For example, the average MAPE for $p = 5$ is 76 percent of the average MAPE for $p = 1$ when a horizon of 18 periods is considered but 89 percent of the average MAPE for $p = 1$ when a horizon of one period is considered.

Do the methods that perform well for short time horizons perform equally well for longer time horizons? In Table 4, the lowest MAPE among all combinations is shown separately for different time horizons. By identifying the best combination for each time horizon, we can improve on the "Low MAPE" curve in Figure 1, but the improvement is relatively slight. For instance, when $p = 5$, the low MAPE in Figure 1 is 16.6, and the overall average of lowest MAPE values in Table 4 is 16.3.

For the subset of 111 series, fourteen individual forecasting methods were considered. The general nature of the results for this subset is consistent with the results for the entire set of 1001 series. The average MAPE and the variance of MAPE are decreasing functions of p, as can be seen from Figures 3 and 4 for the entire subset and from Tables 5 and 6 for different types of series and different time horizons.

TABLE 2

Average MAPE (Variance of MAPE) for All Forecasting Horizons for Yearly, Quarterly, and Monthly Series as a Function of Number of Methods (1001 *Series*)

Number of Methods p	Yearly	Quarterly	Monthly	All
1	19.94 (9.70)	22.35 (48.30)	22.83 (58.40)	22.56 (46.14)
2	18.81 (3.56)	20.21 (13.05)	21.26 (14.41)	20.94 (13.68)
3	18.42 (2.06)	19.16 (5.31)	20.23 (6.09)	19.96 (5.72)
4	18.24 (1.31)	18.47 (2.37)	19.56 (3.10)	19.33 (2.86)
5	18.13 (0.87)	17.96 (1.15)	19.09 (1.76)	18.88 (1.58)
6	18.07 (0.59)	17.58 (0.59)	18.75 (1.03)	18.56 (0.91)
7	18.03 (0.38)	17.27 (0.33)	18.47 (0.60)	18.29 (0.52)
8	18.00 (0.22)	17.04 (0.18)	18.25 (0.33)	18.09 (0.28)
9	17.98 (0.10)	16.84 (0.07)	18.08 (0.14)	17.93 (0.12)
10	17.96 (0.00)	16.68 (0.00)	17.95 (0.00)	17.80 (0.00)

TABLE 3

Average MAPE (Variance of MAPE) for Different Time Horizons as a Function of Number of Methods (1001 *Series*)

Number of Methods p	Horizon 1	Horizon 2	Horizon 3	Horizon 4	Horizon 5	Horizon 6	Horizon 7	Horizon 8	Horizon 9
1	9.9 (4.6)	12.6 (4.0)	14.7 (4.4)	15.8 (3.3)	19.8 (6.4)	22.0 (11.1)	22.5 (37.1)	23.9 (63.2)	19.7 (17.2)
2	9.4 (1.3)	12.0 (2.3)	13.7 (1.4)	14.7 (1.0)	18.5 (1.7)	20.9 (2.8)	20.1 (8.4)	22.4 (16.3)	18.9 (5.5)
3	9.1 (0.5)	11.6 (1.1)	13.2 (0.6)	14.2 (0.4)	17.8 (0.6)	20.2 (0.9)	19.4 (3.8)	21.5 (6.9)	18.2 (2.5)
4	8.9 (0.3)	11.3 (0.6)	12.9 (0.3)	13.9 (0.2)	17.3 (0.3)	19.7 (0.4)	18.9 (2.0)	20.8 (3.3)	17.8 (1.4)
5	8.8 (0.2)	11.2 (0.4)	12.7 (0.2)	13.7 (0.1)	17.1 (0.1)	19.4 (0.2)	18.6 (1.1)	20.4 (1.7)	17.6 (0.9)
6	8.7 (0.1)	11.1 (0.2)	12.6 (0.1)	13.6 (0.1)	16.9 (0.1)	19.2 (0.1)	18.3 (0.6)	20.0 (0.9)	17.4 (0.6)
7	8.6 (0.1)	11.0 (0.1)	12.5 (0.1)	13.5 (0.0)	16.8 (0.0)	19.0 (0.1)	18.1 (0.4)	19.6 (0.5)	17.3 (0.4)
8	8.6 (0.0)	10.9 (0.1)	12.4 (0.0)	13.5 (0.0)	16.7 (0.0)	18.9 (0.0)	18.0 (0.2)	19.3 (0.3)	17.2 (0.2)
9	8.5 (0.0)	10.9 (0.0)	12.3 (0.0)	13.4 (0.0)	16.7 (0.0)	18.9 (0.0)	17.8 (0.1)	19.0 (0.1)	17.1 (0.1)
10	8.5 (0.0)	10.8 (0.0)	12.3 (0.0)	13.4 (0.0)	16.6 (0.0)	18.8 (0.0)	17.7 (0.0)	18.8 (0.0)	17.0 (0.0)

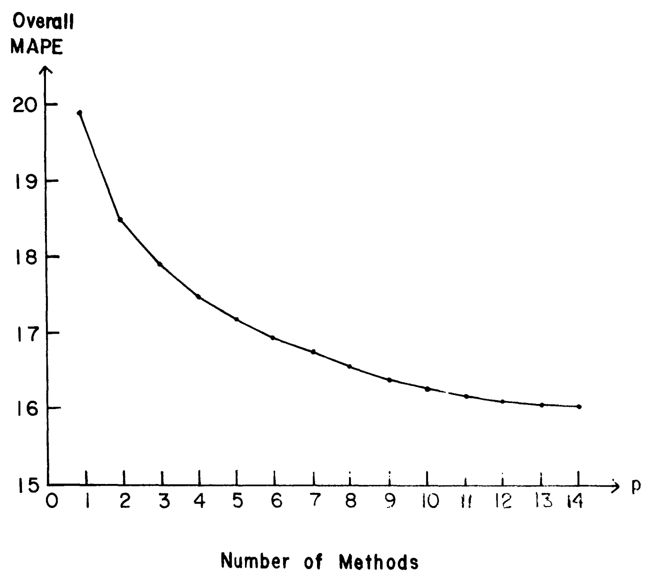

FIGURE 3. Average Overall MAPE as a Function of Number of Methods (111 Series).

TABLE 4

Lowest MAPE for Different Time Horizons as a Function of Number of Methods (1001 *Series*)

Number of Methods p	1	2	3	4	5	6	7	8	9
1	8.6	10.9	13.2	13.7	17.7	19.5	17.3	17.9	15.3
2	8.2	10.3	12.2	13.0	16.5	18.7	16.5	17.9	15.1
3	8.1	10.1	12.0	12.9	16.2	18.6	16.3	17.7	15.2
4	8.0	10.0	12.0	13.0	16.3	18.6	16.4	17.8	15.3
5	8.0	10.1	12.0	13.0	16.2	18.4	16.7	18.0	15.3
6	8.0	10.2	12.0	13.0	16.2	18.2	16.9	18.0	15.4
7	8.1	10.2	12.0	13.1	16.2	18.3	17.0	18.0	15.9
8	8.1	10.3	12.0	13.1	16.3	18.5	17.2	18.2	16.3
9	8.3	10.4	12.0	13.3	16.5	18.7	17.4	18.6	16.6
10	8.5	10.8	12.3	13.4	16.6	18.8	17.7	18.8	17.0

TABLE 3 (*continued*)

Horizon 10	Horizon 11	Horizon 12	Horizon 13	Horizon 14	Horizon 15	Horizon 16	Horizon 17	Horizon 18
23.0 (25.9)	25.9 (42.6)	22.9 (56.3)	25.6 (65.5)	30.4 (93.3)	32.3 (200.9)	35.4 (255.0)	37.4 (352.3)	45.5 (695.1)
20.9 (6.5)	23.3 (21.9)	21.0 (14.9)	23.3 (19.8)	27.1 (28.1)	29.6 (51.2)	30.2 (64.5)	32.6 (85.8)	40.9 (197.3)
20.1 (2.7)	22.1 (9.9)	20.2 (6.5)	22.4 (8.7)	25.7 (11.5)	27.8 (21.3)	28.2 (26.5)	30.5 (57.9)	38.0 (90.4)
19.7 (1.4)	21.3 (5.6)	19.7 (3.5)	21.8 (4.6)	24.8 (5.7)	26.6 (10.6)	26.7 (13.2)	29.0 (20.1)	35.9 (48.9)
19.5 (0.8)	20.8 (3.4)	19.3 (2.0)	21.3 (2.7)	24.1 (3.1)	25.7 (6.0)	25.7 (7.4)	28.0 (11.6)	34.5 (29.1)
19.3 (0.5)	20.5 (2.1)	19.1 (1.2)	21.0 (1.6)	23.7 (1.9)	25.0 (3.6)	24.8 (4.5)	27.3 (6.9)	33.4 (17.4)
19.2 (0.3)	20.2 (1.4)	18.9 (0.8)	20.8 (1.0)	23.3 (1.1)	24.5 (2.1)	24.1 (2.6)	26.6 (4.0)	32.5 (10.0)
19.1 (0.2)	20.0 (0.8)	18.7 (0.4)	20.7 (0.6)	23.0 (0.6)	24.0 (1.1)	23.6 (1.4)	26.1 (2.1)	31.7 (5.5)
19.0 (0.1)	19.8 (0.4)	18.6 (0.2)	20.5 (0.3)	22.8 (0.3)	23.7 (0.5)	23.2 (0.6)	25.7 (0.9)	31.0 (2.3)
18.9 (0.0)	19.7 (0.0)	18.5 (0.0)	20.4 (0.0)	22.7 (0.0)	23.5 (0.0)	22.9 (0.0)	25.3 (0.0)	30.5 (0.0)

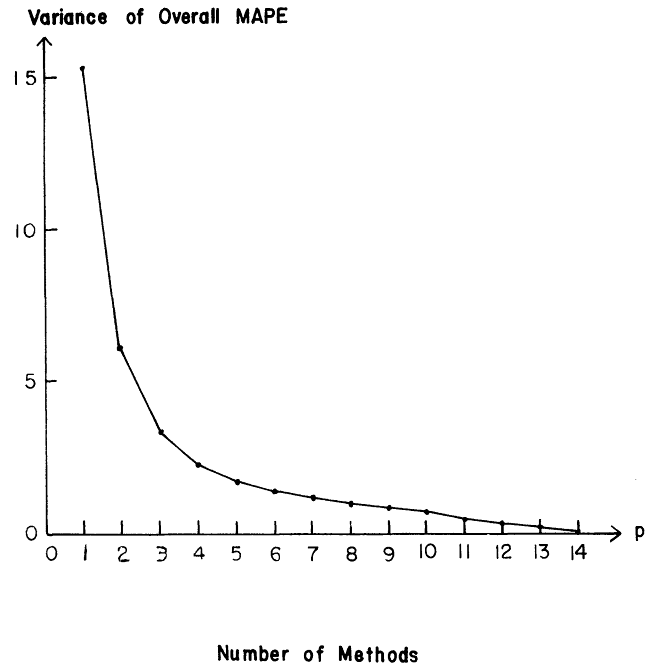

FIGURE 4. Variance of Overall MAPE as a Function of Number of Methods (111 Series).

TABLE 4 (*continued*)

Horizon 10	11	12	13	14	15	16	17	18	Overall average of lowest MAPE
17.1	17.8	16.9	18.2	20.1	21.1	21.2	21.7	26.0	16.9
16.8	17.4	16.3	17.8	19.8	21.2	21.1	21.5	25.5	16.4
16.9	17.3	16.3	17.7	19.8	20.9	20.8	21.5	25.2	16.2
17.0	17.1	16.3	17.6	20.0	20.8	20.8	21.5	25.2	16.2
17.4	17.0	16.3	18.0	20.2	20.9	20.4	21.5	24.9	16.3
17.7	17.0	16.6	18.2	20.7	20.4	20.2	21.3	25.0	16.3
18.0	17.2	16.9	18.5	20.8	20.3	20.8	21.4	25.2	16.4
18.2	18.1	17.3	19.1	21.4	21.5	20.9	22.5	26.2	16.8
18.5	18.8	17.8	19.6	22.0	22.5	22.1	23.8	28.2	17.2
18.9	19.7	18.5	20.4	22.7	23.5	22.9	25.3	30.5	17.8

TABLE 5

Average MAPE (Variance of MAPE) for All Forecasting Horizons for Yearly, Quarterly, and Monthly Series as Function of Number of Methods (111 *Series*)

Number of Methods p	Yearly	Quarterly	Monthly	All
1	15.20 (5.07)	23.63 (22.42)	19.76 (19.09)	19.87 (15.27)
2	14.39 (1.87)	21.47 (8.32)	18.47 (6.58)	18.51 (6.13)
3	14.08 (0.96)	20.65 (5.02)	17.86 (3.71)	17.90 (3.44)
4	13.94 (0.58)	20.02 (3.35)	17.46 (2.57)	17.49 (2.36)
5	13.86 (0.39)	19.67 (2.32)	17.14 (1.98)	17.19 (1.82)
6	13.81 (0.28)	19.43 (1.65)	16.89 (1.63)	16.95 (1.50)
7	13.77 (0.21)	19.25 (1.17)	16.66 (1.36)	16.74 (1.26)
8	13.75 (0.16)	19.12 (0.86)	16.46 (1.17)	16.57 (1.09)
9	13.73 (0.12)	19.01 (0.62)	16.29 (0.95)	16.41 (0.89)
10	13.71 (0.09)	18.93 (0.43)	16.16 (0.73)	16.30 (0.68)
11	13.70 (0.06)	18.86 (0.29)	16.06 (0.51)	16.21 (0.48)
12	13.69 (0.04)	18.80 (0.18)	15.99 (0.32)	16.15 (0.30)
13	13.68 (0.02)	18.75 (0.08)	15.94 (0.15)	16.10 (0.14)
14	13.67 (0.00)	18.72 (0.00)	15.91 (0.00)	16.07 (0.00)

4. Summary and Discussion

In this paper we have investigated the accuracy of combined forecasts consisting of simple averages of forecasts from individual methods. Once the individual forecasts have been generated, it is trivial to find the combined forecast, since no further fitting is required. The combined forecasts considered in this study were more accurate as the number of methods increased, although the accuracy tended to level off somewhat as more methods were added.

The accuracy of combined forecasts depends on both the number of methods in the average and the specific methods being combined. The variability associated with the choice of methods is reduced as more methods are included. If the best individual method could be identified in advance, then any gains from combining it with other methods would be minimal. Also, combining the forecasts from different methods obviously cannot be better than using the "true" underlying model of the process generating the data. But identifying the best method or specifying a "true" model of the data-generating process is not easy. The available data may be limited, the process

TABLE 6

Average MAPE (Variance of MAPE) forDifferent Time Horizons as a Function of Number of Methods (111 *Series*)

Number of Methods p	Horizon 1	Horizon 2	Horizon 3	Horizon 4	Horizon 5	Horizon 6	Horizon 7	Horizon 8	Horizon 9
1	9.7 (2.0)	11.8 (1.8)	13.9 (3.5)	15.6 (2.7)	17.5 (4.9)	18.8 (6.2)	21.4 (19.4)	21.2 (25.3)	16.3 (15.3)
2	8.9 (0.4)	10.7 (0.5)	12.7 (1.1)	14.5 (1.2)	16.2 (1.9)	17.4 (2.5)	21.1 (7.9)	19.7 (9.6)	14.9 (5.4)
3	8.6 (0.2)	10.3 (0.3)	12.2 (0.5)	13.9 (0.7)	15.7 (1.2)	16.8 (1.5)	20.7 (5.0)	19.0 (5.9)	14.5 (3.3)
4	8.4 (0.1)	10.1 (0.2)	11.8 (0.3)	13.7 (0.5)	15.3 (0.9)	16.4 (1.0)	20.4 (3.6)	18.6 (4.2)	14.2 (2.3)
5	8.3 (0.1)	9.9 (0.2)	11.6 (0.2)	13.5 (0.3)	15.1 (0.6)	16.2 (0.7)	20.2 (2.7)	18.4 (3.1)	14.1 (1.7)
6	8.2 (0.1)	9.8 (0.1)	11.5 (0.2)	13.3 (0.2)	15.0 (0.5)	16.0 (0.6)	20.1 (2.1)	18.2 (2.4)	14.0 (1.2)
7	8.2 (0.1)	9.7 (0.1)	11.4 (0.2)	13.2 (0.2)	14.9 (0.4)	15.9 (0.4)	20.1 (1.6)	18.1 (1.9)	13.9 (0.9)
8	8.1 (0.0)	9.7 (0.1)	11.3 (0.1)	13.1 (0.1)	14.8 (0.3)	15.8 (0.3)	20.0 (1.3)	18.0 (1.5)	13.9 (0.7)
9	8.1 (0.0)	9.6 (0.1)	11.3 (0.1)	13.0 (0.1)	14.7 (0.3)	15.8 (0.3)	20.0 (1.0)	17.9 (1.1)	13.8 (0.5)
10	8.1 (0.0)	9.6 (0.1)	11.2 (0.1)	13.0 (0.1)	14.6 (0.2)	15.7 (0.2)	19.9 (0.7)	17.9 (0.8)	13.8 (0.4)
11	8.0 (0.0)	9.6 (0.0)	11.2 (0.0)	12.9 (0.1)	14.6 (0.2)	15.7 (0.2)	19.9 (0.5)	17.9 (0.6)	13.7 (0.2)
12	8.0 (0.0)	9.5 (0.0)	11.2 (0.0)	12.8 (0.0)	14.5 (0.1)	15.6 (0.1)	19.9 (0.3)	17.8 (0.4)	13.7 (0.1)
13	8.0 (0.0)	9.5 (0.0)	11.2 (0.0)	12.8 (0.0)	14.4 (0.1)	15.6 (0.0)	19.8 (0.1)	17.8 (0.2)	13.7 (0.1)
14	8.0 (0.0)	9.5 (0.0)	11.2 (0.0)	12.8 (0.0)	14.4 (0.0)	15.6 (0.0)	19.8 (0.0)	17.8 (0.0)	13.6 (0.0)

may be changing over time, or the analyst may simply decide that the forecasting situation does not necessitate a thorough analysis of the data or the process. Combining forecasts is a viable practical alternative which provides reasonable forecasts. As noted in Koten [2], "Because economists haven't found any one forecasting technique that suceeds consistently . . . a consensus approach offers better long-run performance."

Using an average of forecasts is undoubtedly better than using a "wrong" model or a single poor forecasting method. Therefore, unless an adequate theory exists concerning the data-generating process or strong evidence is available indicating that a particular forecasting method is better than other methods for a given situation, it might be desirable to consider several methods. An advantage of combining forecasts is that when several methods are used, the results do not seem to be highly sensitive to the specific choice of methods. In terms of the worst accuracy among combinations with a given number of methods, the "High MAPE" curve in Figure 1 decreases extremely fast. In this sense, using a combined forecast is safer and less risky than relying on a single method. It may be cheaper, and this study shows that it is less risky, to use a combination of relatively simple methods (e.g., single exponential smoothing) than to use a single method which is more complex and requires personalized data analysis and costly fitting. Or, if the more complex method is used, it may still be better to combine it with one or more simple methods. The results reported here suggest that using averages of forecasts provides considerable practical benefits in terms of improved forecasting accuracy and decreased variability of accuracy.

Finally, it is important to note that averaging forecasts is not confined to combinations utilizing time series methods only, as was the case in the study. The desire to consider any and all available information means that forecasts from different types of sources should be considered. For example, we could combine forecasts from time series methods with forecasts from econometric models and with subjective forecasts from experts. Apparently combinations involving only subjective forecasts are currently used by some companies for major decisions [2]. The combination of time series and econometric forecasts as well as other so-called "objective" forecasts with subjective forecasts is a very promising area requiring further research. We believe that a great deal can be gained by studying the properties of various types of combined forecasts.[1]

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

TABLE 6 (*continued*)

Horizon 10	Horizon 11	Horizon 12	Horizon 13	Horizon 14	Horizon 15	Horizon 16	Horizon 17	Horizon 18
17.4 (13.9)	18.4 (24.5)	17.2 (19.2)	19.3 (45.6)	27.8 (73.3)	31.5 (158.1)	30.7 (163.4)	31.3 (193.4)	36.7 (122.5)
15.8 (4.3)	16.4 (6.5)	15.9 (5.9)	17.8 (10.6)	24.9 (21.6)	32.1 (29.2)	27.9 (29.8)	29.8 (37.7)	34.8 (42.3)
15.3 (2.4)	16.0 (3.9)	15.3 (3.0)	17.1 (6.0)	24.1 (12.8)	31.2 (16.9)	26.9 (16.9)	28.4 (18.6)	33.7 (23.0)
15.1 (1.7)	15.7 (2.8)	14.9 (1.9)	16.6 (4.1)	23.6 (8.7)	30.5 (12.0)	26.2 (11.7)	27.3 (12.9)	32.9 (15.5)
14.9 (1.2)	15.5 (2.1)	14.6 (1.3)	16.3 (3.1)	23.3 (6.4)	29.9 (9.5)	25.6 (9.0)	26.4 (11.4)	32.2 (12.0)
14.7 (0.9)	15.4 (1.6)	14.4 (1.0)	16.1 (2.4)	23.0 (5.0)	29.4 (8.1)	25.1 (7.5)	25.5 (11.5)	31.5 (10.2)
14.6 (0.7)	15.3 (1.3)	14.2 (0.8)	15.9 (2.0)	22.7 (4.0)	28.9 (7.2)	24.6 (6.6)	24.6 (11.1)	30.8 (9.2)
14.5 (0.6)	15.2 (1.0)	14.0 (0.7)	15.7 (1.6)	22.5 (3.2)	28.5 (6.5)	24.1 (5.9)	24.0 (10.8)	30.3 (8.4)
14.4 (0.4)	15.1 (0.8)	13.8 (0.6)	15.5 (1.3)	22.2 (2.5)	28.1 (5.4)	23.7 (5.1)	23.5 (9.0)	29.7 (7.4)
14.4 (0.3)	15.0 (0.6)	13.7 (0.5)	15.3 (1.0)	22.0 (2.0)	27.9 (4.1)	23.3 (4.1)	23.2 (6.7)	29.2 (6.0)
14.3 (0.3)	15.0 (0.4)	13.5 (0.4)	15.2 (0.8)	21.8 (1.4)	27.8 (2.8)	23.1 (2.8)	23.1 (4.6)	28.9 (4.2)
14.2 (0.2)	14.9 (0.3)	13.4 (0.3)	15.0 (0.5)	21.6 (0.9)	27.8 (1.7)	23.0 (1.7)	23.1 (2.8)	28.7 (2.6)
14.2 (0.1)	14.8 (0.1)	13.2 (0.1)	14.9 (0.3)	21.4 (0.4)	27.8 (0.8)	23.0 (0.8)	23.0 (1.3)	28.6 (1.2)
14.1 (0.0)	14.7 (0.0)	13.1 (0.0)	14.8 (0.0)	21.2 (0.0)	27.7 (0.0)	22.9 (0.0)	22.9 (0.0)	28.5 (0.0)

References

1. Bates, J. M. and Granger, C. W. J., "The Combination of Forecasts," *Oper. Res. Quart.*, Vol. 20 (1969), pp. 451–468.
2. Koten, J., "They Say No Two Economists Ever Agree, So Chrysler Tries Averaging Their Opinions," *Wall Street Journal* (Nov. 3, 1981), p. 31.
3. Makridakis, S. et al., "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *J. Forecasting*, Vol. 1 (1982), pp. 111–153.
4. ———, *The Accuracy of Major Extrapolation (Time Series) Methods*, Wiley, London, 1983 (in press).
5. ——— and Wheelwright, S., *Forecasting: Methods and Applications*, New York, Wiley, 1978.
6. Newbold, P. and Granger, C. W. J., "Experience with Forecasting Univariate Time Series and the Combination of Forecasts," *J. Roy. Statist. Soc., Ser.* A, Vol. 137 (1974), pp. 131–165.