MANAGEMENT SCIENCE
Vol. 31, No. 12, December 1985
Printed in U.S.A.

AGGREGATING SUBJECTIVE FORECASTS: SOME EMPIRICAL RESULTS*

ALISON HUBBARD ASHTON AND ROBERT H. ASHTON
Graduate School of Business Administration, New York University, New York, New York 10006

The impact on forecast accuracy of aggregating the subjective forecasts of up to 13 individuals was examined for five forecast weighting methods—equal weighting, two ex post methods that took advantage of prior information about the individuals' relative accuracy, and two ex ante methods based on objective and subjective assessments of relative accuracy. The individuals were executives, managers and sales personnel employed by Time, Inc., and the variable forecasted was the number of advertising pages sold annually by *Time* magazine over a 14-year period. The results show that both the average forecast error and the variance of the error decrease as additional individuals' forecasts are included in the aggregate. Only two to five individuals' forecasts must be included to achieve much of the total improvement available from combining all 13 forecasts. Three of the differential weighting methods produced more accurate forecasts than equal weighting, but the magnitude of the improvement was small. Implications for realistic forecasting situations are discussed, as are conditions under which the use of aggregates seems attractive.
(FORECASTING/APPLICATIONS)

1. Introduction

Improving the forecasts of variables on which business and economic decisions are based can be of great importance to forecast users and others. Forecasts can potentially be improved by enhancing the data on which they are based, and by the use of more sophisticated forecasting models. Another possibility for improving forecasts, and one which is often overlooked, involves the aggregation of different forecasts of the same variable.

Aggregation does not require additional input data beyond that ordinarily used, nor does it require that forecasting models be employed. Instead, the improvement in forecast accuracy attributable to aggregation results from "diversifying away" the random error inherent in the individual forecasts. Therefore, the combination of individual forecasts into an aggregate could be attractive in the case of *subjective* forecasts, i.e., those made in contexts for which modeling is not possible or is not considered feasible.

Some studies have examined the aggregation of econometric forecasts, e.g., Bates and Granger (1969), Makridakis et al. (1982), Makridakis and Winkler (1983), Newbold and Granger (1974), Winkler and Makridakis (1983), and others have focused on the aggregation of subjective forecasts made by individuals, e.g., Goldberg (1970), Libby and Blashfield (1978), Winkler (1968). However, the *number* of forecasts that should be included in the aggregate has seldom been addressed. Two exceptions are a study of econometric forecasts by Makridakis and Winkler (1983) and a study of individuals' forecasts by Libby and Blashfield (1978). In addition, analytical work by Einhorn *et al.* (1977) and Hogarth (1977, 1978) addresses this question. These studies suggest that much of the gain in forecast accuracy that can ultimately be achieved by aggregation can be achieved by combining a small number of forecasts.

*Accepted by Ambar G. Rao; received July 24, 1984. This paper has been with the authors $3\frac{1}{2}$ months for 1 revision.

0025-1909/85/3112/1499$01.25

The studies that address this issue have employed an equal-weighting aggregation method. That is, aggregate forecasts have been formed by taking simple averages of individual forecasts. The impact of using differential weighting methods that take into account the relative accuracy of the individual forecasts has not been investigated.[1] However, differential weighting could result in a reduction of the number of individual forecasts required to achieve a substantial gain in accuracy.

The present study examines the impact of aggregation in a real-world business forecasting task—the forecasting of annual advertising sales at *Time* magazine. Subjective forecasts of executives, managers and sales personnel employed by Time, Inc., who regularly make such forecasts as a normal part of their job responsibilities, provided the data for this study. Using a mean absolute error criterion, the impact on forecast accuracy of including additional individuals' forecasts in the aggregate is examined for five weighting methods—equal weighting, and four differential weighting methods based on ex post and ex ante measures of relative accuracy. §2 describes the data, while §3 explains the choice of weighting methods. The results are presented in §4, and some implications are discussed in §5.

2. Data

The data were taken from an earlier study by Ashton (1982). The task involved short-run forecasts of advertising sales for *Time* magazine. The participants were 13 executives, managers and sales personnel who worked for Time, Inc. magazines other than *Time* (e.g., *Life* and *Sports Illustrated*). The participants had the following titles: Publisher, Advertising Director, Marketing Director (2), Promotion Director, Production Manager, Business Manager (2), Assistant Business Manager, Sales Representative (3), and Advertising Records Manager. Five of the 13 participants actually prepare and update advertising sales forecasts (the publisher, the advertising director, the two business managers, and the assistant business manager), while the remaining eight perform such functions as supervising the compilation of historical data and preparing client-specific sales forecasts that provide inputs to the forecast of total annual sales. The participants were experienced, having been involved in the forecasting process for a minimum of eight years.

Forecasting advertising sales is extremely important at Time, Inc. for two reasons. First, advertising sales is typically the major source of profit for a magazine. Marginal production costs are high relative to circulation revenue, and contribution to profit from circulation revenue alone is typically low or negative. Second, the advertising sales forecast is the key variable in all planning for Time, Inc. magazines. Individual magazine policy determines the acceptable ratio of advertising pages to news and editorial pages in each issue. Thus, the advertising sales forecast is a prerequisite to all editorial and production commitments, as well as to financial budgeting. The importance of this forecast is underscored by the fact that executives' bonuses at Time, Inc. magazines are affected by the accuracy of their forecasts.

The task required the participants to forecast the number of *pages* of advertising sold annually by *Time* magazine over the 14-year period from 1965 to 1978. The forecasts were made after the first, second and third quarter of each year, for a total of 42 forecasts. (There is no reason to forecast after the fourth quarter, since all advertising for the year has been sold by that point.) The task was patterned closely after the process in which the participants are actually involved. Employees of Time, Inc. magazines prepare initial advertising page forecasts in August and update them in October for the following calendar year. As the year progresses, the forecasts are

[1] Winkler and Makridakis (1983) employed differential weighting methods, but they did not investigate the number of forecasts to include in the aggregate.

revised quarterly. The publisher, advertising director and business manager of a magazine prepare independent forecasts prior to arriving at the final forecast, which is agreed on by all three executives.

The variables on which the participants based their forecasts are included in the information set provided to Time, Inc. personnel for making such forecasts in practice. Five variables were made available to the participants: (1) the quarter to which the data applied (first, second or third); (2) the total number of advertising pages appearing in *Time* magazine that quarter; (3) the number of pages of liquor advertising appearing in the magazine that quarter; (4) the number of pages of automobile advertising appearing in the magazine that quarter; (5) the number of advertising pages committed by *Time* advertisers to date for the entire year.

The participants were told that the data provided to them were taken from *Time* magazine records for the years 1965 to 1978. Since the 42 data sets were presented in random order, however, they did not know the year for which they were forecasting, nor could they relate, for example, first- and second-quarter data for the same year. Participation in the study was voluntary. The participants received the materials in their offices and were asked to return them within one week. Many participants provided unsolicited comments that suggested they approached the task with enthusiasm and considered it challenging.

3. Weighting Methods

The mathematical aggregation of forecasts requires the choice of a weighting method, or combination rule, to be applied to the individual forecasts. One approach is to weight the individual forecasts equally, i.e., to compute the simple average of all the individual forecasts included in the aggregate. This weighting method was used by Makridakis and Winkler (1983) and by Libby and Blashfield (1978). Equal weighting offers the advantage of simplicity, but it does not take into account the relative accuracy of the individual forecasts that comprise the aggregate. Weighting methods that incorporate relative accuracy might be expected to yield better aggregate forecasts than equal weighting.

In addition to equal weighting, the present study employed four weighting methods based on relative accuracy assessments. Two of these differential weighting methods were ex post methods that took advantage of prior information—from the earlier study (Ashton 1982)—about the individuals' relative accuracy. These methods were based on two commonly used accuracy measures. One was the mean absolute error (in pages) of the 42 forecasts made by each individual. The other was the correlation, over the 42 observations, between the individual's forecasts and the number of advertising pages actually sold.

In applying each of the ex post weighting methods, the magnitude of the accuracy measure, instead of simply the individual's rank on the measure, was taken into account. Weights based on correlations were formed by dividing each forecaster's correlation by the sum of the correlations of all 13 forecasters, producing a set of normalized weights that summed to 1. Weights based on mean absolute error (MAE) were formed by dividing each forecaster's MAE by the sum of all 13 MAE's, taking the reciprocal of each resulting number, and normalizing the reciprocals.[2] The differences between the correlational and MAE accuracy measures include the squared-error loss function implicit in the correlational measure. The two measures provide somewhat different indications of the relative accuracy of the individuals' forecasts.

[2]Reciprocals were taken so that forecasters with lower (higher) MAE's would receive higher (lower) weights.

Neither of these ex post weighting methods could be used in a true forecasting situation, since the weights were determined from the series of actual forecasts.[3] The reason for employing them in this study is to provide an upper benchmark against which the performance of equal weights and ex ante weights can be compared. Two ex ante weighting methods, which ignored the accuracy information from the earlier study, were employed. An "objective" ex ante method involved ranking the 13 individuals by their respective positions in the organizational hierarchy at Time, Inc., and forming rank-sum weights after allowing for tied ranks. The conversion of ranks to weights was accomplished by assigning the largest rank (13) to the forecaster with the highest organizational rank, the next largest rank to the forecaster with the second-highest organizational rank, and so on. Then each forecaster's assigned rank was divided by the sum of the assigned ranks to produce a set of normalized weights that summed to 1. In using this weighting method, we are assuming that, a priori, it is reasonable to believe that organizational rank is positively correlated with the individuals' forecasting accuracy. Because of the importance of sales forecasting at Time, Inc., as described in §2, we believe this is a reasonable assumption.

A "subjective" ex ante method incorporated the knowledge of an experienced executive at Time, Inc., who ranked the 13 individuals according to the executive's perception of their likely relative accuracy. Rank-sum weighting was again used. The executive had been employed by Time, Inc. for more than ten years, and had worked with the forecasters for periods ranging from six to eight years. Because this weighting method avoided tied ranks, it could potentially provide a more complete differentiation among the individuals than could the "objective" ex ante method. The use of this method assumes, however, that the executive's ranking has some validity.

4. Results

For each of the five weighting methods, all possible combinations of the forecasts of k individuals ($k = 1, \ldots, 13$) were formed, and the mean absolute percentage error (MAPE) was computed for each combination. An example may help to clarify the computation of MAPE. Consider, for example, aggregate forecasts based on combinations of $k = 3$ individuals. There are 286 such combinations (i.e., $n!/(n-k)!\,k! = 13!/10!\,3! = 286$). For the first combination, aggregate forecasts were formed for each of the 42 data sets by applying the appropriate weights to the three individual forecasts, and the absolute errors of these aggregate forecasts, expressed as percentages of actual, were determined. Then, the mean absolute percentage error (MAPE) over the 42 aggregate forecasts was computed. This process was repeated for the other 285 combinations of three individuals, as well as for all combinations of all other levels of k individuals ($k = 1, \ldots, 13$), for each of the five weighting methods. Other descriptive statistics, such as the variance and the median of the MAPE's for all combinations of k individuals, were also computed.

Figure 1 shows forecast errors for the *equal* weighting method as a function of the number of individual forecasts in the aggregate. The worst combination and the best combination of k individuals are indicated by the "high MAPE" and "low MAPE" curves, respectively. The "average MAPE" curve represents the mean MAPE over all combinations of k individuals.

The average MAPE declines from 10.52% when $k = 1$ to 7.19% when $k = 13$, a decrease in average forecast error of approximately one-third. The largest decline occurs in moving from $k = 1$ to $k = 2$. Average MAPE then declines at a decreasing rate, indicating that the marginal impact of adding individuals' forecasts decreases as

[3] One could, of course, use weights based on the *past* relative accuracy of the individual forecasters—if past accuracy were known.

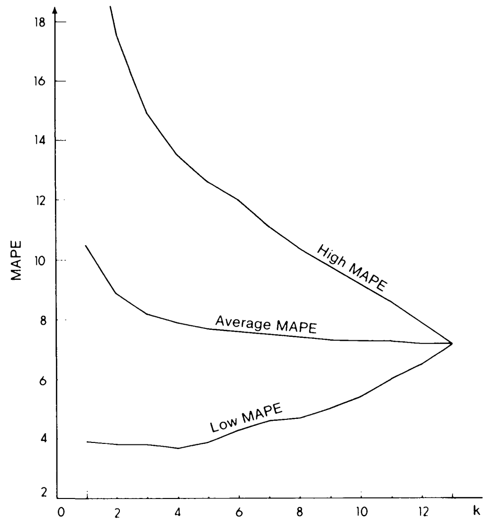

FIGURE 1. MAPE as a Function of the Number of Forecasts—Equal Weighting Method.

the number of individuals increases, and the average MAPE curve flattens out after $k = 5$.

The high MAPE curve depicts much greater improvement in forecast accuracy as k increases, declining from 20.24% when $k = 1$ to 7.90% when $k = 13$. In fact, the *worst* combinations of eight or more individuals produce lower MAPE's than the average MAPE for a single individual. The low MAPE curve, depicting the *best* forecasts at each level of k, decreases slightly from $k = 1$ to $k = 4$, indicating that even the best individual forecasts can be improved by aggregation. Beyond $k = 4$, the low MAPE curve increases as the forecasts of more individuals are added. This increase is not surprising, as Makridakis and Winkler (1983, p. 989) note, since both the number of combinations and the differences among combinations are smaller as the upper limit of k is approached.[4]

The variability of MAPE decreases as the forecasts of more individuals are added to the aggregate, as shown in Figure 2. At $k = 1$, the variance of MAPE is 0.19% (standard deviation of 4.34%). Consistent with the average MAPE, the variance declines most rapidly from $k = 1$ to $k = 2$ and the curve begins to flatten out after $k = 5$. When six or more individuals are included, the variance is less than 0.02%. Table 1 presents the mean and variance of MAPE for all k under the equal weighting method, as well as the high, low, median and quartile values.

Average MAPE's for the four *differential* weighting methods are shown in Figure 3, along with that of the equal weighting method (EQL) for comparison purposes. The overall shapes of these four average MAPE curves are similar to that of EQL. Both the ex post weighting method based on mean absolute error (MAE) and the ex post method based on correlational accuracy (COR) produce lower average MAPE's than equal weighting, at all levels of k. The MAE method results in lower average MAPE's than does COR at all levels of k, and is the best of the five methods considered. This might be expected since the aggregate forecasts are evaluated here in terms of mean absolute error.

[4]It is interesting to compare Figures 1 and 2 with the comparable figures in Makridakis and Winkler (1983, pp. 990–991). Although the magnitudes of both the average MAPE and the variance of MAPE are considerably smaller in the present study, the shapes of the respective curves in the two studies are extremely similar.

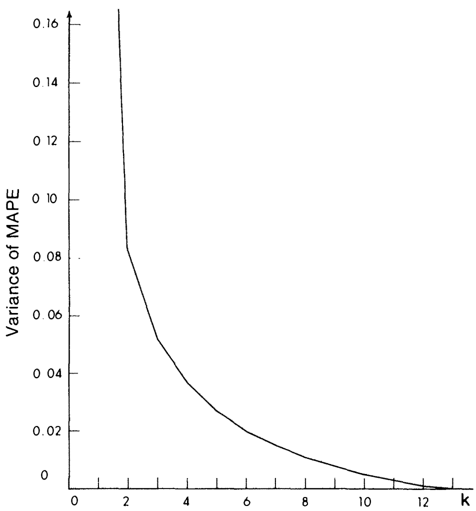

FIGURE 2. Variance of MAPE as a Function of the Number of Forecasts—Equal Weighting Method.

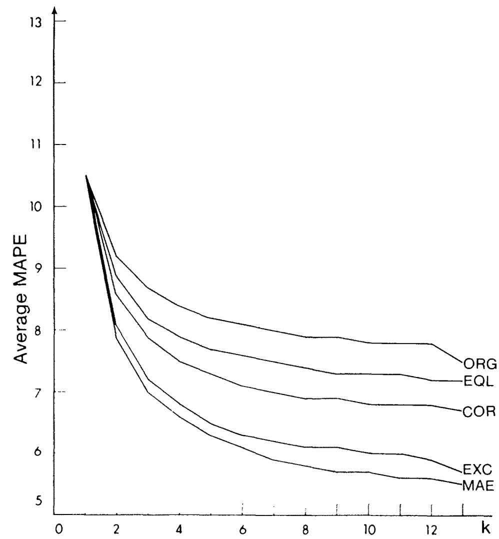

FIGURE 3. Average MAPE as a Function of the Number of Forecasts—for Five Weighting Methods.

TABLE 1

Descriptive Statistics for MAPE—Equal Weighting Method (EQL)

Number of Individuals (k)	Average	Variance	Low	First Quartile	Median	Third Quartile	High
1	10.52	0.1880	3.85	7.93	10.56	13.15	20.24
2	8.89	0.0833	3.79	6.56	8.41	10.46	17.46
3	8.23	0.0520	3.81	6.52	7.98	9.65	14.92
4	7.89	0.0365	3.65	6.45	7.70	9.11	13.48
5	7.69	0.0266	3.94	6.50	7.56	8.76	12.61
6	7.55	0.0198	4.31	6.52	7.45	8.50	11.97
7	7.45	0.0148	4.61	6.55	7.37	8.28	11.16
8	7.38	0.0109	4.73	6.60	7.35	8.13	10.40
9	7.33	0.0077	4.95	6.69	7.30	7.94	9.76
10	7.28	0.0052	5.43	6.74	7.27	7.77	9.15
11	7.25	0.0032	5.97	6.89	7.26	7.61	8.61
12	7.22	0.0014	6.46	6.96	7.31	7.41	7.90
13	7.19	—	—	—	—	—	—

The "objective" ex ante weights, based on the individuals' positions in the organizational hierarchy (ORG), produce the least accurate forecasts at all levels of k. In fact, the MAPE based on the combination of all 13 individuals for the ORG method is worse than the average MAPE based on combinations of only seven individuals for the EQL method. The "subjective" ex ante weights, based on the expected-accuracy rankings of the experienced Time, Inc. executive (EXC), result in average MAPE's that are only slightly higher than those produced by the best ex post method (MAE). The MAPE's associated with both the MAE and EXC weights eventually fall below 6.00%, while this level of accuracy is never reached by COR, ORG or EQL.

Descriptive statistics for the four differential weighting methods are presented in Table 2. It is interesting to note that the high MAPE's produced by the subjective ex ante weights (EXC) are the lowest of all weighting methods at each level of k. In other words, if only the "worst case" aggregates are considered, the weights provided by the experienced Time, Inc. executive result in better forecasts than any other weighting

TABLE 2

Descriptive Statistics for MAPE—Four Differential Weighting Methods

Number of Individuals	Method MAE				Method COR				Method ORG				Method EXC			
(k)	Average	Variance	Low	High	Average	Variance	Low	High	Average	Variance	Low	High	Average	Variance	Low	High
1	10.52	0.1880	3.85	20.24	10.52	0.1880	3.85	20.24	10.52	0.1880	3.85	20.24	10.52	0.1880	3.85	20.24
2	7.94	0.0849	3.44	17.06	8.58	0.0865	3.76	17.38	9.18	0.1317	3.46	19.09	8.11	0.0765	3.48	15.93
3	7.03	0.0532	3.39	14.08	7.85	0.0543	3.80	14.82	8.65	0.0953	3.44	18.05	7.21	0.0415	3.49	13.67
4	6.55	0.0371	3.35	12.76	7.48	0.0375	3.60	13.47	8.37	0.0699	3.74	16.29	6.76	0.0259	3.62	12.12
5	6.26	0.0270	3.62	11.87	7.26	0.0271	3.83	12.61	8.20	0.0519	3.88	14.97	6.49	0.0177	3.79	11.40
6	6.06	0.0200	3.76	11.25	7.11	0.0200	4.20	12.02	8.08	0.0389	3.94	13.63	6.32	0.0126	3.93	10.85
7	5.92	0.0149	3.93	10.19	7.00	0.0148	4.41	10.98	7.99	0.0290	4.15	12.75	6.21	0.0091	4.12	9.76
8	5.81	0.0110	4.10	9.44	6.93	0.0108	4.50	10.19	7.92	0.0213	4.50	12.09	6.12	0.0066	4.37	8.95
9	5.73	0.0079	4.26	8.77	6.87	0.0077	4.70	9.47	7.87	0.0153	4.88	11.08	6.05	0.0047	4.57	8.33
10	5.67	0.0054	4.59	8.28	6.83	0.0052	5.13	8.94	7.83	0.0103	5.30	10.29	6.00	0.0032	4.79	7.71
11	5.62	0.0033	4.91	7.53	6.79	0.0031	5.63	8.35	7.79	0.0063	5.85	9.64	5.95	0.0020	5.07	7.17
12	5.59	0.0015	5.24	6.76	6.76	0.0014	6.10	7.55	7.76	0.0028	6.35	8.71	5.92	0.0009	5.49	6.71
13	5.48	—	—	—	6.74	—	—	—	7.53	—	—	—	5.73	—	—	—

method. In addition, the variances of the MAPE's produced by the EXC weights are the smallest of all weighting methods.

As Figure 3 indicates, three of the differential weighting methods result in lower average MAPE's than does equal weighting. An additional question is the *magnitude* of the increase in forecast accuracy achieved by the differential methods. This issue can be addressed by comparing the average MAPE's produced by EQL with those produced by the better of the two ex post and ex ante methods (MAE and EXC). The difference between the MAPE's produced by the MAE and EQL methods is 0.95% when $k = 2$, and it increases to 1.71% when $k = 13$. The corresponding numbers are 0.78% and 1.46% for $k = 2$ and $k = 13$, respectively, when the EXC and EQL methods are compared. These numbers can be put in perspective by considering only the EQL method, and noting that average MAPE improves by 1.63% from $k = 1$ to $k = 2$ and by 2.29% from $k = 1$ to $k = 3$. While the two techniques for improving forecasts are not mutually exclusive, this analysis does suggest that more is to be gained by aggregation than by differential weighting.

5. Discussion

The principal results of this study can be summarized as follows: (1) Aggregates of subjective forecasts were found to be more accurate than the individuals' forecasts that comprised the aggregates. (2) Weighting the individuals' forecasts differentially produced better aggregate forecasts than did equal weighting for each of two ex post weighting methods, but for only one of two ex ante methods. (3) The incremental accuracy of the better ex post or ex ante weighting method over equal weighting was small. (4) Regardless of the weighting method employed, much of the total gain in forecast accuracy attributable to aggregation was achieved by combining a small number of individual forecasts. Implications of these results for forecast aggregation in general, and for the specific application of forecasting advertising sales for *Time* magazine, are explored below.

Generally, the results agree with those of other studies in their implication that a useful approach for improving forecast accuracy is to aggregate the forecasts of more than one individual. This approach could be especially advantageous for avoiding the detrimental effects of unknowingly relying on the *worst* of the available individuals. However, the results also show that it is possible to improve on the forecasts of the *best* individual by combining his or her forecasts with those of a second person.

The simple approach of weighting all the individual forecasts equally appears to be a promising solution to the problem of choosing a forecast weighting method. This conclusion is supported by the finding that the incremental accuracy of differential over equal weighting was small in the present study.[5] Moreover, ex post weights—which were found to be superior to ex ante weights in this study—may not be available in realistic forecasting contexts, since the relative accuracy of the individual forecasters might not be known. In such cases, differential weighting must rely on ex ante assessments of the *expected* relative accuracy of the individuals. Of course, the use of ex ante weights assumes that one can successfully differentiate the expertise of the individuals involved. The ability to do this is likely to differ from one practical situation to another. Therefore, the incremental accuracy provided by differential (ex ante) weighting might not be reliable in some situations.

The accuracy of small aggregates is an important result in subjective forecasting contexts because the costs associated with the time of the individual forecasters could be substantial. Although accuracy can be improved by continuing to include addi-

[5] Winkler and Makridakis (1983) also found differential weighting to be only slightly superior to equal weighting.

tional forecasts in the aggregate, the trade-off between increased accuracy and the costs associated with the individuals' time must be considered. However, the exact number of individual forecasts that should be included in the aggregate cannot be specified without knowledge of the loss function inherent in the specific forecasting context.

In a specific context such as the forecasting of advertising sales for *Time* magazine, one way to evaluate the increased accuracy provided by aggregates is to explicitly consider the level of improvement that would be regarded as "significant" in that context. We pursued this approach by asking the Time, Inc. executive who provided the EXC weights to specify a level of forecast improvement that would be viewed as significant for *Time* magazine. The executive replied that an improvement of 100 pages would be considered extremely important, but that an improvement of at least 60 pages would also be viewed as significant. These numbers represent 3.51% and 2.10%, respectively, of the mean number of advertising pages sold annually by *Time* magazine during the 14 years covered by the study (2,852 pages).

The data in Tables 1 and 2 indicate that a 60-page improvement was possible, on average, with an aggregate of only two individuals assuming MAE or EXC weighting, and with an aggregate of only three individuals assuming equal weighting. The tables also show that a 100-page improvement was possible, on average, with a 4-person aggregate assuming MAE or EXC weighting, but could not have been achieved with equal weighting as the maximum improvement in that case (at $k = 13$) was 95 pages. This analysis suggests that aggregate subjective forecasts based on a small number of individuals could have proved useful in this specific context.

A more formal benchmark against which to evaluate aggregate forecasts for *Time* magazine is the level of accuracy that could have been achieved by using a statistical forecasting model. To investigate this issue, linear multiple regression models were developed by regressing the actual number of advertising pages sold on the five independent variables described in §2. A model was based on the 30 quarters of data from 1965 to 1974, and used to forecast the number of advertising pages sold in 1975. This model was updated 11 times by adding one additional quarter of data each time, and the updated models were used to forecast future advertising page sales. The average MAPE for the 12 resulting forecasts is 4.22%, which is slightly better than the MAPE of 5.48% produced by the 13-person aggregate based on the best differential weighting method (MAE). This comparison ignores the costs associated with modeling (and with aggregation), which must be considered in making a rational choice between the two techniques.

In some practical situations, problems associated with statistical models will tend to increase the attractiveness of aggregates vis-à-vis models. For example, the data-generating process on which a model is based may change over time, or individuals within the organization may exhibit "resistance" to the use of models. Moreover, modeling may not be possible in some contexts because sufficient historical data on the relevant independent variables are unavailable, or because the dependent variable is unobservable. In such situations, the use of aggregates may be the only viable alternative for improving the accuracy of forecasts.[6]

[6]We are indebted to Robert Ching and Vicky Heiman for research assistance.

References

Ashton, A. H., "An Empirical Study of Budget-Related Predictions of Corporate Executives," *J. Accounting Res.*, 20, II (1982), 440–449.

Bates, J. M. and C. W. J. Granger, "The Combination of Forecasts," *Oper. Res. Quart.*, 20 (1969), 451–468.

Einhorn, H. J. et al., "Quality of Group Judgment," *Psychological Bull.*, 84 (1977), 158–172.

GOLDBERG, L. R., "Man Versus Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences," *Psychological Bull.*, 73 (1970), 422–432.

HOGARTH, R. M., "Methods for Aggregating Opinions," in H. Jungermann and G. de Zeeuw (Eds.), *Decision Making and Change in Human Affairs*, D. Reidel, Dordrecht, 1977, 231–255.

———, "A Note on Aggregating Opinions," *Organizational Behavior and Human Performance*, 21 (1978), 40–46.

LIBBY, R. AND R. K. BLASHFIELD, "Performance of a Composite as a Function of the Number of Judges," *Organizational Behavior and Human Performance*, 21 (1978), 121–129.

MAKRIDAKIS, S. ET AL., "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *J. Forecasting*, 1 (1982), 111–153.

——— AND R. L. WINKLER, "Averages of Forecasts: Some Empirical Results," *Management Sci.*, 29 (1983), 987–996.

NEWBOLD, P. AND C. W. J. GRANGER, "Experience with Forecasting Univariate Time Series and the Combination of Forecasts," *J. Roy. Statist. Soc., Ser. A*, 137, 2 (1974), 131–146.

WINKLER, R. L., "The Consensus of Subjective Probability Distributions," *Management Sci.*, 15 (1968), 61–75.

——— AND S. MAKRIDAKIS, "The Combination of Forecasts," *J. Roy. Statist. Soc., Ser. A*, 146, 2 (1983), 150–157.