Accuracy of Forecasting: An Empirical Investigation

Author(s): Spyros Makridakis, Michele Hibon and Claus Moser

Source: *Journal of the Royal Statistical Society. Series A (General)*, 1979, Vol. 142, No. 2 (1979), pp. 97-145

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2345077

J. R. Statist. Soc. A (1979),
142, *Part* 2, *pp.* 97–145

Accuracy of Forecasting: An Empirical Investigation

By Spyros Makridakis and Michèle Hibon

INSEAD—The European Institute of Business Administration

[Read before the Royal Statistical Society on Wednesday, December 13th, 1978, the President, Sir Claus Moser in the Chair]

Summary

In this study, the authors used 111 time series to examine the accuracy of various forecasting methods, particularly time-series methods. The study shows, at least for time series, why some methods achieve greater accuracy than others for different types of data. The authors offer some explanation of the seemingly conflicting conclusions of past empirical research on the accuracy of forecasting. One novel contribution of the paper is the development of regression equations expressing accuracy as a function of factors such as randomness, seasonality, trend-cycle and the number of data points describing the series. Surprisingly, the study shows that for these 111 series simpler methods perform well in comparison to the more complex and statistically sophisticated ARMA models.

Keywords: FORECASTING; TIME SERIES; FORECASTING ACCURACY

0. Introduction

The ultimate test of any forecast is whether or not it is capable of predicting future events accurately. Planners and decision makers have a wide choice of ways to forecast, ranging from purely intuitive or judgemental approaches to highly structured and complex quantitative methods. In between, there are innumerable possibilities that differ in their underlying philosophies, their cost, their complexity and their accuracy. Unfortunately, since information about these differences is not usually available, objective selection among forecasting methods is extremely difficult. The major purpose of this paper is to deal with one important aspect of choosing a forecasting methodology: accuracy. Section 1 will survey past research on accuracy and will look into the reasons why the reported accuracies of different studies vary, often significantly. Section 2 will report our own empirical findings on the accuracy of 111 time series and show how the results could be consistent with the conflicting conclusions of three major previous studies on the subject of accuracy. Section 3 will develop regression equations that express accuracy as a function of a number of factors related to the internal characteristics of the time series used. Section 4 will examine the reasons for variations in the relative accuracy of various methods and suggest needs for future research and Section 5 attempts a brief conclusion. The appendices describe the methods and technical details used to make the comparisons.

1. Research Findings on Accuracy Performance

The first choice to be made in predicting the future is whether to use a formal forecasting method or to rely on judgemental processes. Evidence from the psychological literature asserts that in repetitive situations, quantitative methods outperform clinical judgement (Goldberg, 1970; Hogarth, 1975; Sarbin, 1943; Sawyer, 1966; Slovic, 1972). In his review of the literature, Meehl (1954, 1965), for example, concluded that he found only one case in which clinical judgement was superior to a statistical model. This case and a subsequent one (Libby, 1976) have been disputed by Goldberg (1976) who reversed the exceptional findings by simply transforming the data. In a recent article, Dawes (1977) says that he knows of no other findings that have been reported showing the superiority of clinical judgement.

The idea that decision makers' models can perform better than decision makers themselves is difficult to accept; acting on this idea arouses emotional reactions and suggests dehumanizing overtones. Slovic (1972), Kahneman and Tversky (1973), Tversky (1974) Tversky and Kahneman (1974) and Dawes (1977) have investigated the causes behind the inability of clinical judgement to outperform formal models. The main difficulties with judgemental predictions are stated to be lack of application of valid principles, anchoring effects, regression biases, the assumption that specific cases can be generalized (representation biases), lack of reliability and the basing of predictions on irrelevant information.

Outside the psychological literature, judgemental forecasts of earnings per share have been analysed in some detail. A number of researchers have compiled histories of the forecasts of earnings per share made by analysts (representing judgemental predictions) and compared them with the results obtained from quantitative methods (Green and Segall, 1967; Cragg and Malkiel, 1968; Elton and Gruber, 1972; Niederhoffer and Regan, 1972). In these four studies, the researchers conclude that analysts do not perform as well in forecasting earnings per share as do quantitative techniques. However, these conclusions are not unanimous.

Johnson and Schmitt (1974) claim that analysts can do better than quantitative methods provided that they have accurate economic and industrial information. But they make no attempt to find out if a model can also do better when this extra information is incorporated into it. Furthermore, several studies of professionally managed funds indicate that these funds do worse or the same as the market as a whole (Bauman, 1965; Fama, 1965; Jensen, 1968; O'Brien, 1970; Slovic, 1972).

Mabert (1975) reports a direct comparison between judgemental and quantitative methods of forecasting. He found that forecasts based on opinions of the sales force and corporate executives gave less accurate results over the five-year period covered by the study than did the quantitative forecasts of three methods (exponential smoothing, harmonic smoothing and Box–Jenkins). He also found that the quantitative techniques cost less and took less time than the subjective estimates. Adam and Ebert (1976), in a more complete study involving many subjects in a controlled setting, found that Winters' method produced forecasts that were statistically more accurate than those of human forecasters. This conclusion held true under different kinds of experimental variations.

When the forecasting accuracy of quantitative methods is compared, there is less agreement as to which method does best. Quantitative forecasting can be classified into two types: econometric (explanatory) and time series (mechanistic). A major reason for the popularity of ARMA models has been that several studies have found them to be at least as accurate as the complex econometric models of the USA economy. Cooper (1972, p. 920), for example, concludes that "mechanical (ARMA) forecasting models can be constructed which predict economic variables about as well as econometric models". Naylor *et al.* (1972, p. 153) report that "the Box–Jenkins results were significantly better in all cases and, except for GNP, they provide better forecasts by a factor of almost two to one". Similarly, Nelson (1972, p. 915) concludes that "if the mean square error were an appropriate measure of loss, an unweighted assessment clearly indicates that a decision maker would have been best off relying simply on ARMA predictions in the post-sample period".

Other studies by Christ (1951), Steckler (1968), Cleary and Fryk (1974), Narasimham *et al.* (1974, 1975), Cooper and Nelson (1975) and McWhorter (1975) reach similar conclusions even though the findings do not always suggest a clear superiority of ARMA models. The suggested reasons for the better performance of ARMA models are the inability of econometric models to accommodate structural changes in the economy (Christ, 1951; Steckler, 1968; Cooper, 1972; Nelson, 1972). Other authors suggest that this lack of accuracy of econometric models could be due to spurious correlations (Granger, 1969; Granger and Newbold, 1974); some even question the fundamental process of obtaining functional relationships in econometrics (Pierce, 1977). Econometricians disagree with these conclusions. They find fault with the methodology and the technicalities involved (Goldfeld, 1972; McCarthy, 1972; Howrey *et al.*,

1974) and insist that the comparisons made are unfair to econometric models, whose main purpose is not predictive but explanatory (see Leser, 1968). In at least one study (Christ, 1975) found that ARMA models are the "poorest" of those compared.

Comparisons between the accuracy of econometric models and anticipatory surveys of investment spending in the USA have also been reported. Liebling and Russell (1969) and Jorgenson *et al.* (1970) conclude that anticipation surveys were at least as successful (and frequently more so) in forecasting investment in the 1950s and 1960s. A recent study by Liebling *et al.* (1976, p. 451) reaches similar conclusions.

Within econometric models there is a fair amount of agreement as to the relative accuracy of individual techniques. Leser (1966), Jorgenson *et al.* (1970), Cooper (1972), Fromm and Klein (1973), Christ (1975) and McNees (1975) accept that no single econometric model is overwhelmingly superior to all others. McNees (1975, pp. 30–31) refers to comparisons of econometric models as follows: "Differences ... typically did not fall into systematic patterns. Valid generalizations are extremely difficult to make." Another finding seems to be that bigger econometric models with more equations do not necessarily produce more accurate forecasts than simpler models, involving only a few structural equations (von Hohenbalken and Tintner, 1962; Leser, 1966; Theil, 1966).

Comparing the relative accuracy of time series methods is more difficult because there are many more techniques to be compared and different researchers use different sets of methods when they make the comparisons. Time series methods can be placed into four groupings: smoothing, decomposition, filtering and ARMA (Makridakis, 1976). Some early studies compare the accuracy of smoothing methods, and some later ones compare smoothing and ARMA (Box–Jenkins) models. Unfortunately, nothing has been reported comparing all four groups of methods or, even better, including econometric models as well as judgemental predictions in such a comparison.

In a study reported by Kirby (1966), three different time series methods were compared: moving averages, exponential smoothing and regression (trend fitting). Kirby found that in terms of month-to-month forecasting accuracy, the exponential smoothing methods did best, with moving averages and exponential smoothing giving similar results when the forecasting horizon was increased to six months. The regression model (trend fitting) included in that study was the best method for longer-term forecasts of one year or more.

In a study reported by Levine (1967), the same three forecasting methods examined by Kirby were compared. Levine concluded that although the moving average method had the advantage of simplicity, exponential smoothing offered the best potential accuracy for short-term forecasting. Other studies reported by Gross and Ray (1965), Raine (1971) and Krampf (1972) have arrived at similar conclusions.

There is considerable disagreement concerning comparisons between ARMA and smoothing methods. Newbold and Granger (1974, p. 143) report: "The Box–Jenkins forecasts do seem to be better than those derived from two fully automatic procedures—the Holt–Winters method and stepwise autoregression—for a sizeable majority of the time series in our sample."

A similar conclusion was reached by Reid (1969, p. 266), who reported: "Thus the Box–Jenkins method is clearly better than Brown's general exponential smoothing, even when the latter is modified for serially correlated errors, for practically all series."

A study by Groff (1973, p. 30), however, arrived at a rather different conclusion: "The forecasting errors of the best of the Box–Jenkins models that were tested are either approximately equal to or greater than the errors of the corresponding exponentially smoothed models for most series."

Similarly, Guerts and Ibrahim (1975, p. 187), although they examined a single series only, found that the "exponentially smoothed models patterned on Brown's model and the Box–Jenkins approach seem to perform equally well".

Other studies dealing with accuracy (McNees, 1976; Makridakis and Wheelwright 1977) report results showing that the most accurate method varies from one set of data to another and from one time period to the next. These seemingly conflicting findings regarding the accuracy record of forecasting methods are a disturbing factor for both academics and practitioners when they have to choose between alternative forecasting methods.

2. FORECASTING ACCURACY FOR 111 TIME SERIES

The 111 time series used in this study were collected from a variety of sources, including several countries, industries and companies. These time series also represent different periods of time and time intervals (monthly, quarterly or yearly). Some of the series were seasonal and others were not. In addition, the length of both the series and seasonality (for seasonal series only) varies. This selection was made in order to minimize any bias that could arise from the use of a single source of data. However, it cannot be considered a random or even representative choice because the majority of the series come from French sources, have monthly time horizons and involve observations taken during the 1970s.

Taking n_j as the number of data points in the jth series, we used n_j-12 points to develop a forecasting model, and subsequently 12 forecasts were obtained. The error, e_{tj}, is defined as

$$e_{tj} = X_t - \hat{X}_{tj}, \quad t = n_j-11, n_j-10, \ldots, n_j, \tag{1}$$

where X_t is the actual value at period t, $\hat{X}_{tj}$ is the value forecast by the jth method, and e_{tj} is the forecast error. Accuracy, undoubtedly, is related negatively to e_t, except that it is up to the decision maker to assign different loss functions involving e_t. The most common measures of accuracy are the mean square error (m.s.e.), Theil's U-coefficient (Theil, 1966) and the mean absolute percentage error (m.a.p.e.).

The m.s.e. involves a quadratic loss function and is preferred when more weight is to be given to big errors. Its disadvantage is that it does not allow for comparisons across methods, since it is an absolute measure related to a specific series. The U-coefficient is a relative measure, it assumes a quadratic loss function and allows comparisons with the naive ($\hat{X}_t = X_{t-1}$) or random walk model. In addition, it has several other properties that make its use attractive (Theil, 1966, pp. 21–36). Its disadvantage is that its interpretation is more difficult than the m.a.p.e. Moreover, the U-statistic has no upper bound, so a few very large values can easily distort the comparisons. We will present the forecasting errors for the 111 series when both Theil's U-statistic and the m.a.p.e. were used. In addition, a ranking of the percentage of time that Naive 1, Naive 2 and ARMA are better than all other methods (without respect to the magnitude of the error) will be given.

A forecasting model using n_j-12 data points was estimated for each of twelve time series methods and one naive method (Naive 1, $\hat{X}_t = X_{t-1}$). In addition, the data (n_j-12 values) were deseasonalized and a forecasting model involving the eight nonseasonal methods was estimated. Finally, Naive 2 was defined as the most recent seasonally adjusted observation ($\hat{X}_t = X'_{t-1}$ where X'_{t-1} is seasonally adjusted). In all, forecasting models were developed for twenty-two methods (see Appendix A for a description of each method and the technical details of developing a forecasting model and obtaining predictions). No attempt was made to choose an appropriate method for each time series. (That is not necessary for several methods—Harrison's, Winters', adaptive filtering and ARMA—that can deal with all types of data.) The major difficulty arises with nonseasonal methods when the data involved are seasonal. In these latter cases, however, the data were seasonally adjusted (nonseasonal data obviously did not need adjustment.) After this adjustment there was no need to choose an appropriate method for each time series and introduce biases into the results through such a choice.

Once a forecasting model was developed, several m.a.p.e. measures were computed. First, the m.a.p.e. of fitting a model to the data was found by the following equation:

$$\text{m.a.p.e.}_{oj} = \frac{1}{n_j-12} \sum_{t=1}^{n_j-12} \frac{|X_t - \hat{X}_{tj}|}{X_t} (100), \quad j = 1, 2, 3, \ldots, 22, \tag{2}$$

where $\hat{X}_{tj}$ is the one-period-ahead forecast of period t by the jth method. The m.a.p.e.$_{oj}$ tells us how well method j does in fitting a model to *existing data* and is of questionable value since some model can always be found (as long as n_j is finite) to make equation (2) equal to zero.

Mean absolute percentage errors up to period k ($k = 1, 2, 3, 4, 5, 6, 9, 12$) were also calculated for each of the twenty-two methods and the 111 series. These errors, denoted as m.a.p.e.$_{ij}$, are in effect average absolute cumulative errors from period $n-11$ to $n-11+k$—that is, the average cumulative forecasting error of up to k periods ahead. In our opinion, this way of measuring accuracy is of higher practical relevance than multiple lead-time forecasting errors (for example, in budgeting, production planning, inventory management and so on.) Furthermore, no research has been carried out, establishing that multiple lead-time forecasts produce smaller errors than average cumulative ones.

The mean absolute percentage errors of forecasting for each of 111 series were found by:

$$\text{m.a.p.e.}_{ij} = \frac{1}{i}\sum_{k=1}^{i}\frac{|X_{n-12+k} - \hat{X}_{n-12+k,j}|}{X_{n-12+k}}(100) \tag{3}$$

where $i = 1, 2, 3, 4, 5, 6, 9, 12$; $j = 1, 2, 3, \ldots, 22$.

Table 1 shows the average of the mean absolute percentage errors for the fitted models and m.a.p.e.s for 1, 2, 3, 4, 5, 6, 9 and 12 periods ahead. The average m.a.p.e.s have been calculated over all 111 series as:

$$\text{average}\quad \text{m.a.p.e.}_{ij} = \frac{1}{111}\sum_{s=1}^{111}\text{m.a.p.e.}_{sij}, \tag{4}$$

where $i = 1, 2, 3, 4, 5, 6, 9, 12$; $j = 1, 2, 3, \ldots, 22$.

Table 1 indicates that if a single user had to forecast for all 111 series, he would have achieved the best results by using exponential smoothing methods after adjusting the data for seasonality (the seasonal indices being calculated by a decomposition method—see Appendix B). Thus a combination of decomposition (which cannot provide forecasts on its own) and exponential smoothing would have produced the best results (assuming a linear loss error function). The table also shows that the m.a.p.e.s of simpler methods are surprisingly close to those of the more statistically sophisticated ones. A final observation is that if the trend fitting and harmonic smoothing methods are excluded, the m.a.p.e.s of the remaining methodologies within each category (nonseasonal, seasonal and seasonally adjusted) are rather similar.

In order to get a wider perspective on these results, we should ask if the m.a.p.e.s of Table 1 are in agreement with those reported elsewhere. Unfortunately, not many direct comparisons are possible. To our knowledge, no one has used seasonally adjusted data with the exception of Geurts and Ibrahim (1975) who tested only a single series. Their conclusion agrees with the results of Table 1. Groff (1973) used more variations of exponential smoothing models than we did; however, both the magnitude of percentage errors he found and his conclusion that exponential smoothing models are at least as good as the Box–Jenkins methodology agree with our conclusion about the comparison between Winters' exponential smoothing and the ARMA models (the only two similar methods employed by both studies). There is also a certain amount of agreement between our conclusion and those of Reid (1969) regarding Harrison's harmonic smoothing*, in that both feel that its performance is disappointingly bad. There is further agreement in the findings of Gross and Ray (1965), Kirby (1966) and Levine (1967), and our own findings concerning the fact that exponential smoothing methods are in general more accurate than moving average ones in forecasts up to five or six periods ahead, after which they become the same or worse. Rather striking differences appear between our conclusion and those of Newbold and Granger (1974), and Reid (1969), who have found that Box–Jenkins models do better in most cases than exponential smoothing ones.

* It should be noted that it was *not* the improved version of Harrison's smoothing (i.e. SEATREND) that was used in this study (see Appendix A).

We tried classifying the data into categories to see if that would produce significantly different results. The classification involving the extent of randomness—that is, fluctuations that cannot be classified as seasonal or caused by changes in the level of economic activity—produced the most significant differences. Table 2, for example, shows the averages of the m.a.p.e. for 64 series whose randomness is less than 10 per cent. Comparing Tables 1 and 2, we see that ARMA does relatively better in Table 2 than in Table 1. If we go one step further and compute the average m.a.p.e.s of the 31 series with less than 5 per cent randomness (see Table 3), we see that ARMA models do considerably better than Naive 1 and better than Winters' method. This result is consistent with the findings of Newbold and Granger (1974) that ARMA models do better for shorter forecasting horizons, even though they did not do as well in percentages in our studies (see Tables 4, 5 and 6) as Newbold and Granger (1974) found or as Reid (1971) reported.

Tables 1, 2 and 3 leave little doubt that the extent of randomness in these 111 data series does influence the relative performance of forecasting methods. Furthermore, it could explain discrepancies in research findings. Groff (1973) used sales series whose randomness is high (the extent of randomness could be even higher in our series, many of which include data from the 1974–75 major recession), and his conclusions are in agreement with those of Table 1. Reid (1969, 1971) used macro-series whose randomness is even smaller than those in Table 3. His results are not in perfect agreement with those of Table 3, but they are in the same direction. Newbold and Granger (1974) do not give m.a.p.e. measures nor do they provide a breakdown of the macro- and micro-series used, or indications of the magnitude of percentage errors involved; however, we believe that their series were mostly of a macro type, having little randomness, but more than that in the series used by Reid. It is interesting to note that the percentage of times the Box–Jenkins method did better than exponential smoothing is much higher in Reid's study than in Newbold and Granger's. The latter results should be more in agreement with those of Table 3.

The m.a.p.e. is one of several alternative measures of expressing accuracy. Tables 4, 5 and 6 use another approach involving relative ranking of the methods used in relation to Naive 1, Naive 2 and the ARMA models respectively. They show the percentage of time that Naive 1, Naive 2 or ARMA is better than the remaining methods. Such a comparison takes no account of the magnitude of the errors; it only indicates when the error of one method is smaller than that of the other. (The comparisons shown in Tables 4, 5 and 6 are given because previous research findings have been expressed in a similar form.)

The final comparison uses Theil's U-statistic, which is based on a quadratic loss function, and provides a comparison over a "no change" model—that is, the Naive 1 or random walk method. Tables 7 and 8 show the average of U-coefficients for all series, 64 series whose randomness is less than 10 per cent and 31 series with less than 5 per cent randomness. A value of 1 in these coefficients means that Naive 1 is as good as the method it is compared with. A U value of greater than 1 indicates that Naive 1 does better than the forecasting method it is compared with. Finally, when U is smaller than 1, the other method does better than Naive 1. Table 8 differs from Table 7 in that the U-coefficients have been adjusted (any value greater than 2 has been set to 2). This adjustment allows for more meaningful comparisons, since a few high U-values tend to distort the averages shown in Table 7.

The next section will formally test the hypothesis that randomness is indeed an important factor influencing accuracy and the relative performance of different forecasting methods. Furthermore, it will look for additional factors that affect accuracy and relative performance, and will attempt to specify and measure such relationships through regression equations.

3. EXPLAINING VARIATIONS IN THE ACCURACY OF FORECASTING

Section 2 has shown that at least one factor (randomness) affects accuracy and may explain differences in the relative performance of various methods. To go one step further, we

TABLE 1

The average of the mean absolute percentage errors (m.a.p.e.) of all series (111)

	Forecasting Method	Model Fitting	Forecasting Horizons 1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1	21.9	15.5	18.4	20.4	27.9	28.8	28.6	32.2	34.1
	2 Single moving average	19.5	13.8	16.4	18.7	27.2	28.2	27.8	30.7	32.3
	3 Single exponential smoothing	19.5	14.4	16.6	19.0	27.3	28.1	27.9	31.3	33.3
	4 Adaptive response rate exponential smoothing	21.2	13.5	15.4	18.0	25.8	26.4	26.0	28.6	30.5
	5 Linear moving average	22.2	17.1	20.3	23.6	34.2	36.5	37.1	44.1	49.6
	6 Brown's linear exponential smoothing	20.2	13.2	15.8	18.4	26.5	27.7	27.3	31.2	34.7
	7 Holt's (2 parameters) linear exp. smoothing	20.5	13.3	15.6	18.1	26.2	27.7	27.5	30.5	32.5
	8 Brown's quadratic exponential smoothing	20.8	13.6	15.9	18.1	26.2	28.4	29.0	36.4	43.3
	9 Linear trend (regression fit)	22.5	19.0	19.8	22.3	30.8	31.3	30.6	34.8	38.0
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	11.0	26.4	26.3	27.6	27.4	28.0	29.3	32.2	34.2
	11 Winters' linear and seasonal exp. smoothing	10.9	13.8	14.8	15.4	16.2	17.1	18.4	21.3	23.6
	12 Adaptive filtering	11.7	15.6	16.7	16.8	18.9	18.7	19.5	22.9	24.5
	13 Autoregressive moving average (Box-Jenkins)	10.6	14.7	15.0	15.7	16.6	17.1	18.1	21.6	24.3
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	10.0	14.5	15.0	15.1	15.3	15.6	16.6	19.0	21.0
	15 Single moving average	8.4	12.9	13.6	13.7	13.8	14.3	15.3	17.7	19.8
	16 Single exponential smoothing	8.5	12.8	13.4	13.8	14.0	14.3	15.6	18.1	20.2
	17 Adaptive response rate exponential smoothing	9.2	13.0	14.0	14.5	14.7	15.2	16.2	18.5	20.4
	18 Linear moving average	9.1	15.0	15.6	16.3	16.6	17.4	18.6	22.6	26.4
	19 Brown's linear exponential smoothing	8.5	12.9	14.3	14.6	14.9	15.9	17.1	20.3	23.5
	20 Holt's (2 parameters) linear exp. smoothing	9.0	12.0	12.8	13.2	13.7	14.8	16.0	19.7	23.0
	21 Brown's quadratic exponential smoothing	8.7	12.5	14.0	14.7	15.6	17.0	18.6	23.6	28.9
	22 Linear trend (regression fit)	11.4	19.6	20.4	21.1	21.1	21.9	22.8	25.3	27.4

TABLE 2

The average of the mean absolute percentage errors (m.a.p.e.) for series whose randomness is less than 10 *per cent* (64 *series*)

	Forecasting Method	Model Fitting	Forecasting Horizons 1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1	10.4	9.3	11.0	11.4	12.6	13.6	13.8	17.3	18.6
	2 Single moving average	9.6	8.4	10.6	11.4	12.5	13.5	13.7	17.0	18.1
	3 Single exponential smoothing	9.5	8.5	10.4	11.2	12.3	13.3	13.5	16.9	18.2
	4 Adaptive response rate exponential smoothing	10.3	8.9	10.4	11.2	12.4	13.4	14.0	17.2	18.2
	5 Linear moving average	10.5	9.1	12.7	14.1	15.8	17.4	18.2	23.6	26.9
	6 Brown's linear exponential smoothing	9.6	7.8	10.4	11.6	12.8	14.2	14.4	18.4	20.8
	7 Holt's (2 parameters) linear exp. smoothing	9.8	7.7	10.4	11.2	12.4	13.5	13.7	16.7	17.9
	8 Brown's quadratic exponential smoothing	9.9	8.8	11.2	12.7	14.3	16.4	17.8	24.4	30.4
	9 Linear trend (regression fit)	10.7	15.3	15.5	16.1	17.2	18.3	18.6	22.4	23.4
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	6.2	16.6	17.2	17.7	18.1	18.7	19.5	21.8	23.1
	11 Winters' linear and seasonal exp. smoothing	5.9	7.2	8.0	8.4	9.2	10.0	11.0	13.6	15.4
	12 Adaptive filtering	5.9	6.7	8.1	8.8	9.5	10.1	10.7	13.2	14.4
	13 Autoregressive moving average (Box-Jenkins)	5.3	6.5	7.3	8.1	8.7	9.5	10.3	14.0	16.8
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	5.2	6.5	6.5	6.9	7.5	8.2	8.9	11.5	13.6
	15 Single moving average	4.6	6.4	6.9	7.3	7.8	8.6	9.3	11.8	13.8
	16 Single exponential smoothing	4.6	6.3	6.7	7.1	7.5	8.3	9.0	11.5	13.6
	17 Adaptive response rate exponential smoothing	5.2	6.6	7.6	8.0	8.7	9.6	10.6	13.1	14.8
	18 Linear moving average	4.7	8.2	9.0	9.5	10.0	11.0	11.8	14.8	18.1
	19 Brown's linear exponential smoothing	4.4	6.4	7.5	8.1	8.6	9.6	10.3	13.0	15.7
	20 Holt's (2 parameters) linear exp. smoothing	4.6	5.8	7.1	7.7	8.2	9.2	10.0	12.9	15.8
	21 Brown's quadratic exponential smoothing	4.5	6.2	7.7	8.4	9.3	10.7	11.8	16.2	20.9
	22 Linear trend (regression fit)	6.4	14.0	13.7	14.0	14.4	15.3	16.0	18.3	19.7

TABLE 3

The average of the mean absolute percentage errors (m.a.p.e.) for series whose randomness is less than 5 per cent (31 *series*)

	Forecasting Method	Model Fitting	Forecasting Horizons 1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1	5.0	4.9	5.9	6.8	7.8	8.8	10.2	11.9	12.6
	2 Single moving average	4.8	4.7	5.7	6.6	7.6	8.7	10.2	11.8	12.5
	3 Single exponential smoothing	4.9	4.7	5.7	6.6	7.7	8.8	10.2	11.8	12.6
	4 Adaptive response rate exponential smoothing	5.9	6.0	7.6	7.9	8.8	9.6	10.9	12.2	13.0
	5 Linear moving average	2.5	3.7	4.0	4.9	5.5	6.3	7.3	9.1	10.0
	6 Brown's linear exponential smoothing	4.7	4.6	6.0	6.8	8.0	9.2	10.3	13.1	15.3
	7 Holt's (2 parameters) linear exp. smoothing	4.7	4.8	5.9	6.6	7.6	8.6	9.9	11.0	11.4
	8 Brown's quadratic exponential smoothing	5.1	4.9	6.9	8.9	11.4	14.0	16.9	24.2	32.1
	9 Linear trend (regression fit)	6.7	10.7	10.7	10.8	11.0	11.7	12.8	13.7	14.4
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	4.3	9.4	9.6	10.2	10.5	10.9	11.5	12.3	12.9
	11 Winters' Linear and seasonal exp. smoothing	3.5	3.3	3.7	4.8	5.4	6.0	6.8	8.0	8.6
	12 Adaptive filtering	3.0	3.6	4.3	5.0	5.5	5.9	6.6	7.3	7.7
	13 Autoregressive moving average (Box-Jenkins)	2.5	3.1	3.3	4.3	4.8	5.5	6.4	8.3	10.5
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	2.9	2.7	3.2	4.3	5.1	6.0	6.9	8.7	10.1
	15 Single moving average	2.7	3.3	3.5	4.6	5.4	6.2	7.1	8.9	10.2
	16 Single exponential smoothing	2.8	2.9	3.4	4.4	5.2	6.0	7.0	8.8	10.1
	17 Adaptive response rate exponential smoothing	3.3	3.2	4.4	5.3	6.3	7.0	8.1	9.7	11.0
	18 Linear moving average	2.5	3.7	4.0	4.9	5.5	6.3	7.3	9.1	10.0
	19 Brown's linear exponential smoothing	2.3	3.1	3.4	4.1	4.7	5.3	6.0	7.2	7.7
	20 Holt's (2 parameters) linear exp. smoothing	2.4	2.7	3.2	3.9	4.5	5.1	5.9	7.2	8.1
	21 Brown's quadratic exponential smoothing	2.4	3.1	3.5	4.2	5.0	5.9	6.9	9.4	11.7
	22 Linear trend (regression fit)	4.5	9.5	9.4	9.9	10.2	10.7	11.3	12.2	13.1

TABLE 4

Percentage of time that Naive 1 *is better than other methods listed* (111 *series*)

	Forecasting Method	Model Fitting	Forecasting Horizons 1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1									
	2 Single moving average	6.4	34.5	30.9	34.5	37.3	37.3	37.3	36.4	30.9
	3 Single exponential smoothing	8.2	35.5	30.9	35.5	37.3	38.2	38.2	36.4	35.5
	4 Adaptive response rate exponential smoothing	39.1	41.8	47.3	50.0	54.5	53.6	55.5	50.9	50.0
	5 Linear moving average	53.6	53.6	55.5	58.2	60.9	62.7	61.8	62.7	65.5
	6 Brown's linear exponential smoothing	26.4	43.6	47.3	48.2	52.7	52.7	54.5	51.8	56.4
	7 Holt's (2 parameters) linear exp. smoothing	30.0	43.6	48.2	48.2	49.1	50.9	51.8	50.0	44.5
	8 Brown's quadratic exponential smoothing	34.5	43.6	44.5	46.4	46.4	50.0	50.9	54.5	56.4
	9 Linear trend (regression fit)	44.5	57.3	56.4	55.5	55.5	55.5	58.2	50.9	52.7
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	18.2	55.5	57.3	60.9	54.5	57.3	56.4	50.9	50.0
	11 Winters' linear and seasonal exp. smoothing	12.7	42.7	45.5	48.2	40.0	40.9	41.8	32.7	30.9
	12 Adaptive filtering	9.1	41.8	43.6	44.5	40.0	36.4	41.8	38.2	36.4
	13 Autoregressive moving average (Box-Jenkins)	1.8	40.9	36.4	42.7	36.4	35.5	40.9	36.4	32.7
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	1.8	34.5	35.5	28.2	23.6	25.5	27.3	19.1	17.3
	15 Single moving average	1.8	34.5	37.3	32.7	30.0	30.9	31.8	23.6	20.9
	16 Single exponential smoothing	1.8	32.7	35.5	31.8	30.9	32.7	33.6	26.4	21.8
	17 Adaptive response rate exponential smoothing	13.6	38.2	45.5	43.6	38.2	36.4	37.3	30.9	29.1
	18 Linear moving average	7.3	47.3	47.3	47.3	40.0	39.1	40.0	33.6	33.6
	19 Brown's linear exponential smoothing	1.8	37.3	41.8	39.1	31.8	35.5	39.1	32.7	30.9
	20 Holt's (2 parameters) linear exp. smoothing	2.7	34.5	38.2	34.5	30.9	30.0	36.4	27.3	25.5
	21 Brown's quadratic exponential smoothing	3.6	35.5	41.8	38.2	35.5	33.6	37.3	35.5	37.3
	22 Linear trend (regression fit)	21.8	53.6	54.5	51.8	45.5	45.5	47.3	39.1	36.4

TABLE 5

Percentage of time that Naive 2 is better than other methods listed (111 *series*)

	Forecasting Method	Model Fit-ting	Forecasting Horizons 1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1	82.7	50.0	49.1	56.4	60.9	59.1	57.3	65.5	67.3
	2 Single moving average	78.2	51.8	53.6	53.6	60.9	62.7	60.9	65.5	67.3
	3 Single exponential smoothing	79.1	50.9	51.8	56.4	62.7	61.8	62.7	66.4	70.9
	4 Adaptive response rate exponential smoothing	94.5	52.7	56.4	58.2	67.3	69.1	69.1	73.6	74.5
	5 Linear moving average	85.5	60.9	68.2	65.5	70.0	75.5	71.8	72.7	71.8
	6 Brown's linear exponential smoothing	80.9	47.3	54.5	58.2	63.6	67.3	63.6	73.6	78.2
	7 Holt's (2 parameters) linear exp. smoothing	83.6	45.5	53.6	54.5	62.7	64.5	60.9	68.2	70.0
	8 Brown's quadratic exponential smoothing	81.8	52.7	55.5	56.4	60.0	65.5	61.8	67.3	74.5
	9 Linear trend (regression fit)	87.3	65.5	61.8	64.5	68.2	72.7	70.0	72.7	75.5
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	54.5	63.6	70.9	71.8	69.1	70.0	70.9	71.8	70.0
	11 Winters' Linear and Seasonal Exp. Smoothing	67.3	49.1	57.3	54.5	51.8	53.6	55.5	57.3	57.3
	12 Adaptive Filtering	71.8	47.3	60.0	60.0	61.8	62.7	60.9	61.8	60.0
	13 Autoregressive moving average (Box-Jenkins)	59.1	46.4	53.6	55.5	57.3	51.8	58.2	61.8	60.9
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	5.5	33.6	39.1	40.0	39.1	43.6	42.7	42.7	37.3
	15 Single moving average	4.5	40.9	44.5	50.0	47.3	47.3	48.2	48.2	48.2
	16 Single exponential smoothing	40.0	42.7	59.1	63.6	63.6	64.5	67.3	62.7	60.0
	17 Adaptive response rate exponential smoothing	32.7	53.6	58.2	60.0	61.8	61.8	63.6	62.7	66.4
	18 Linear moving average	17.3	45.5	52.7	55.5	55.5	56.4	56.4	58.2	59.1
	19 Brown's linear exponential smoothing	18.2	43.6	47.3	49.1	50.0	51.8	54.5	51.8	50.9
	20 Holt's (2 parameters) linear exp. smoothing	20.0	46.4	54.5	55.5	56.4	55.5	53.6	61.8	61.8
	21 Brown's quadratic exponential smoothing	49.1	61.8	66.4	65.5	61.8	63.6	61.8	60.9	59.1
	22 Linear trend (regression fit)									

TABLE 6

Percentage of time that the ARMA *method is better than other methods listed* (111 *series*)

	Forecasting Method	Model Fitting	Forecasting Horizons 1	2	3	4	5	6	9	12
Original Data: Nonseasonal Methods	1 Naive 1	98.2	59.1	63.6	57.3	63.6	64.5	59.1	63.6	67.3
	2 Single moving average	92.7	49.1	56.4	53.6	60.9	64.5	61.8	65.5	66.4
	3 Single exponential smoothing	90.9	53.6	57.3	55.5	60.0	62.7	62.7	65.5	66.4
	4 Adaptive response rate exponential smoothing	95.5	49.1	56.4	56.4	60.0	62.7	64.5	62.7	67.3
	5 Linear moving average	97.3	63.6	63.6	63.6	69.1	70.9	70.0	70.0	70.9
	6 Brown's linear exponential smoothing	92.7	42.7	51.8	50.0	53.6	59.1	59.1	67.3	70.9
	7 Holt's (2 parameters) linear exp. smoothing	90.9	50.0	55.5	50.0	54.5	62.7	61.8	62.7	66.4
	8 Brown's quadratic exponential smoothing	91.8	48.2	51.8	50.9	57.3	61.8	63.6	65.5	73.6
	9 Linear trend (regression fit)	90.9	60.9	61.8	63.6	65.5	69.1	68.2	71.8	72.7
Seasonal and Non-seasonal Methods	10 Harrison's harmonic smoothing	49.1	62.7	68.2	72.7	70.0	70.0	73.6	71.8	65.5
	11 Winters' linear and seasonal exp. smoothing	69.1	47.3	49.1	51.8	50.0	50.0	50.9	46.4	45.5
	12 Adaptive filtering	70.9	49.1	54.5	52.7	55.5	56.4	56.4	58.2	51.8
	13 Autoregressive moving average (Box-Jenkins)									
Seasonally Adjusted Data: Nonseasonal Methods	14 Naive 2	40.9	53.6	46.4	44.5	42.7	48.2	41.8	38.2	39.1
	15 Single moving average	19.1	46.4	42.7	43.6	39.1	42.7	39.1	37.3	38.2
	16 Single exponential smoothing	19.1	50.0	42.7	42.7	38.2	43.6	39.1	41.8	40.0
	17 Adaptive response rate exponential smoothing	30.9	44.5	46.4	42.7	40.9	47.3	48.2	44.5	41.8
	18 Linear moving average	26.4	55.5	50.9	50.9	49.1	47.3	49.1	51.8	51.8
	19 Brown's linear exponential smoothing	19.1	42.7	48.2	44.5	44.5	49.1	47.3	49.1	46.4
	20 Holt's (2 parameters) linear exp. smoothing	20.9	41.8	43.6	42.7	40.0	46.4	44.5	46.4	45.5
	21 Brown's quadratic exponential smoothing	20.0	41.8	44.5	44.5	45.5	47.3	47.3	50.9	59.1
	22 Linear trend (regression fit)	41.8	63.6	59.1	62.7	60.0	59.1	57.3	53.6	50.0

TABLE 7

Theil's U-coefficient

		Forecasting Method		Model fitting	Forecasting Horizons 1	2	3	4	5	6	9	12	Average
Original Data: Nonseasonal Methods	1	Naive 1	* A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
			10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
			5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2	Single moving average	A	0.88	1.46	1.37	1.46	1.51	1.84	1.37	1.20	2.14	1.54
			10	0.91	1.47	1.46	1.47	1.48	1.46	1.48	1.08	2.73	1.58
			5	0.95	1.77	1.04	0.97	1.00	1.19	1.02	0.97	1.01	1.12
	3	Single exponential smoothing	A	0.88	1.47	1.27	1.35	1.43	1.60	1.15	1.07	1.15	1.31
			10	0.91	1.26	1.32	1.31	1.27	1.27	1.14	1.00	1.13	1.21
			5	0.97	1.41	1.01	0.99	1.02	1.06	1.00	1.00	1.00	1.06
	4	Adaptive response rate exponential smoothing	A	0.97	1.32	1.49	1.54	1.75	1.62	1.60	1.43	2.41	1.64
			10	1.02	1.13	1.61	1.53	1.97	1.60	1.94	1.17	2.68	1.70
			5	1.13	1.14	1.87	1.06	1.38	1.27	1.16	1.25	1.10	1.28
	5	Linear moving average	A	0.96	2.56	2.59	2.85	2.96	4.51	2.52	5.13	10.78	4.24
			10	0.96	1.76	3.02	2.83	3.08	3.16	2.10	4.27	11.49	3.96
			5	0.95	2.07	3.55	2.56	2.40	3.11	2.20	6.10	4.84	3.35
	6	Brown's linear exponential smoothing	A	0.86	1.49	1.79	1.85	1.79	2.04	1.68	2.39	4.93	2.24
			10	0.87	1.15	2.17	2.12	2.05	2.13	2.04	2.78	6.59	2.63
			5	0.86	1.50	2.51	1.68	1.93	2.35	1.81	4.38	3.41	2.45
	7	Holt's (2 parameters) linear exp. smoothing	A	0.87	1.16	1.58	1.68	1.71	2.80	1.73	1.65	5.23	2.19
			10	0.87	0.98	1.79	1.83	1.95	2.01	1.89	1.43	7.07	2.37
			5	0.84	1.15	1.89	1.28	1.59	2.22	1.65	1.51	1.66	1.62
	8	Brown's quadratic exponential smoothing	A	0.90	1.46	1.89	2.15	2.34	2.90	1.93	4.88	9.51	3.38
			10	0.92	1.49	2.33	2.53	3.02	2.69	2.17	6.26	12.15	4.08
			5	0.93	1.94	2.73	2.40	4.03	3.73	2.29	10.90	11.77	4.98
	9	Linear trend (regression fit)	A	1.03	2.84	2.57	2.44	2.45	3.04	2.12	1.86	3.67	2.62
			10	1.16	3.01	3.00	2.87	2.87	2.75	2.48	1.60	4.32	2.86
			5	1.45	3.69	4.18	2.36	1.91	2.28	1.67	1.95	1.57	2.45
Seasonal and Non-seasonal Methods	10	Harrison's harmonic smoothing	A	0.75	6.61	3.60	3.55	3.43	5.04	2.82	2.49	5.92	4.18
			10	0.88	7.48	3.75	4.39	3.68	3.18	3.08	2.05	8.12	4.47
			5	1.16	11.97	4.08	3.96	2.57	2.20	2.54	2.27	1.98	3.95
	11	Winters' linear and seasonal exp. smoothing	A	0.68	1.47	1.76	1.55	2.21	2.66	1.65	1.76	3.53	2.07
			10	0.69	1.36	1.75	1.52	2.27	2.12	1.79	1.60	4.82	2.15
			5	0.78	1.58	1.88	1.35	1.84	2.54	1.57	1.88	1.19	1.73
	12	Adaptive filtering	A	0.69	1.93	1.97	1.46	2.23	2.26	1.67	1.74	3.37	2.08
			10	0.70	1.36	1.82	1.47	2.22	2.41	1.69	1.63	4.36	2.12
			5	0.75	1.41	2.14	1.13	1.63	2.66	1.37	1.50	1.45	1.66
	13	Autoregressive moving average (Box-Jenkins)	A	0.61	1.51	1.68	1.73	1.91	2.22	1.77	1.88	3.04	1.97
			10	0.58	1.45	1.45	1.50	1.82	1.87	1.69	1.75	4.17	1.96
			5	0.57	1.91	1.48	1.36	1.63	2.37	1.44	1.78	1.21	1.65
Seasonally Adjusted Data: Nonseasonal Methods	14	Naive 2	A	0.66	1.62	1.09	1.14	1.30	1.68	1.20	1.17	1.00	1.28
			10	0.64	1.03	0.92	1.14	1.40	1.31	1.32	1.14	1.00	1.16
			5	0.70	0.99	0.72	1.10	1.27	1.49	1.04	1.28	1.00	1.11
	15	Single moving average	A	0.57	1.61	1.35	1.25	1.47	1.93	1.30	1.35	2.87	1.64
			10	0.57	1.42	1.15	1.24	1.45	1.43	1.33	1.17	3.91	1.64
			5	0.66	1.87	0.80	1.14	1.28	1.60	1.05	1.31	1.00	1.26
	16	Single exp. smoothing	A	0.57	1.66	1.27	1.13	1.32	1.68	1.23	1.25	1.05	1.32
			10	0.58	1.22	1.13	1.03	1.36	1.37	1.31	1.14	1.05	1.20
			5	0.68	1.51	0.77	1.11	1.27	1.49	1.04	1.28	1.00	1.18
	17	Adaptive response rate exponential smoothing	A	0.63	1.59	1.61	1.39	1.83	2.17	1.58	1.65	3.11	1.87
			10	0.66	1.09	1.62	1.31	2.02	1.82	1.74	1.34	3.92	1.86
			5	0.79	1.01	1.61	1.23	1.50	1.70	1.26	1.51	1.12	1.37
	18	Linear moving average	A	0.58	1.77	2.30	1.80	1.96	2.29	2.01	2.41	7.19	2.72
			10	0.56	1.43	2.67	1.79	2.39	2.32	2.38	1.92	10.35	3.16
			5	0.58	1.51	3.00	1.16	1.77	2.44	1.53	1.92	2.36	1.96
	19	Brown's linear exponential smoothing	A	0.53	1.49	1.95	1.51	1.88	2.41	1.55	1.64	5.42	2.23
			10	0.51	1.14	2.09	1.52	2.02	2.19	1.67	1.49	7.94	2.51
			5	0.53	1.41	2.11	1.02	1.62	2.15	1.31	1.57	1.21	1.55
	20	Holt's (2 parameters) linear exp. smoothing	A	0.55	1.41	1.57	1.37	1.79	2.59	1.52	1.61	4.95	2.10
			10	0.52	0.94	1.65	1.46	1.96	2.06	1.69	1.21	6.47	2.18
			5	0.53	1.06	1.51	1.03	1.64	2.04	1.47	1.15	1.34	1.40
	21	Brown's quadratic exponential smoothing	A	0.55	1.67	1.96	1.59	2.11	3.03	1.81	1.67	6.13	2.50
			10	0.53	1.22	2.14	1.65	2.34	2.68	2.06	1.62	8.71	2.80
			5	0.56	1.60	2.18	1.15	1.67	2.59	1.17	1.33	2.25	1.74
	22	Linear trend (regression fit)	A	0.76	2.96	2.85	2.60	2.48	3.68	2.13	2.23	5.13	3.01
			10	0.88	2.86	3.09	3.00	2.81	3.06	2.26	1.97	6.94	3.25
			5	1.19	3.30	4.05	2.68	1.88	2.52	1.70	2.46	1.25	2.48

* A = all series (111).
10 = series whose randomness is less than 10% (64).
5 = series whose randomness is less than 5% (31).

TABLE 8

Theil's U-coefficient adjusted

		Forecasting Method		Model fitting	Forecasting Horizons 1	2	3	4	5	6	9	12	Average
Original Data: Nonseasonal Methods	1	Naive 1	*A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
			10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
			5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2	Single moving average	A	0.88	1.01	1.05	1.06	1.07	1.11	1.01	1.00	1.06	1.04
			10	0.91	1.00	1.09	1.07	1.05	1.06	1.05	0.96	1.02	1.04
			5	0.95	0.99	1.01	0.97	1.00	1.03	1.02	0.97	1.01	1.00
	3	Single exponential smoothing	A	0.88	1.00	1.01	1.09	1.07	1.06	1.03	1.01	1.05	1.04
			10	0.91	0.97	1.05	1.07	1.08	1.06	1.03	0.99	1.00	1.03
			5	0.97	0.94	1.01	0.99	1.02	1.03	1.00	1.00	1.00	1.00
	4	Adaptive response rate exponential smoothing	A	0.97	1.00	1.02	1.09	1.07	1.09	1.03	0.99	1.10	1.05
			10	1.02	1.02	1.12	1.05	1.20	1.11	1.13	0.98	1.06	1.08
			5	1.13	1.05	1.30	1.00	1.19	1.09	1.06	1.03	1.07	1.10
	5	Linear moving average	A	0.96	1.16	1.18	1.29	1.28	1.33	1.25	1.34	1.43	1.28
			10	0.96	1.11	1.23	1.27	1.28	1.29	1.24	1.26	1.28	1.24
			5	0.95	1.03	1.15	1.26	1.24	1.28	1.28	1.23	1.23	1.21
	6	Brown's linear exponential smoothing	A	0.86	0.92	1.04	1.11	1.09	1.11	1.04	1.11	1.28	1.09
			10	0.87	0.89	1.09	1.13	1.14	1.17	1.12	1.06	1.20	1.10
			5	0.86	0.96	1.07	1.04	1.09	1.10	1.11	1.09	1.07	1.07
	7	Holt's (2 parameters) linear exp. smoothing	A	0.87	0.95	1.02	1.09	1.05	1.15	1.03	1.04	1.19	1.07
			10	0.87	0.89	1.09	1.10	1.10	1.07	0.99	0.96	1.11	1.04
			5	0.84	0.99	1.08	1.01	0.99	0.99	0.98	0.91	1.03	1.00
	8	Brown's quadratic exponential smoothing	A	0.90	0.96	1.01	1.11	1.12	1.25	1.16	1.22	1.43	1.16
			10	0.92	0.94	1.08	1.15	1.20	1.28	1.22	1.23	1.35	1.18
			5	0.93	0.96	1.07	1.19	1.22	1.31	1.28	1.23	1.29	1.19
	9	Linear trend (regression fit)	A	1.03	1.22	1.14	1.14	1.10	1.15	1.07	1.07	1.20	1.14
			10	1.16	1.35	1.26	1.17	1.19	1.23	1.11	1.00	1.14	1.18
			5	1.45	1.56	1.31	1.13	1.06	1.12	0.96	0.95	1.13	1.15
Seasonal and Non-seasonal Methods	10	Harrison's harmonic smoothing	A	0.75	1.18	1.22	1.23	1.14	1.23	1.21	1.09	1.27	1.20
			10	0.88	1.15	1.26	1.23	1.26	1.24	1.20	1.08	1.18	1.20
			5	1.16	1.11	1.09	1.11	1.10	1.16	0.97	0.98	1.04	1.07
	11	Winters' linear and seasonal exp. sm.	A	0.68	0.94	0.97	0.99	0.97	1.03	1.08	0.99	1.09	1.01
			10	0.69	0.88	0.96	0.98	1.03	1.01	1.08	0.93	1.04	0.99
			5	0.78	0.89	0.86	0.92	0.83	0.86	0.82	0.86	0.93	0.87
	12	Adaptive filtering	A	0.69	1.04	1.03	1.06	1.02	1.10	1.08	1.09	1.19	1.08
			10	0.70	1.01	0.98	1.08	1.05	1.03	1.04	0.95	1.10	1.03
			5	0.75	1.00	0.94	0.88	0.88	0.83	0.80	0.71	0.82	0.86
	13	Autoregressive moving average (Box-Jenkins)	A	0.61	0.97	0.94	1.01	0.97	0.96	1.03	1.03	1.10	1.00
			10	0.58	0.87	0.88	1.05	1.01	0.93	1.08	0.97	1.06	0.98
			5	0.57	0.86	0.81	1.04	0.82	0.86	0.86	0.90	0.85	0.88
Seasonally Adjusted Data: Nonseasonal Methods	14	Naive 2	A	0.66	0.94	0.87	0.92	0.89	0.94	0.98	0.94	1.00	0.93
			10	0.64	0.88	0.77	0.93	0.93	0.90	1.01	0.95	1.00	0.92
			5	0.70	0.82	0.72	0.94	0.91	0.89	0.88	0.92	1.00	0.89
	15	Single moving average	A	0.57	0.93	0.88	0.90	0.87	0.94	0.98	0.95	1.02	0.93
			10	0.57	0.86	0.83	0.95	0.94	0.93	1.02	0.95	0.97	0.93
			5	0.66	0.90	0.76	0.98	0.92	0.90	0.88	0.92	1.00	0.91
	16	Single exp. smoothing	A	0.57	0.91	0.87	0.90	0.87	0.89	0.98	0.93	1.02	0.92
			10	0.58	0.85	0.82	0.93	0.94	0.89	1.02	0.95	1.02	0.93
			5	0.68	0.87	0.76	0.95	0.91	0.89	0.88	0.92	1.00	0.90
	17	Adaptive response rate exponential smoothing	A	0.63	0.90	0.98	1.00	0.93	0.97	1.03	0.93	1.08	0.98
			10	0.66	0.83	1.01	1.01	1.08	0.95	1.12	0.95	1.03	1.00
			5	0.79	0.87	1.09	1.00	1.14	0.91	1.00	0.94	1.08	1.00
	18	Linear moving average	A	0.58	1.01	0.96	1.02	0.92	1.09	1.08	1.12	1.29	1.06
			10	0.56	0.96	0.93	0.99	1.00	1.07	1.09	1.04	1.17	1.03
			5	0.58	0.89	0.77	0.89	0.90	0.95	0.96	0.97	0.97	0.91
	19	Brown's linear exponential smoothing	A	0.53	0.89	0.90	0.95	0.93	1.02	1.04	1.01	1.16	0.99
			10	0.51	0.80	0.86	0.98	1.03	1.03	1.05	0.94	1.09	0.97
			5	0.53	0.77	0.70	0.84	0.80	0.82	0.79	0.86	0.88	0.81
	20	Holt's (2 parameters) linear exp. smoothing	A	0.55	0.85	0.84	0.90	0.85	1.02	0.96	1.02	1.16	0.95
			10	0.52	0.76	0.86	0.89	0.95	0.96	0.98	0.92	1.10	0.93
			5	0.53	0.73	0.71	0.79	0.77	0.79	0.74	0.82	0.96	0.79
	21	Brown's quadratic exponential smoothing	A	0.55	0.87	0.90	0.95	0.92	1.05	1.06	1.11	1.26	1.02
			10	0.53	0.79	0.88	0.96	1.01	1.04	1.08	1.09	1.19	1.01
			5	0.56	0.77	0.67	0.88	0.82	0.85	0.83	1.04	1.14	0.87
	22	Linear trend (regression fit)	A	0.76	1.16	1.11	1.15	1.05	1.09	1.09	1.02	1.12	1.10
			10	0.88	1.21	1.22	1.21	1.21	1.16	1.14	0.98	1.04	1.15
			5	1.19	1.39	1.20	1.19	1.13	1.05	0.88	0.89	0.96	1.09

* A = all series (111).
10 = series whose randomness is less than 10% (64),
5 = series whose randomness is less than 5% (31 series)

TABLE 9

Regression coefficients

(Dependent variable is the mean absolute percentage error (m.a.p.e.) of model fitting)

		Forecasting Method	R^2	Constant	Independent Variables				Standard Error of Estimate	F-Test
					No. of Data	Mean Absolute % Change				
						Trend Cycle	Randomness	Seasonality		
Original Data: Nonseasonal Methods	1	Naive 1	.98	0			.95 (38.29)*	1.02 (51.51)		2051
	2	Single moving average	.95	1.40 (3.24)			.69 (18.08)	.92 (33.84)	2.48	974
	3	Single exponential smoothing	.96	1.35 (3.06)			.71 (17.99)	.90 (36.80)	2.58	1081
	4	Adaptive response rate exponential smoothing	.93	2.02 (3.37)			.75 (13.94)	.95 (28.43)	3.49	646
	5	Linear moving average	.94	1.55 (2.74)			.80 (15.88)	1.04 (28.96)	3.26	724
	6	Brown's linear exponential smoothing	.95	1.04 (2.15)			.76 (17.65)	.94 (34.77)	2.83	984
	7	Holt's (2 parameters) linear exp. smoothing	.97	.79 (2.03)			.80 (23.20)	.97 (39.33)	2.26	1393
	8	Brown's quadratic exponential smoothing	.95	1.24 (2.36)			.77 (16.44)	.97 (33.43)	3.05	895
	9	Linear trend (regression fit)	.76	0		2.72 (4.83)	.64 (5.4)	.90 (14.03)	7.01	104
Seasonal and non-seasonal Methods	10	Harrison's harmonic smoothing	.76	1.06 (1.75)		1.65 (5.66)	.58 (10.54)	.10 (4.04)	3.27	108
	11	Winters' linear and seasonal exp. smoothing	.90	3.61 (5.55)	−.02 (−3.72)		.87 (27.71)	.06 (3.72)	2.00	310
	12	Adaptive filtering	.86	2.09 (2.36)	−.01 (−1.82)		.99 (22.49)	.10 (4.60)	2.70	212
	13	Autoregressive moving average (Box-Jenkins)	.84	2.30 (2.44)	−.02 (−1.86)		.94 (21.25)	.07 (3.13)	2.89	178
Seasonally Adjusted Data: Nonseasonal Methods	14	Naive 2	.96	.61 (2.94)		.33 (3.49)	.90 (46.22)		1.15	1384
	15	Single moving average	.92	.59 (2.29)		.71 (5.82)	.68 (28.22)		1.47	619
	16	Single exp. smoothing	.93	.71 (3.00)		.62 (5.49)	.69 (31.15)		1.36	730
	17	Adaptive response rate exponential smoothing	.92	1.04 (3.86)		.73 (5.76)	.72 (28.45)		1.53	627
	18	Linear moving average	.91	.58 (1.83)		.30 (2.03)	.82 (27.66)		1.80	518
	19	Brown's linear exponential smoothing	.93	.48 (1.88)		.28 (2.31)	.77 (32.04)		1.47	695
	20	Holt's (2 parameters) linear exp. smoothing	.95	.82 (3.74)			.83 (42.29)		1.34	1788
	21	Brown's quadratic exponential smoothing	.93	.55 (2.16)		.31 (2.55)	.78 (32.57)		1.46	722
	22	Linear trend (regression fit)	.72	1.72 (2.87)		1.44 (4.72)	.62 (12.29)		3.33	126

* Numbers in parentheses are the t-tests of regression coefficients.

TABLE 10

Regression coefficients

(Dependent variable is the mean absolute percentage error (m.a.p.e.) of forecasts)

		Forecasting Method	R^2	Constant	Mean Absolute % change			Absolute % Change of Trend-Cycle at Period n = 12	Number of Periods Ahead Forecasting	Standard Error of Estimate	F-Test
					Trend-Cycle	Randomness	Seasonality				
Original Data: Nonseasonal Methods	1	Naive 1	.47	-5.91 (-4.68)*	2.22 (4.81)	1.11 (11.69)	.12 (2.48)	.66 (9.74)	1.44 (9.52)	15.06	145
	2	Single moving average	.42	-4.02 (-3.47)	1.71 (4.03)	.73 (8.68)	.30 (6.96)	.45 (7.23)	1.51 (10.90)	13.80	121
	3	Single exponential smoothing	.48	-5.79 (-5.12)	2.51 (6.06)	.81 (9.84)	.25 (5.82)	.54 (8.97)	1.54 (11.36)	13.45	149
	4	Adaptive response rate exponential smoothing	.38	-2.19 (-2.02)	2.27 (5.70)	.83 (10.53)	.20 (4.78)	.13 (2.26)	1.26 (9.68)	12.89	101
	5	Linear moving average	.37	-5.69 (-3.27)		1.07 (9.05)	.63 (10.25)	.22 (2.18)	2.51 (11.42)	22.00	123
	6	Brown's linear exponential smoothing	.41	-5.02 (-4.12)	.99 (2.21)	.79 (8.87)	.29 (6.57)	.42 (6.47)	1.86 (12.73)	14.62	117
	7	Holt's (2 parameters) linear exp. smoothing	.40	-3.91 (-3.19)	1.43 (3.19)	.85 (9.06)	.18 (4.05)	.43 (6.50)	1.60 (10.84)	14.74	109
	8	Brown's quadratic exponential smoothing	.34	-5.46 (-3.68)	1.66 (3.04)	.72 (6.65)	.18 (3.45)	.37 (4.71)	2.49 (13.97)	17.76	84
	9	Linear trend (regression fit)	.36	-4.10 (-2.71)	5.30 (9.53)	.39 (3.61)	.46 (8.15)	.40 (5.06)	1.57 (8.70)	18.01	92
Seasonal and Non-seasonal Methods	10	Harrison harmonic smoothing	.38	0	5.50 (9.88)	.62 (5.48)	.37 (7.02)	.63 (7.39)	.75 (5.09)	18.80	103
	11	Winters' linear and seasonal exp. smoothing	.43	-2.45 (-2.69)	1.84 (5.52)	.65 (9.77)	.09 (3.02)	.50 (10.16)	.91 (8.42)	10.77	122
	12	Adaptive filtering	.48	-3.06 (-3.33)	1.52 (4.54)	.82 (11.85)	.17 (5.30)	.49 (9.63)	.91 (8.33)	10.95	155
	13	Autoregressive moving average (Box-Jenkins)	.43	-3.85 (-4.09)	2.08 (5.87)	.69 (10.00)	.19 (5.83)	.37 (6.97)	.98 (8.68)	11.14	123
Seasonally Adjusted Data: Nonseasonal Methods	14	Naive 2	.47	-2.33 (-2.93)	2.17 (7.83)	.78 (14.19)		.43 (10.33)	.62 (6.47)	9.48	182
	15	Single moving average	.43	-1.31 (-1.81)	1.98 (7.46)	.54 (10.27)	.15 (6.06)	.22 (5.68)	.72 (8.25)	8.65	126
	16	Single exp. smoothing	.46	-1.76 (-2.44)	2.18 (8.26)	.64 (12.07)	.12 (4.85)	.22 (5.71)	.65 (7.67)	8.48	141
	17	Adaptive response rate exponential smoothing	.41	0	2.15 (8.18)	.53 (9.44)	.14 (5.74)	.17 (4.24)	.66 (9.36)	8.89	114
	18	Linear moving average	.39	-2.57 (-2.52)	1.62 (4.32)	.68 (9.24)	.08 (2.38)	.56 (9.92)	1.07 (8.83)	12.00	105
	19	Brown's linear exponential smoothing	.46	-2.37 (-2.86)	.95 (3.13)	.77 (12.80)	.09 (3.17)	.44 (9.88)	.92 (9.28)	9.84	142
	20	Holt's (2 parameters) linear exp. smoothing	.40	-2.75 (-3.21)	1.57 (5.02)	.65 (9.97)	.08 (2.86)	.26 (5.60)	1.07 (10.46)	10.18	111
	21	Brown's quadratic exponential smoothing	.41	-3.95 (-3.88)	1.61 (4.46)	.62 (9.04)		.65 (12.38)	1.40 (11.46)	12.12	144
	22	Linear trend (regression fit)	.32	0	4.33 (9.61)	.31 (3.41)	.25 (5.99)	.46 (6.77)	.82 (6.86)	15.16	77.30

* Numbers in parentheses are the t-tests of regression coefficients.

assumed that the accuracy of a forecasting method depends upon several factors and that these factors could be isolated and quantified, and their influence measured.

In this and an earlier study (Makridakis and Vandenburgh, 1975), several sets of factors were explored for their ability to influence accuracy. Two sets produced the best regression equations (that is, statistically significant coefficients, highest R^2). Both involve the m.a.p.e.s as the dependent variable and two combinations of independent variables. The first set relates the m.a.p.e. of the fitted model, $n_j - 12$ values, to combinations of the mean absolute percentage change of the trend-cycle, randomness, seasonality (as calculated by the decomposition method shown in Appendix B) and the number of data points. The second relates the m.a.p.e.s of the forecasts (1, 2, 3, 4, 5, 6, 9 and 12 periods ahead) to the mean absolute percentage change in the trend-cycle, randomness, seasonality, the absolute percentage change of the trend-cycle at the last period of data ($n_j - 12$), and the number of periods ahead of forecasting. The coefficients of both sets of regressions and the related statistics can be seen in Tables 9 and 10 respectively. Almost all coefficients are statistically significant at a 99 per cent level; only a few exceptions involve a 95 per cent confidence level, and these usually refer to the regression constant.

In total, there are 44 regression equations. Two for each of the 22 methods—one when the m.a.p.e.s of the fitted model are the dependent variable and one when the m.a.p.e.s of the forecasts are the dependent variable.

For example, the regression equation for the m.a.p.e. of the fitted model for ARMA (see Table 9) is

$$\text{m.a.p.e.}_{o(\text{ARMA})} = 2{\cdot}30 - {\cdot}02X_1 + {\cdot}94X_3 + {\cdot}97X_4, \tag{5}$$

where X_1 is the number of data points in the jth series, X_3 is the mean absolute percentage change in randomness of the jth series, and X_4 is the mean absolute percentage change in seasonality of the jth series; the $R^2 = {\cdot}84$, $\hat{\sigma}_u = 2{\cdot}8$, and the F-statistic = 178.

Similarly, the corresponding equations of Winters' model is:

$$\text{m.a.p.e.}_{o(\text{Winters})} = 3{\cdot}61 - {\cdot}02X_1 + {\cdot}89X_3 + {\cdot}06X_4, \tag{6}$$

with an $R^2 = {\cdot}90$, $\hat{\sigma}_u = 2{\cdot}0$, and the F-statistic = 310.

In equation (5), when randomness increases by 1 per cent (all other things being equal), the m.a.p.e. of the model will increase by ·94 per cent. In equation (6), this increase is only ·89 per cent. However, the constant in equation (6) is considerably bigger than in equation (5), which also explains why ARMA models are relatively more accurate than Winters', when data with a low level of randomness are involved.

Another interesting observation is the magnitude of $\hat{\sigma}_u$. In equation (5), this value is 2·8, and in equation (6) it is 2·0. This means that the chances of going above or below the average of the regression equation are higher in equation (5) than in equation (6). In other words, a conservative forecaster who would like to take risk into account would be more inclined towards equation (6) than equation (5), all other things being equal. On the other hand, there is more room for improvement through equation (5) (the ARMA method) than through equation (6), if we assume that the user has some control over the accuracy of the time series to be forecast.

Table 10 lists the regression coefficients and related statistics of the regression equations involving the m.a.p.e.s of forecasts as the dependent variable. The form of the equation (with a few exceptions) is:

$$\text{m.a.p.e.}_{ij} = a + b_2 X_2 + b_3 X_3 + b_4 X_4 b_5 X_5 + b_6 X_6, \tag{7}$$

where a is the constant, b_2, b_3, b_4, b_5 and b_6 are the regression coefficients, X_2 is the mean absolute percentage change in the trend-cycle of the jth data series, X_3 is the mean absolute percentage change in the randomness of the jth data series, X_4 is the mean absolute percentage change in the seasonality of the jth data series, X_5 is the absolute percentage change in the trend-cycle of the jth data series at period $n-12$, X_6 is the number of periods ahead of forecasts (that is, 1, 2, 3, 4, 5, 6, 9 or 12). Obviously, intrapolation for the values 7, 8, 10 and 11 is possible, and maybe extrapolation beyond 12 might be justified.

The equation relating the forecasting performance of the ARMA method is

$$\text{m.a.p.e.}_{i(\text{ARMA})} = -3{\cdot}85 + 2{\cdot}08X_2 + {\cdot}69X_3 + {\cdot}19X_4 + {\cdot}37X_5 + 98X_6 \tag{8}$$

with the $R^2 = {\cdot}43$, $\hat{\sigma}_u = 11{\cdot}14$, and the F-statistic = 123.

The corresponding values of Winters' smoothing is

$$\text{m.a.p.e.}_{i(\text{Winters})} = -2{\cdot}45 + 1{\cdot}84X_2 + {\cdot}65X_3 + {\cdot}09X_4 + {\cdot}50X_5 + 91X_6 \tag{9}$$

with the $R^2 = {\cdot}43$, $\hat{\sigma}_u = 10{\cdot}77$, and the F-statistic = 122.

The parameters of equations (5) and (6), or (8) and (9), as well as those corresponding to other methodologies, could provide some revealing information about the relative forecasting performance of the various methodologies. Unlike the regression coefficients of Table 9, those of Table 10 have relatively higher standard errors, which raises questions as to the statistical significance of the differences in the coefficients. The approach of specifying and measuring the relationship, however, is valid even though its usefulness can be enlarged by the following improvements, which would require further work:

1. Some random selection of the series used must be made.
2. The sample size (number of series used) must be increased considerably.
3. More independent variables are probably required so that a higher percentage of total variations in m.a.p.e.s could be explained by the forecasting equations.
4. In addition to m.a.p.e., quadratic accuracy measures such as Theil's U-coefficients should be introduced as the dependent variable. However, there will be difficulty in specifying appropriate quadratic measures to describe the different series.

If we ignore for the moment the fact that some of the differences may not be significant, the examination of the coefficients of equations (8) and (9) can reveal that the constant corresponding to the ARMA method has a smaller value than that of Winters' model. Similarly, the coefficient corresponding to the trend-cycle change of the last period is smaller in ARMA than in the regression equation of Winters' model whose coefficients are smaller for the remaining variables. This means that the ARMA method will be relatively more accurate when the numerical values of the variables are small and when there are stronger fluctuations in the level of the economic activity (trend-cycle *in the last* period of data (that is, ARMA models will predict trend-cycle changes more accurately than Winters' model). Winters' method, on the other hand, will do better, in terms of smaller m.a.p.e., when there are high mean absolute percentage changes in trend-cycle, randomness and seasonality.

4. THE CAUSES OF VARIATIONS IN ACCURACY

After noting these variations in accuracy, we went on to examine possible causes for these differences. The m.a.p.e. of model fitting is determined by a different process from that which influences the m.a.p.e. of forecasts. The variations in the former can be precisely explained, in the sense that the R^2 values of the regression equations are usually larger than ·90 (see Table 9), by two or three factors that account for most of the changes in the level of m.a.p.e. The extent of randomness in the series (see Appendix B) seems to be the overwhelming factor influencing the accuracy of all methods used. That is not surprising, of course, except that we would expect the regression coefficients of randomness of the more sophisticated methods to be smaller than those of the simpler methods. But the opposite is true. Nonseasonal methods, as expected, do include significant coefficients for the variable "seasonality"; seasonal methods and seasonally adjusted data do not, which indicates that both groups of methods sufficiently eliminate seasonality so that it does not, statistically, influence accuracy in the model-fitting process. Apart from linear trend fitting and Harrison's harmonic smoothing, all methods that do not use seasonally adjusted data seem to predict the trend-cycle fairly well. Even Naive 1 is quite effective, which suggests that period-to-period changes in the trend-cycle are not important enough, in the fitting phase, to produce statistically significant coefficients. The equations for the seasonally adjusted data include a significant coefficient for the variable "trend-cycle". That is somewhat surprising and might be caused by the decomposition process which seems to result in nonrandom residuals (Burman, 1965; Cleveland and Tiao, 1976). Finally, only three of the more sophisticated methods have a significant coefficient for the number of data points in the series. The values of these coefficients, however, are small and the t-statistics are not very high, which indicates that the number of data points is not as important a factor as might have been thought.

Given the magnitude of R^2 of the regression equations in Table 9, it can be asserted that these equations can be used to predict the level and explain the variations in the m.a.p.e. of the model fitting fairly well. In addition, the standard errors of the estimate are not large. Thus it can be said that these equations can even be used to forecast the m.a.p.e. of the fitting phase of various series/methods. Obviously, such information can be extremely helpful in selecting forecasting methodologies, if it is assumed that the decision makers are concerned with linear errors in the model-fitting phase.

The R^2 of the regression equations for the m.a.p.e. of forecasts do not explain as much of the total variation in accuracy. (The average of all R^2 of Table 9 is ·91 while that of Table 10 is only ·41.) Furthermore, the standard errors of estimate are rather large, which makes the use of these equations unreliable for predicting the m.a.p.e. of forecasts, since the confidence limits will be large. Finally, it is surprising to see how little the factors included in the equations and the magnitude of their coefficients vary in the great majority of methods. No general patterns can be inferred, and differences between groups of methods are hard to explain from a statistical point of view. There are two additional factors of importance in the equations in Table 10: the first is the absolute percentage change in the trend-cycle in the last period of data $(n-12)$. This factor does make sense and is consistent with experience and findings in the literature (McNees, 1976; Makridakis and Majani, 1977); it is included as a sign of impending change in the level of economic activity. The second factor is the number of periods ahead of forecasts $(k = 1, 2, 3, 4, 5, 6, 9, 12)$. The interpretation of this factor is obvious in that as the time horizon of forecasting increases, accuracy decreases. The variable that is missing from Table 10 is "number of data points". The influence of this variable was either nonsignificant statistically, or when it was significant, its sign was positive, which made no sense. It was therefore excluded.

Many other factors (autocorrelation coefficients, standard error, coefficient of variation, mean absolute percentage change in the original series, seasonal indices, intercept and slope of regression equation (trend fitting), the absolute change in trend-cycle at periods $n-13$, $n-14$, $n-15$, and so forth) were tested, but (i) they were statistically nonsignificant, (ii) they resulted in R^2 that were inferior to those shown in Tables 9 and 10, or (iii) they were multicollinear with other variables. None of this additional information available up to period $n-12$ could be used to understand or explain differences in variations in accuracy, so it is not presented here.

Naive 1 and, to a lesser extent, the exponential smoothing methods seem to do well because they "hedge" their forecasts towards the middle. In this respect the chances of large errors are smaller when the pattern changes. The more sophisticated ARMA methods, on the other hand, attempt to follow the pattern as closely as possible. When there is no change from the previous pattern they are very accurate, which can be clearly seen by examining well-behaved series. When they miss, however, the errors are large. These large errors are shown in several ways. The coefficients of variation of the m.a.p.e. of, for example, Naive 1, single exponential smoothing, Holt's exponential smoothing, Winters' exponential smoothing, adaptive filtering and ARMA are 1·15, 1·29, 1·2, 1·27, 1·65 and 1·55 respectively. These coefficients indicate that the m.a.p.e. of the more sophisticated methods of adaptive filtering and ARMA are more dispersed around the mean values than the smoothing methods. Furthermore, Naive 1 has the smallest coefficient. With similar reasoning it can be seen that Naive 1 does the best in almost all cases when the unconstrained quadratic loss function is used (see Table 7). Smoothing methods on the original data are also relatively more accurate than on the same data after they have been seasonally adjusted, even when series with high randomness are excluded. This higher accuracy means that seasonal adjustments, which attempt to predict the seasonal pattern, can be inaccurate, which may cause large errors. The relative accuracy comparison changes in Table 8, which adjusts the value of Theil's U-statistic (it sets it equal to 2 when it exceeds 2). The extent of large random values then becomes less important, and the seasonal methods (excluding Harrison's harmonic smoothing) do better than Naive 2 and about the same as the methods using seasonally adjusted data. When a linear loss function is used (see Tables 1, 2 and 3), the difference is more pronounced. For example, in Table 3 the ARMA method is at least as good as any of the remaining methods in terms of the average of the m.a.p.e.s and considerably better than Naive 1.

It may be that when high randomness is present in a data series, more sophisticated methods such as ARMA may overfit a model to these data. This overfit could occur while the mean square error is being minimized when stationarity is being achieved through and/or seasonal differencing, or when an ARMA (p,q) model is selected. Lack of randomness in residuals, for instance, does not always mean better forecasting results. In our study, 28 of the first identified appropriate ARMA (p,q) models for the 111 series did not produce random residuals, which means that the model fitted was not adequate. Alternative models were later selected, and in the final analysis the residuals from all series were random. Surprisingly, however, the new average m.a.p.e.s for the forecasts were no better than when 28 series did not have random residuals. Smoothing methods, on the other hand, do not require random residuals; instead the optimal model is determined by minimizing the m.s.e. Decomposition methods are not even concerned with minimizing mean square errors, or any other kind of loss function, except possibly implicitly through the averaging process they employ.

Finally, the difference between *model fitting* and *forecasting* as well as the type of loss function should be mentioned. In Table 7 which is based on an unconstrained standardized quadratic loss function, the seasonal methods do as well as those using seasonally adjusted data and much better than the nonseasonal methods. In forecasting, however, that is not true. The best methods vary depending upon the loss function assumed and the amount of randomness present in the series. It is also interesting to note that the m.a.p.e. of the forecasts is often smaller than the m.a.p.e. of model fitting (see Tables 1, 2 and 3). That is not so with Tables 7 and 8 which utilize a quadratic accuracy measure.

5. CONCLUSIONS

A decision maker using these 111 series who wanted to apply a single forecasting method would have obtained very different results depending upon what loss function he wanted to minimize and whether he wanted to minimize the errors in the model fitting or in the forecasting phase. Overall, however, he would have done as well by using simpler rather than more sophisticated methods.

Further research will be required to shed light on the "mystery" of why, under certain circumstances, simpler methods do as well or better than sophisticated ones. Obviously, this result is contrary to expectations and the previous experience of the authors. We believe it probably happened because the series used include observations from the 1974–75 major recession. At the same time, this occurrence is not an isolated instance; it has been reported in several other studies (Chatfield and Prothero, 1973; Groff, 1973; Dawes and Corrigan, 1974; Geurts and Ibrahim, 1975; McWhorter, 1975; Narasimham *et al.*, 1975; McCoubrey and McKenzie, 1976).

In addition to more research, several other improvements should be considered:

(i) Combining quantitative forecasts (Bates and Granger, 1969; Newbold and Granger, 1974) and subjective and quantitative predictions in a Bayesian framework (Pankoff and Robert, 1968) may improve forecasting efficiency.

(ii) Developing alternative forecasting approaches, based on different methodologies such as anticipatory surveys, Delphi and so forth (Friend and Jones, 1964; Ripe *et al.*, 1976), can improve results.

(iii) Being able to deal with situations caused by unusual events (for example, the oil embargo, legislation; see Box and Tiao, 1975) or to anticipate changes in the structural pattern of the data (Box and Tiao, 1976) will undoubtedly result in lower errors.

(iv) Switching methods (for example, using an exponential smoothing model or Naive 1 if a recession is imminent) may also reduce forecasting errors or provide more conservative forecasts.

In the final analysis, however, the findings of this and other studies cannot be generalized until more is known about the reasons and causes affecting the accuracy of forecasting. We hope that determining and measuring the factors affecting accuracy, as described in Section 3, can be of great help in finding the most accurate forecasting methods for specific situations. What is needed is an understanding of *when* and *under what circumstances* one method is to be preferred over the others.

ACKNOWLEDGEMENTS

The authors would like to thank Professors Ahmet Aykac, Robert Fildes, Robin Hogarth, Jean Claude Larréché, Harry Roberts, Bernard Tate and many others, too numerous to mention, for their comments and useful suggestions.

APPENDIX A—THE METHODS

This appendix presents the methods used to make the analyses described in the body of the paper.

Two sets of errors were calculated for each method. The first was arrived at by fitting a model to the first $n-12$ values of each of the 111 series and calculating the error e_t as follows:

$$e_t = X_t - \hat{X}_t, \tag{A-1}$$

where X_t is the actual value, and $\hat{X}_t$ is the one-period-ahead forecasted value.

Three so-called errors of "model fitting" were also calculated as in (a)–(c) below, where all summations go from 1 to $n-12$.

(a) The mean average percentage error (m.a.p.e.) $= (n-12)^{-1}\Sigma(|e_t|/X_t)(100)$. (A-2)

(b) The mean square error (m.s.e.) $= (n-12)^{-1}\Sigma e_t^2$. (A-3)

(c) Theil's U-statistic: $= \sqrt{\{\Sigma e_t^2/\Sigma(X_t - X_{t+1})^2\}}$. (A-4)

The percentage of time the error in method i was smaller than in method m was also recorded.

The second set of errors involves the last 12 values, which were utilized as post-sample measures to determine the magnitude of the errors. The three measurements shown in equations (A-2), (A-3), (A-4), as well as the percentage of time method i was better than method m, were also computed for up to k forecasting horizons, starting at period $n-11$. For convenience, k takes the values of 1, 2, 3, 4, 5, 6, 9 and 12 only. In no instance have the last $n-11$ values been used to develop a forecasting model or estimate its parameters.

The following methods were used in the comparison. A more complete description of them given by Makridakis and Wheelwright (1977) also provides additional references.

(1) *Naive 1*

$$\text{Model fitting:}\quad \hat{X}_{t+1} = X_t, \tag{A-5}$$

where $t = 1, 2, 3, \ldots, n-12$.

$$\text{Forecasts:}\quad \hat{X}_{n-12+k} = X_{n-12}, \tag{A-6}$$

where $k = 1, 2, 3, 4, 5, 6, 9, 12$.

(2) *Simple moving average*

$$\text{Model fitting:}\quad \hat{X}_{t+1} = \frac{X_t + X_{t-1} + X_{t-2} + \ldots + X_{t-N+1}}{N}, \tag{A-7}$$

where N is chosen so as to minimize Σe_t^2, again summing over t from 1 to $n-12$.

$$\text{Forecasts:}\quad X_{n-12+k} = \frac{X_{n-12+k-1} + X_{n-12+k-2} + \ldots + X_{n-12+k-N}}{N}. \tag{A-8}$$

When the subscript of X on the right-hand side of (A-8) is larger than $n-12$, the corresponding forecasted value is substituted.

(3) *Single exponential smoothing*

$$\text{Model fitting:}\quad \hat{X}_{t+1} = \alpha X_t + (1-\alpha)\,\hat{X}_t, \tag{A-9}$$

where α is chosen so as to minimise Σe_t, the mean square error where again summing is over t from 1 to $n-12$.

$$\text{Forecasts:}\quad \hat{X}_{n-12+k} = \alpha X_{n-12} + (1-\alpha)\,\hat{X}_{n-12+K-1}. \tag{A-10}$$

(4) *Adaptive response rate exponential smoothing*

The equations are exactly the same as in (A-9) and (A-10), except α varies with t. The value of α_t is found by

$$\alpha_t = |E_t/M_t|, \tag{A-11}$$

where $E_t = \beta e_t + (1-\beta)\,E_{t-1}$ and $M_t = \beta\,|e_t| + (1-\beta)\,M_{t-1}$. β is set at ·2.

(5) *Linear moving average*

$$\text{Model fitting:}\quad S_t' = \frac{X_{t-1} + X_{t-2} + \ldots + X_{t-N}}{N},$$

$$S_t'' = \frac{S_{t-1}' + S_{t-2}' + \ldots + S_{t-N}'}{N}, \quad \hat{X}_{t+1} = a_t + b_t, \tag{A-12}$$

where $a_t = 2S_t' - S_t''$ and $b_t = 2(N-1)^{-1}(S_t' - S_t'')$.

The value of N is chosen so as to minimize the mean square error.

$$\text{Forecasts:}\quad \hat{X}_{n-12+k} = a_{n-12} + b_{n-12}(k). \tag{A-13}$$

(6) *Brown's one-parameter linear exponential smoothing*

$$\text{Model fitting:}\quad S_t' = \alpha X_t + (1-\alpha)\,S_{t-1}', \quad S_t'' = \alpha S_t' + (1-\alpha)\,S_{t-1}'', \quad \hat{X}_{t+1} = a_t + b_t, \tag{A-14}$$

where $a_t = 2S_t' - S_t''$ and $b_t = (1-\alpha)^{-1}(S_t' - S_t'')$.

The value of α is chosen so as to minimize the mean square error.

$$\text{Forecasts:}\quad \hat{X}_{n-12+k} = a_{n-12} + b_{n-12}(k). \tag{A-15}$$

(7) *Holt's two-parameter linear exponential smoothing*

$$\text{Model fitting:} \quad S_t = \alpha X_t + (1-\alpha)(S_{t-1}+T_{t-1}), \tag{A-16}$$

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}, \tag{A-17}$$

$$\hat{X}_{t+1} = S_t + T_t. \tag{A-18}$$

The values of α and β are chosen so as to minimise the mean square error. Marquardt's (1963) nonlinear optimization algorithm is employed for this purpose.

$$\text{Forecasts:} \quad \hat{X}_{n-12+k} = S_{n-12} + T_{n-12}(k). \tag{A-19}$$

(8) *Brown's one-parameter quadratic exponential smoothing*

$$\text{Model fitting:} \quad S_t' = \alpha X_t + (1-\alpha)S_{t-1}', \tag{A-20}$$

$$S_t'' = \alpha S_t' + (1-\alpha)S_{t-1}'', \tag{A-21}$$

$$S_t''' = \alpha S_t'' + (1-\alpha)S_{t-1}''', \tag{A-22}$$

$$\hat{X}_{t+1} = a_t + b_t + 1/2c_t, \tag{A-23}$$

where

$$a_t = 3S_t' - 3S_t'' + S_t''', \quad b_t = \alpha\{2(1-\alpha)^2\}^{-1}\{(6-5\alpha)S_t' - (10-8\alpha)S_t'' + (4-3\alpha)S_t'''\}$$

and

$$c_t = \alpha(1-\alpha)^{-2}(S_t' - 2S_t'' + S_t'''.)$$

The value of α is chosen so as to minimize the mean square error.

$$\text{Forecasts:} \quad \hat{X}_{n-12+k} = a_{n-12+k} + b_{n-12+k}(k) + 1/2c_{n-12+k}(k)^2. \tag{A-24}$$

(9) *Linear regression trend fitting*

$$\text{Model fitting:} \quad \hat{X}_t = a + bt, \tag{A-25}$$

where $t = 1, 2, 3, \ldots, n-12$, and a and b are chosen so as to minimize the sum of the square errors by solving the normal equations:

$$a = \frac{\Sigma X}{n-12} - b\frac{\Sigma t}{n-12}, \quad b = \frac{(n-12)\Sigma tX - t\Sigma X}{(n-12)\Sigma t^2},$$

where all summations go from 1 to $n-12$.

$$\text{Forecasts:} \quad \hat{X}_{n-12+k} = a + b(n-12+k). \tag{A-26}$$

(10) *Harrison's harmonic smoothing*

(a) *Removal of trend-cycle*. The trend-cycle is removed by calculating an L-period moving average (where L is the length of seasonality) and dividing it into the time series X_t. Using the moving average elimiates seasonality and randomness. Thus,

$$M_t = T_t C_t, \tag{A-27}$$

where T_t denotes trend and C_t denotes cycle.

(b) *Calculation of crude seasonal estimates*. Equation (A-27) can be divided into X, yielding

$$R_t^* = \frac{X_t}{M_t} = \frac{S_t T_t C_t R_t}{T_t C_t} = S_t R_t, \tag{A-28}$$

where S_t denotes seasonality and R_t denotes randomness. The R_t^*'s are estimates of seasonality with some randomness remaining.

In order to eliminate randomness, R_t^* is arranged in such a way that similar months can be averaged together. Letting r_{ij} denote the ith year and the jth period (month), one could average all the periods together as follows:

$$r_j = \Sigma r_{ij} m_j^{-1}, \tag{A-29}$$

summing over i from 1 to m_j, where m_j is the number of observations available for the jth period and the r_j are the crude seasonal estimates.

The standard deviation of the crude seasonal estimate is

$$s = \sqrt{\left\{\sum_{i=1}^{m_j} \sum_{j=1}^{L} (r_{ij} - r_j)^2\right\} \Big/ \left(\sum_{j=1}^{L} m_j - L\right)}. \tag{A-30}$$

(c) *Obtaining smoothed seasonal estimates through Fourier analysis.* The smoothed seasonal estimates, r_j, can be obtained by applying

$$\hat{r}_j = 1 + \sum_{k=1}^{L} a_k \cos(kd_j) + b_k \sin(kd_j), \quad j = 1, 2, \ldots, L, \tag{A-31}$$

where

$$d_j = \frac{2(j-1)\pi}{L/2} - \pi, \quad j = 1, 2, \ldots, L, \tag{A-32}$$

$$a_k = \frac{1}{L/2} \sum_{j=1}^{L} r_j \cos(KX_j), \tag{A-33}$$

$$b_k = \frac{1}{L/2} \sum_{j=1}^{L} r_j \sin(KX_j). \tag{A-34}$$

In equations (A-33) and (A-34), the crude seasonal estimates, r_j, and the actual values, X_j, are used to estimate the Fourier coefficients, which in turn are used in equation (A-31) to calculate the smoothed seasonal estimates. Only those $\hat{r}_j$ in equation (A-31) that are significantly different from zero are included in the final calculations; the remainder are discarded.

(d) *Replacement of the outliers.* If the difference between r_{ij}, the original estimates of seasonality, and $\hat{r}_j$, the smoothed estimates, is greater than 2·50 (see (A-30)), the corresponding X_t value is replaced by some more likely value (for example, $\hat{r}_j R_t^*$). When outliers are replaced, the effect of unusual events such as strikes, wars and total breakdowns is eliminated from the series, which allows a more realistic estimate.

After the outliers are replaced, steps (a), (b) and (c) above are recomputed with the replaced values. Thus, new smoothed seasonal estimates are calculated with equation (A-31).

(e) *Measurement of the adequacy of the seasonal fit.* The adequacy of the seasonal fit is tested by an F-test. The F-value is computed as the ratio of the mean square errors between the variations caused by seasonal effects, M_s, and that caused by unexplained residual factors, M_r.

$$F\text{-test} = M_s/M_r.$$

$$\text{Forecasts:} \quad \hat{X}_{n-12+k} = T_{n-12+k} \hat{r}_j, \tag{A-35}$$

where T_{n-12+k} is an estimate of the trend-cycle at period $n-12$ and $\hat{r}_j$ is the corresponding seasonal index for period $n-12+k$.

T_{n-12+k} is calculated by different methods depending on whether or not there is a statistically significant trend in the data.

(11) *Winters' three-parameter linear and seasonal exponential smoothing*

$$\text{Model fitting:}\quad S_t = \alpha\frac{X_t}{I_{t-L}} + (1-\alpha)(S_{t-1}+T_{t-1}), \tag{A-36}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1},\quad I_t = \beta\frac{X_t}{S_t} + (1-\beta)I_{t-L},\quad \hat{X}_{t+1} = (S_t+T_t)I_{t-L+1},$$

where L is the length of seasonality.

The values of α, β and γ are chosen so as to minimize the mean square error by Marquardt's (1963) nonlinear optimization steepest descent algorithm.

$$\text{Forecasts:}\quad \hat{X}_{n-12+k} = (S_{n-12} + kT_{n-12})I_{n-12+k}. \tag{A-37}$$

(12) *Adaptive filtering*

$$\text{Model fitting:}\quad X_t = \phi_{1t}X_{t-1} + \phi_{2t}X_{t-2} + \phi_{3t}X_{t-3} + \ldots + \phi_{pt}X_{t-p}. \tag{A-38}$$

The values of the parameters ϕ_{it} are modified by equation (A-39) so as to minimize the mean square error.

$$\phi_{ti} = \phi_{(t-1)i} + 2Ke_t^* X_{t-i}^{**}, \tag{A-39}$$

where K is a learning constant set equal to $1/p$ and e_t^* and X_{t-1}^* are standardised values of e_t and X_{t-1} correspondingly.

The order of the autoregressive equation (A-38) is set automatically. Thus, $p = 3$ for nonseasonal data and $p = L$ for seasonal stationary data. If the data are nonstationary and seasonal, the first difference is found, then p is set equal to L.

Forecasts:

$$\hat{X}_{n-12+k} = \phi_{1(n-12)}X_{n-11+k} + \phi_{2(n-12)}X_{n-10+k} + \ldots + \phi_{p(n-12)}X_{n-p+k}. \tag{A-40}$$

If the first difference has been taken, equation (A-39) is adjusted accordingly. When the subscript of X on the right-hand side of equation (A-40) exceeds $n-12$, the forecasted value $\hat{X}$ is substituted in equation (A-40).

(13) *Autoregressive moving average (Box–Jenkins) methodology*

Model fitting:

$$\hat{X}_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots \phi_p X_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}. \tag{A-41}$$

The parameters of equation (A-41) are estimated, using Marquardt's (1963) nonlinear optimization procedure so as to minimize the mean square error. Initial estimates of the parameters are provided to start the algorithm by solving the Yule–Walker equations for the the ϕ_i. An iterative procedure is used for initial estimates of ϕ_i. If there is no convergence in the iterative process (which is unusual), the values of ϕ_i are set at ·2, ·3, ·4, ...

The orders of p and q are determined by examining the autocorrelations and partial autocorrelations of the first $n-12$ data points of each series. This determination was made judgementally. If the data were seasonal, a multiplicative seasonal model was specified by incorporating seasonal parameters in equation (A-41).

Before applying equation (A-41), two further steps were taken. (i) Two transformations were applied to the data to assure stationarity in variance. Square root and logarithmic transformations were compared to no transformation, and the one that produced the most homogeneous variance was chosen. The method of range mean plot described by Box and Jenkins (1973) was employed to select the best transformation. It was found, however, that the difference in accuracy between the transformed and untransformed data was marginal, if any, and was restricted to the longer forecasting horizons of 9 and 12 periods ahead. (ii) Next, regular (short) and seasonal (long) differences were applied to the data, if necessary, so as to achieve stationarity of the mean.

Once the best transformation (or no transformation) had been made and the appropriate level of differences chosen, expression (A-41) or its seasonal extension was used to estimate the optimal ARMA parameters. Then the autocorrelation of residuals was examined. If the autocorrelations showed random residuals, the tentatively identified model was considered adequate; otherwise, another model was specified. It may be of interest that 75 per cent of the tentatively identified models produced random residuals the first time. About half of the remaining 25 per cent were correctly identified the second time, and only one-fifth of the remainder the third time. It took 6 or 7 trials to achieve random residuals for all of the series. The biggest difficulty in correct identification arose through having specified the wrong level of differencing and not getting enough clues from the autocorrelations to perceive the error. The residuals were considered random when all of their autocorrelations, up to 24 time lags, were within the limits of $\pm 1{\cdot}96\{\sqrt{(n-12)}\}^{-1}$ and the Box–Pierce statistic was smaller than the corresponding value of the χ^2 distribution for the lags 1–12 and 1–24. There were rarely autocorrelations outside the 95 per cent confidence limits, but in no case was the Box–Pierce greater than the corresponding χ^2 value (for both time lags 1–12 and 1–24).

Only the ARMA method required a considerable amount of judgemental input; other methods were run automatically. Finally, no backforecasting was used to estimate initial values for the error terms e_{t-i}.

Forecasting: Forecasts were made after the estimates were retransformed and the differences taken into consideration. Forecasted values were used when subscripts of X on the right-hand side of equation (A-42) were larger than $n-12$.

$$\hat{X}_{n-12+k} = \phi_1 X_{n-11+k} + \phi_2 X_{n-10+k} + \ldots + \phi_p X_{n-p+k} - \theta_1 e_{n-11+k} - \theta_2 e_{n-10+k} - \theta_q e_{n-q+k}. \tag{A-42}$$

(14) *Seasonally adjusted data for 22 methods*

Model fitting: The *$n-12$ data* of each of the 111 series were first adjusted for seasonality, as

$$X_t' = X_t/S_t', \tag{A-43}$$

where X_t' is the seasonally adjusted value at period t, and S_j' is the corresponding seasonal factor for period t.

Forecasts: The forecasts for $\hat{X}'_{n-11}, \hat{X}'_{n-10}, \ldots, \hat{X}'_n$ were reseasonalized as

$$\hat{X}_{n-12+k} = \hat{X}_{n-12+k}(S_j'). \tag{A-44}$$

Equations (A-43) and (A-44) were applied for all non seasonal methods 1 to 9 above.

The indices S_j' were computed (see Appendix B) by applying the simple ratio-to-moving average decomposition procedure. It was found that the seasonal indices and the one-year-ahead forecasts of the seasonal indices computed by the Census II decomposition method of the USA Bureau of Labor Statistics did not produce better forecasting results than those of the simple ratio-to-moving average method.

APPENDIX B—DECOMPOSING A TIME SERIES

If a multiplicative relationship is assumed of the form:

$$X_t = S_t T_t C_t R_t, \tag{A-45}$$

where S_t is the seasonality at period t, T_t is the trend at period t, C_t is the cycle at period t and

R_t is the randomness at period t, then by computing a centered moving average of equation (A-45), whose length is equal to the length of seasonality, we get

$$M_t = T_t C_t. \tag{A-46}$$

Dividing equation (A-46) into equation (A-45) yields

$$\frac{X_t}{M_t} = \frac{S_t T_t C_t R_t}{T_t C_t} = S_t R_t. \tag{A-47}$$

Averaging equation (A-47) will eliminate randomness and yield seasonal estimates, S'_j. Dividing S'_j into equation (A-45) yields

$$X_t^* = T_t C_t R_t. \tag{A-48}$$

Computing a low order moving average (the one used was a double 3×3 moving average) on the X_t^* values of equation (A-48) results in the elimination of randomness and yields another set of trend-cycle values M'_t or

$$M'_t = T_t C_t. \tag{A-49}$$

Finally, dividing equation (A-49) into equation (A-48) will isolate randomness, R_t. Thus, randomness is defined as whatever fluctuation remains after seasonality and trend-cycle fluctuations have been removed. It is a residual term in this sense and is equal to

$$\frac{X_t^*}{M'_t} = \frac{T_t C_t R_t}{T_t C_t} = R_t. \tag{A-50}$$

The concept of randomness as used throughout this paper is defined by equation (A-50) above.

The seasonal indices, S'_j, were used in equation (A-43) and (A-44) for adjusting and re-adjusting for seasonality. Furthermore, mean absolute percentage changes for each of the components of the time series were computed as

$$\text{m.a.p.c. } Y = 100\left(\frac{1}{n-12}\sum_{t=2}^{n-12}\frac{|Y_t - Y_{t-1}|}{Y_{t-1}}\right), \tag{A-51}$$

where Y_t can be any of the components of the time series. The m.a.p.c. of the components are used as the independent variables of the regression equations shown in Tables 9 and 10.

REFERENCES

ADAM, E. E. and EBERT, R. J. (1976). A comparison of human and statistical forecasting. *AIITE Trans.*, **8**, 120–127.

BATES, J. M. and GRANGER, C. W. J. (1969). Combination of forecasts. *Operat. Res. Q.*, **20**, 451–468.

BAUMAN, W. S. (1965). The less popular stocks versus the most popular stocks. *Financial Analysts J.*, **21**, 61–69.

BOX, G. E. P. and JENKINS, G. M. (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall. *J. R. Statist. Soc.*, **139**, 295–336.

BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Ass.*, **70**, 70–79.

—— (1976). Comparison of forecast and actuality. *Appl. Statist.*, **25**, 195–200.

BURMAN, J. P. (1965). Moving seasonal adjustment of economic time series. *J. R. Statist. Soc.* A, **128**, 534–558.

CHATFIELD, C. and PROTHERO, D. L. (1973). Box–Jenkins seasonal forecasting: problems in a case study. *J. R. Statist. Soc.* A, **136**, 295–336.

CHRIST, C. F. (1951). A test of an econometric model of the United States, 1921–1974. In *Conference on Business Cycles*. New York: National Bureau of Economic Research.

—— (1975). Judging the performance of econometric models of the U.S. economy. *Int. Econ. Rev.*, **16**, (1), 57–81.

CLEARLY, J. P. and FRYK, D. A. (1974). A comparison of ARIMA and econometric models for telephone demand. *Proc. Amer. Statist. Ass.*, Business and Economics Section, pp. 448–450.

CLEVELAND, W. P. and TIAO, G. C. (1976). Decomposition of seasonal time series: a model for the census X-11 program. *J. Amer. Statist. Ass.*, **71**, 581–587.

COOPER, J. P. and NELSON, C. R. (1975). The ex ante performance of the St. Louis and FRB–MIT–PENN econometric models and some results on composite predictors. *J. of Money, Credit and Banking*, February issue, pp. 1–32.

COOPER, R. L. (1972). The predictive performance of quarterly econometric models of the United States. In *Econometric Models of Cyclical Behavior* (B. G. Hickman, ed.) New York: National Bureau of Economic Research.

CRAGG, J. and MALKIEL, B. (1968). The consensus and accuracy of some predictions of the growth in corporate earnings. *J. of Finance*, March issue, pp. 67–84.

DAWES, R. M. (1977). In *Shallow Psychology* (J. S. Carroll and J. W. Payne, eds).Hilldale, N. J.: Laurence Erlbaunn.

DAWES, R. M. and CORRIGAN, B. (1974). Linear models in decision making. *Psychol. Bull.*, **81**, 95–106.

ELTON, E. J. and GRUBER, M. J. (1972). Earnings estimates and the accuracy of expectational data. *Management Sci.* B, **18**, 409–424.

FAMA, E. F. (1965). The behaviour of the stock market prices. *J. Business*, **38**, 34–105.

FRIEND, J. and JONES, R. C. (1964). Short-run forecasting models incorporating anticipatory data. In *Models of Income Determination* (NBER). Princeton: Princeton University Press.

FROMM, G. and KLEIN, L. R. (1973). A comparison of eleven econometric models of the United States. *Amer. Econ. Ass.*, May issue, 385–401.

GEURTS, M. D. and IBRAHIM, I. B. (1975). Comparing the Box–Jenkins approach with the exponentially smoothed forecasting model application to Hawaii tourists. *J. Marketing Res.*, **12**, 182–188.

GOLDBERG, L. R. (1970). Man versus model of man: a rationale, plus some evidence, for a method of improving clinical inferences. *Psychol. Bull.*, **73**, 422–432.

—— (1976). Man versus model of man: just how conflicting is that evidence? *Organizational Behavior and Human Performance*, Vol. 16, 13–22.

GOLDFELD, S. M. (1972). The predictive performance of quarterly econometric models of the United States: Comments. In *Econometric Models of Cyclical Behavior* (B. G. Hickman, ed.). (Studies in Income and Wealth, no. 36). New York: Columbia University Press for NBER.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica*, **37**, 424–438.

—— (1974). *Spectral Analysis of Economic Time Series*. Princeton: Princeton University Press.

GRANGER, C. W. J. and NEWBOLD, P. J. (1974). Spurious regressions in econometrics. *J. Econometrics*, **2**, 111–120.

GREEN, D. and SEGALL, J. (1967). The predictive power of first-quarter earnings reports. *J. Business*, **40**, January issue, 44–55.

GROFF, G. K. (1973). Empirical comparison of models for short-range forecasting. *Management Sci.*, **20** (1), 22–31.

GROSS, D. and RAY, J. L. (1965). A general purpose forecasting simulator. *Management Sci.* B, **11** (6), 119–135.

HOGARTH, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *J. Amer. Statist. Ass.*, **70**, 271–290.

VON HOHENBALKEN, B. and TINTNER, G. (1962). Econometric models of the O.E.E.C. member countries, the United States and Canada, and their application to economic policy. *Weltwirtschaftliches Archiv*, **89**, 29–86.

HOWREY, E. P., KLEIN, L. R. and MCCARTHY, M. D. (1974). Notes on testing the predictive performance of econometric models. *In Econ. Rev.*, **15**, 2, 376–377.

JENSEN, M. C. (1968). The performance of mutual funds in the period 1945–1956. *J. of Finance*, **23**, 389–416.

JOHNSON, T. E. and SCHMITT, T. G. (1974). Effectiveness of earnings per share forecasts. *Financial Management*, summer issue, 64–72.

JORGENSON, D. W., HUNTER, J. and NADIRI, M. I. (1970). A comparison of alternative econometric models of quarterly investment behavior. *Econometrica*, **38**, 187–213.

KAHNEMAN, D. and TVERSKY, A. (1973). On the psychology of prediction. *Psychol. Rev.*, **80**, no. 4, 237–251.

KIRBY, R. M. (1966). A comparison of short and medium-range statistical forecasting methods. *Management Sci* B, 15 no. 4, 202–210.

KRAMPF, R. F. (1972). The turning point problem in smoothing models. Ph.D. Thesis, University of Cincinnati.

LESER, C. E. V. (1966). The role of macro-models in short-term forecasting. *Econometrica*, **34**, 862–872.

—— (1968). A survey of econometrics. *J. R. Statist. Soc.* A, **131**, 530–566.

LEVINE, A. H. (1967). Forecasting techniques. *Management Accounting*, January issue, 86–95.

LIBBY, R. (1976). Man versus model of man: some conflicting evidence. *Organizational Behaviour and Human Performance*, **16**, 1–12.

LIEBLING, H. I. and RUSSELL, J. M. (1969). Forecasting business investment by anticipation surveys and econometric models in 1968–69. *1969 Proceedings of Business and Economic Statistics Section*, American Statistical Association, pp. 250–255.

LIEBLING, H. I, BIDWELL, P.T. and HALL, K. E. (1976). The recent performance of anticipation surveys and econometric model projections of investment spending in the United States. *J. of Business*, **49**, no. 4., 451–478.

MABERT, V. A. (1975). Statistical versus sales force—executive opinion short-range forecasts: a time-series analysis case study. Krannert Graduate School, Purdue University (working paper).

MCCARTHY, M. D. (1972). The predictive performance of quarterly econometric models of the United States: comments. In *Econometric Models of Cyclical Behavior* (B.G. Hickman, ed.). (Studies in Income and Wealth, no. 36.) New York: Columbia University Press for NBER.

MCCOUBREY, C. A. and MCKENZIE, W. Jr (1976). Forecasting collections using expert opinions in a deterministic model. ORSA/TIMS Conference, Miami, Fl.

MCNEES, S. K. (1975). An evaluation of economic forecasts. *New England Econ.Rev.*, November–December issue.

—— (1976). The forecasting performance in the early 1970s. *New England Econ. Rev.*, July–August 29–40.

MCWHORTER, A. Jr (1975). Time-series forecasting using the Kalman filter: an empirical study. *Proc. Amer. Statist. Ass.*, Business and Economics Section, pp. 436–446.

MAKRIDAKIS, S. (1976). A survey of time series. *Int. Statist. Rev.*, **44**, 29–70.

MAKRIDAKIS, S. and MAJANI, B. (1977). Can recession be predicted? *Long-range Planning J.*, April issue.

MAKRIDAKIS, S. and VANDEBURGH, H. M. (1975). The accuracy and cost of nonseasonal time-series forecasting methods. *INSEAD Research Papers Series*, no. 143.

MAKRIDAKIS, S. and WHEELWRIGHT, S. (1978). *Interactive Forecasting: Univeriate and Multivariate Methods.* San Francisco: Holden-Day.

MARQUARDT, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. *Soc. Indust. Appl. Math.*, **11**, 431–441.

MEEHL, P. E. (1954). *Clinical versus Statistical Prediction: A Theoretical Analysis and Review of the Literature.* Minneapolis: University of Minnesota Press.

—— (1965). Clinical versus statistical prediction. *J. Exper. Res. in Personality*, **1**, 27–32.

NARASIMHAM, G. V. L. *et al.* (1974). On the predictive performance of the BEA quarterly econometric model and a Box–Jenkins type of ARIMA model. *Proc. Amer. Statist. Assoc.* Business and Economics Section, pp. 448–450.

—— (1975). A comparison of predictive performance of alternative forecasting techniques: time series versus an econometric model. *Proc. Amer. Statist. Ass.*, Business and Economics Section, pp. 459–464.

NAYLOR, T. H. *et al.* (1972). Box–Jenkins methods: an alternative to econometric forecasting. *Int. Statist. Rev.*, **40**, 123–137.

NELSON, C. R. (1972). The prediction performance of the FRB–MIT–PENN model of the U.S. economy. *Amer. Econ. Rev.*, **62**, 902–917.

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.

NIEDERHOFFER, V. and REGAN, D. (1972). Summarized in *Barron's Magazine*, December issue.

O'BRIEN, J. W. (1970). How market theory can help investors set goals, select investment managers and appraise investment performance. *Financial Analysts J.*, **26**, 91–103.

PANKOFF, L. D. and ROBERTS, H. V. (1968). Bayesian synthesis of clinical and statistical prediction. *Psychol. Bull.*, **70**, 762–773.

PIERCE, D. A. (1977). Relationships—and the lack thereof—between economic time series, with special reference to money, reserves, and interest rates. *J. Amer. Statist. Ass.*, **72**, 11–27.

RAINE, J. E. (1971). Self-adaptive forecasting considered. *Decision Sci.*, 2, no. 2., 264–279.

REID, D. J. (1969). A comparative study of time-series prediction techniques on economic data. Ph.D. Thesis, University of Nottingham.

—— (1977). A survey of statistical forecasting techniques with empirical comparisons. Paper presented at the *I.E.E. Symposium on Statistical Model Building for Precition and Control.*

RIPE, R., WILKINSON, M. and MORRISON, D. (1976). Industrial market forecasting with anticipatory data. *Management Sci.*, **22**, 639–651.

SARBIN, T. R. (1943). Contribution to the study of actuarial and individual methods of prediction. *Amer. J. Sociol.*, **48**, 593–602.

SAWYER, J. (1966). Measurement and prediction, clinical and statistical. *Psychol. Bull.*, **66**, 178–200.

SLOVIC, P. (1972). Psychological study of human judgement: implications for investment decision making. *J. of Finance*, **27**, 779–799.

—— (1974). Limitations of the mind of man: implications for decision-making in the nuclear age. *Oregon Research Institute Research Bulletin*, **11**, no. 17.

STECKLER, H. O. (1968). Forecasting with econometric models: an evaluation. *Econometrica*, **36**, 437–463.

THEIL, H. (1966). *Applied Economic Forecasting*, pp. 26–32. Amsterdam: North Holland.

TVERSKY, A. (1974). Assessing uncertainty (with Discussion). *J. R. Soc.* B, **36**, 148–159.

TVERSKY, A. and KAHNEMAN, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, **185**, 1124–1131.

DISCUSSION OF THE PAPER BY PROFESSOR MAKRIDAKIS AND DR HIBON

DR W. G. GILCHRIST (Sheffield City Polytechnic): It gives me pleasure to propose the vote of thanks to this paper. Modern man is fascinated with the subject of forecasting. The authors' first sections list a large number of references of people's attempts to compare methods. One reference which is not given is that from the Operational Research Society Forecasting Study Group. About 10 or 11 years ago it gathered together various sets of data and sent them out to different firms, asking them to forecast these data using their standard methodology. The reference is not in the paper because, having received the results, the Study Group was so horrified by them that they were brushed under the carpet, the carpet patted down, and no more was said about them. It is all very well, as statisticians, to say that there is a proper methodology for doing forecasting, that it is all laid out nicely in Box–Jenkins and various other books and papers, but we have to accept that most firms have vast numbers of time series to forecast, they have standard methods which are in their computer packages and they use these more or less regardless. Therefore, although we might say this is not the way to do it, the facts of life are such that this is the way most people do it: they take their data in large quantities and push it through their own standard methods. Therefore, there is real value in seeing how these very commonly used methods work when large varieties of different sorts of time series are put through them—because this is what happens in practice.

In the Operational Research Society exercise some things went disastrously wrong because of series which clearly became highly unstable just at the end—and the answer should have been that there should not be a forecast, but we should think about the situation. Another major source of error arose because seasonal data were used with non-seasonal methods. The authors carefully broke up the data according to randomness, looking at the subset of data with low randomness. I thought that they should have done the same with the seasonal and non-seasonal data. Looking at the Tables, the non-seasonal methods were given very short shrift because of the fact that most of the data that went through them, I suspect, were essentially seasonal data.

It also emerged from the Operational Research Society exercise that there was a need to do a fair amount of forecasting before it was possible properly to assess its quality. I was slightly concerned about the number of forecast errors used by the authors as the basis for their arguments. It was argued this evening that the problem is one of re-identifying the model—but, if this is what people do in practice, perhaps this is what the authors should have done. What concerns me is that analysis is based on the 12 errors $e_1, e_2, \ldots, e_{12}$. When forecasting one time unit into the future, if I understood correctly, it was just the single error e_1 that was used as a basis for comparing approaches and methods. Using only that one error for one step ahead slightly concerns me. The authors adopted a methodology of averaging the absolute percentage errors for sets of different lead times. This has effects which I do not really understand. Clearly, it creates quantities which are highly correlated to each other, which makes their interpretation difficult.

If we plot, for some of the better methods, mean absolute percentage error (m.a.p.e.) against lead-time (forecast horizon), one obtains remarkably linear plots. I am interested to know how linear they really are. The averaging operation will tend to make it much more difficult to observe deviations from linearity if they are fairly small.

I should, therefore, like to ask the authors to tell us a little more about this methodology of using mean absolute percentage errors because I feel that it does not help in understanding the effect of lead-time.

I rather query some of the terminology. When the authors talk about "fitting", in fact, they are using one-step ahead forecasting, but forecasts based on using all the data up to the last 12 to fit the model or to optimize the parameters. I should like to know what the model fitting results would have been if they had used the actual residuals between the fitted model and the observations at that point.

One other aspect of methodology I want to query is one which their predecessors who compared methods also adopted. It worries me that when comparisons are made the dice are very much loaded in favour of the autoregressive moving average (ARMA) models. A whole family of models is first defined, this is followed by choosing the one that fits best. In the authors' case, they also looked at various possible transformations. The chosen model is used for that particular set of data, and different models are used for different sets of data. When we turn to the polynomial models, a constant mean is used on all 111 models—linear trend on all 111, quadratic on all 111 and so on—and they are all then compared. It seems to me that we are not really comparing like with like. If one methodology is adopted for Box–Jenkins, perhaps we ought to consider for each time series which of the family of polynomials ought to be used, whether transformations should be used etc. The family of ARMA models would then be compared with the family of polynomial models.

Again, returning to the authors' general approach, it is fair to say that they are comparing what people use in practice. It is amazing to me, however, that after all this exercise in identifying models, transforming and so on, that the autoregressive moving averages come out so badly. I wonder whether it might be partly due to the authors not using the backwards forecasting approach to obtain the intial errors.

The approach and objective of the authors in trying to ask the questions: what is it that produces a good method, what is it that produces a bad one, and what characteristics of a set of data might guide us in choosing our methods, is worthwhile. Athough it is possible to criticize the way in which it has been done in the paper, as a first attempt it is an exercise that is well worth doing. If some method could be devised for characterizing sets of data which could then be plugged into automatic forecasting methods this might improve those methods.

May I thank the authors for an interesting paper, for raising some interesting issues and for underlining once again that, in reality, methods which are simple and fairly robust, fairly responsive to the non-stationary unstable world in which we live, are well worth considering, even though they lack the statistical finesse of the methods we like to teach our students. I should like to propose the vote of thanks.

Professor M. B. PRIESTLEY (University of Manchester Institute of Science and Technology): The subject of forecasting has aroused considerable interest over the last fifteen years or so and it has now acquired a massive literature, most of which is concerned with its theoretical aspects. It is therefore timely and refreshing to have presented to us the results of a substantial empirical study of various forecasting techniques, and in this respect the authors have rendered the Society a valuable service.

As in all empirical studies, however, we must resist the temptation to read too much into the results of the analyses. Let us first clarify some basic concepts. *There is no such thing as a forecasting "method"; there is no such thing as an* ARMA *(or Box–Jenkins) forcasting "method"*. There *is* something called a "least-squares" forecasting method, and this, in fact, provides the basis for virtually all theoretical studies. What the authors refer to as the "Box–Jenkins method" is simply a recursive algorithm for calculating *least-squares* forecasts, and, if the model for the series was known precisely, Box–Jenkins "method" would lead to exactly the same answer as the Wiener "method", the Kolmogorov "method" or even the Kalman filter "method". (All these "methods" are simply different computational procedures for calculating the same quantity, namely, the least-squares forecast of a future value from linear combinations of the past data.)

It is also extremely important to distinguish very clearly between a *model* and a *forecasting procedure*. *A model (such as an* ARMA *scheme) does not automatically lead to a particular forecasting formula.* A model represents a stochastic description of a series, and given a particular model it is up to the user to decide how to use this stochastic description so as to obtain that form of forecast which is optimal according to a chosen loss function. Of course, if we choose the mean-square error as our loss function, and if we are dealing with an ARMA model in which residuals are independent (rather than merely uncorrelated), then the optimal forecast is very easily constructed from the model by simply setting to zero the values of those residuals, ε_t, which correspond to future time points. This procedure produces the conditional expectation of the variable to be forecast (given the past data), and whilst the conditional expectation is optimal for loss functions other than the mean square error (for example, it is optimal for any loss function which is an even power of the forecast error, e_t, provided the conditional distribution is symmetric), it is certainly not optimal for an arbitrary choice of loss function and arbitrary distribution. To see this we need only remind ourselves of the well-known result that the expression $E[\,|X-c|\,]$ is minimized by choosing c as the median (not the mean) of X. This would certainly affect the form of the forecasts corresponding to the loss function, $E[\,|e_t|\,]$, if the series is non-Gaussian.

Thus, before one can start to consider the optimal forecast derived from a particular model one must first specify the chosen form of loss function. The authors apparently follow the reverse procedure, i.e. they first calculate the forecast, and consider afterward how best to assess its accuracy. In fact, the forecast formula used for ARMA models (A-42) is the one which is appropriate for the *mean square error* loss function, and the same is true for most of the other forecasting "methods" used, since their associated parameters (such as, for example, the parameter α in the "single exponential smoothing" technique) are all estimated by minimizing the mean square error.

The question now arises as to whether it is valid to assess the relative accuracies of the different methods according to the m.a.p.e. criterion which the authors select as their main basis for comparison. (They dismiss the mean square error criterion by saying that it cannot be used *across* series, but it can certainly be used to compare different forecasting techniques applied to the same series.) In principle, the optimal m.a.p.e. forecast could be determined from a given ARMA model, although it would be extremely difficult to obtain an analytical expression. However, it is certainly safe to say that it is by no means obvious that the optimal m.a.p.e. forecast would be identical with the optimal least-squares forecast! This prompts one to ask the question: are the authors' general conclusions heavily dependent on their use of the m.a.p.e. criterion?

Suppose that instead of using the m.a.p.e. they had used the mean-square error. In this case the optimal (linear) forecast is uniquely determined by the second-order properties of the series, or equivalently, by its model, and there is no scope for debating whether or not method A is superior to method B. *If the series conforms to an* ARMA *model, and the model has been fitted correctly, then the forecast based on this* ARMA *model must, by definition, be optimal.* (Apart from the ARMA model, all the other forecasting methods considered are of an *ad hoc* nature. The ARMA method involves model fitting and its performance depends to a large extent on the ability of the user to identify correctly the underlying model.)

I am afraid that I really cannot believe that the regression models described by the authors in Section 3 reflect some fundamental law of nature. Suppose we know (or generate) a seasonal + trend + residual series from a seasonal ARMA model. If we again use the the mean square error as a criterion then the forecast constructed from the mosel *must* be optimal—irrespective of the percentage variations due to the different components. The "single exponential smoothing" technique (more commonly referred to as "exponentially weighted moving averages") is optimal in the mean sense for a series whose first differences satisfy an MA(1) model. *If* the series is of this form, then single exponential smoothing will be superior to a forecast obtained from, for example, an (incorrectly fitted) ARMA(2, 3) model—but this tells us nothing about the relative merits of the two forecasting techniques. If the authors argue that these theoretical considerations do not apply to their study because in many cases the structures of their series were changing over time, then what they are really saying is that it is the *degree of non-stationary* (suitably measured) which mainly affects the forecasting performance. However, it may be noted that the optimal least-squares forecasts can be constructed for non-stationary series provided their structure is changing "smoothly" over time—see Abdrabbo and Priestley (1967).

The authors tell us that the regression equations were obtained by starting with a large number of "independent" variables, (X_1, X_2, X_3, ...) and then dropping those which were deemed to be "non-significant". It would be interesting to know how variables were originally included, how these were selected and whether this selection was determined before or after the forecasting results were available. However, the fact that the authors started out with a large number of variables suggests that the "significance" of those finally chosen could be somewhat dubious. (Ideally, the authors should have considered the $F_{\max}$ test rather than the standard F test.) I note also that in Section 4, third paragraph, the authors disclose the fact that the variable corresponding to the "number of data points" was dropped in the regression model for "forecasts" because, when significant, its coefficient had the "wrong sign". This also suggests that there may be some spurious features in the construction of these regression models.

I do not believe that it is very fruitful to attempt to classify series according to which forecasting techniques perform "best". The performance of any particular technique when applied to a particular series depends essentially on (a) the model which the series obeys; (b) our ability to identify and fit this model correctly and (c) the criterion chosen to measure forecasting accuracy.

Despite these reservations I am impressed by the magnitude of the authors' study and by the care which they have taken in scrutinizing the literature and comparing their qualitative results with those of previous authors. I was pleased to hear that the authors are willing to make available the full results of their analysis to others who may be interested in studying forecasting, in which case the substance of their study will provide a very valuable background for further work.

I have much pleasure in seconding the vote of thanks.

The vote of thanks was carried by acclamation.

Mr J. P. BURMAN (Bank of England): I should also like to thank the authors for a very interesting paper. The results are, indeed, provoking, but I find it difficult to draw conclusions because they are presented in summary form. It might almost be said that it is a "black box". For instance, we do not know how long the series are. In the case of the Box–Jenkins models, it is my experience, and I am sure that of many others, that a considerable amount of effort is required for model identification before a successful result is obtained. This is why I think it is difficult in a study on a large scale of over 100 series to apply these sorts of models successfully—they are not really suitable for automatic handling.

Dr Gilchrist has already made the point about distinguishing between seasonal and non-seasonal series. As I understand it, the first block of data in all the Tables contains two entirely disparate things: the application of non-seasonal methods to non-seasonal series and the application of non-seasonal methods to seasonal series. In the third block is the application of the non-seasonal methods to seasonally adjusted series. Presumably, the series that had no seasonal pattern are simply repeated from the first group.

Secondly, about the naive forecasts and stationarity. I assume that, although the authors talk about ARMA models, in fact they also include a good many non-stationary series. Therefore, my concept of a naive forecast for a series which requires first differencing would be

$$\Delta X_t = \Delta X_{t-1}.$$

I wonder, therefore, whether the authors, when referring to the Naive 1 and Naive 2 methods, have applied the appropriate naive method depending on the degree of differencing.

Nevertheless, it is surely surprising and disappointing that for the less noisy series the Naive 2 method gives the best results for nearly all the forecasting horizons, in terms of the m.a.p.e.—although, as Professor Priestley said, this may not be the appropriate criterion, given the way in which the series were fitted. I am drawing this conclusion from Table 2.

Does this cast some doubt on the adequacy of the seasonal adjustment? The method of seasonal adjustment described in Appendix B seems to be a version of X-9—that is, the method two before the present X-11 which most people use. It carries out smoothing by a 3×3 moving average. I wonder why they did not use the standard X-11 with 3×5 moving averages, and whether they included any modification for extremes.

There is a statement in Section 4, paragraph 5:

> "Smoothing methods on the original data are also relatively more accurate than on the same data after they have been seasonally adjusted, even when series with high randomness are excluded."

However, this is not true of the m.a.p.e. comparisons in Tables 2 and 3, nor of those based on the adjusted Theil coefficient in Table 8.

Dr C. CHATFIELD (Bath University): Today's paper provides further evidence that simple forecasting methods are often to be preferred to sophisticated procedures, a view that I have advanced for some time. Yet, paradoxically, I cannot help feeling that today's results go too far in their assessment of relative accuracies. My reservations about Box–Jenkins arise because the extra complexity of the method is not always justified by the improvement in accuracy, and because it is easy for the method to be misapplied. But I find it hard to believe that Box–Jenkins, if properly applied, can actually be worse than so many of the simple methods. Today's results certainly clash with the results of some previous empirical studies such as Newbold and Granger (1974) and Reid (1975).

Why do empirical studies sometimes give different answers? It may depend on the selected sample of time series, but I suspect it is more likely to depend on the skill of the analysts and on their individual interpretations of what is meant by Method X.

For example, consider Groff (1973) who found exponential smoothing to be as good as Box–Jenkins. If you read his paper you will see that Groff's version of "Box–Jenkins" bears little resemblance to that originally proposed by Box and Jenkins (1976). So we may disregard Groff's results except insofar as they tell us that the Box–Jenkins procedure may be misinterpreted.

Consider the study of Newbold and Granger (1974) where Box–Jenkins came out better than automatic forecasting procedures. The methods used in the study are clearly defined and I was able to carry out a follow-up study on some of the series which showed (Chatfield, 1978a) that a non-automatic version of the Holt–Winters procedure gives more accurate results than the automatic procedure. Thus in using a phrase like "The Holt–Winters Forecasting Procedure", it is clearly essential to define precisely what is intended.

One drawback of today's paper is that the forecasting procedures are not described in sufficient detail. For example, the authors' method 11, which I call the (automatic) Holt–Winters procedure, depends heavily on the starting values selected for the mean, trend and seasonals, but we are not told here, or in Makridakis and Wheelwright (1978a), what formulae are used.

I find Appendices A and B unhelpful. I do not understand Appendix B at all. Compare equations (A-46) with (A-49). The descriptions of the methods look confused and incomplete. For example, S_t sometimes denotes a seasonal factor and sometimes without warning a mean level. Formulae (A-10), the formula for b after (A-25), (A-42) and (A-44) look incorrect.

Because the procedures are imprecisely defined, it would be difficult to reproduce the authors' results, even if given access to the series. It therefore appears that the evidence in today's paper is inconclusive as to whether we should believe the results. This being so, the only alternative appears to be to look at the authors' other published work in order to build up a prior probability as to whether their results are likely to be accurate. It is, of course, unusual to comment on other work in this way, but today's empirical study seems to justify such an exception.

As it happens, I have recently (Chatfield, 1978b) been examining a forecasting procedure called "adaptive filtering" which has been proposed by Professor Makridakis in collaboration with S. C. Wheelwright. In common with a number of other writers (Ekern, 1976; Golder and Settle, 1976; Montgomery and Contreras, 1977) I was unhappy about the method, and I decided to check the claim in Makridakis and Wheelwright (1977) that adaptive filtering gives better forecasts than the Box–Jenkins procedure on the famous airline passengers data. While their comparisons were numerically correct, it turned out that their comparisons were made for a single base period. When forecasts were made for different base periods in the forecast period, it turned out that adaptive filtering was worse on average than both Box–Jenkins and Holt–Winters. Here it is relevant to point out that the forecast comparisons in today's paper are made for a single base period at time $(n-12)$ whereas the Newbold–Granger study, for example, made comparisons over all the forecast period. I would like to ask the speakers why they choose to make their comparisons in such an unusual way.

My views on today's paper are also influenced by the two books on forecasting co-authored by Professors Makridakis and Wheelwright. As expressed in my review (Chatfield, 1978c) of Wheelwright and Makridakis (1977), I formed the impression that these authors are more at home with simple procedures than with Box–Jenkins, whereas the reverse might be said of Newbold and Granger. Thus I would argue that empirical studies say more about the respective analysts than they do about the methods. I regret to conclude that I have to remain sceptical about today's results, particularly in regard to the conclusions relating to the Box–Jenkins procedure.

Mr G. Stern (ICL, London): The authors have supplied much interesting information, but I feel that their approach needs modification, otherwise the proposed examination of 1001 series could be a labour of Sisyphus. The basis of their approach is revealed in Section 1, where they contrast judgemental with mathematical forecasts. The same thought is carried on in later sections where they compare forecasting methods as if they were "treatments" in an experiment, to be applied blindly by an experimenter who is not supposed to take cognizance of anything except the rules for applying the method and the numbers being analysed.

But, to use an Ehrenbergism, we ought to preach what we practise, and I maintain that forecasting is frequently not an expert in forecasting techniques applying a method to data which he knows nothing about, but rather an expert in the data using mathematical methods and judgement to make a prediction. I suggest that the man who successfully forecasts sales of shoes for the next 12 months, month by month, is a shoe expert. He then applies simple methods which seem to have a physical meaning and to be appropriate such as a linear or logarithmic trend together with seasonal factors. He would then start "bending": he may decide to disregard or alter the 1974 figures because of the 3-day week, and he may alter the final forecast because he believes that there will be a boom next year. Economic forecasting works this way: Surrey (1971) points out that one can never go back over an econometric forecast and re-run it to see what the result would have been had one known what the values of the explanatory variables were, because the user of an econometric forecast always "bends". Again, I would maintain that Box–Jenkins, properly applied, is really a sophisticated way of doing this "bending" and I think this explains at least part of its success as compared to automatic methods. (I do not accept Groff's contrary findings as I believe he has not used Box–Jenkins correctly and I suspect this may apply to the authors' work also). My criticism of Box–Jenkins is that I suspect that a far simpler system along the lines sketched above would give nearly as good results and be accessible to more experts in the item being forecast. Chatfield (1978a) shows that Holt–Winters can be greatly improved by manual intervention.

The authors quote a study showing that purely judgemental forecasts are less good than mathematical forecasts, but they do not ask why. I suggest that the reason is that many time series are subject to high inflation rates and to multiplicative seasonability. I believe that a purely judgemental forecaster will often underestimate both these factors: however often he is told that 20 per cent inflation means a doubling over less than 4 years, he still resists accepting the result, and similarly people tend to feel that seasonal effects are additive when in fact they are multiplicative and so more important in an inflationary context. A better test would have been to compare "judgemental plus trend and seasonals" with the other methods. I would recommend something of this sort for the author's future 1001 series tests. I would like to see, for example, if Chatfield's non-automatic Holt–Winters performs as well as Box–Jenkins. I would like to see a comparison between, say, the UK government's model of the economy and a much simpler model with very few variables and manual intervention. And I think it would be valuable if the authors were to persuade experts in some of the series to forecast using only, say, trend and seasonal factors with added judgement, and see if any expert forecaster not expert in the data but using any methods can do better.

Dr F. H. Hansford-Miller (Inner London Education Authority): I should like to thank Professor Makridakis, and his co-author, for their excellent and important paper, which not only impinges on statistical theory, but upon us as citizens. Are we not all now very much the victims of forecasting? Forecasts are made, and are used for crucial decisions by politicians in Westminster and Brussels. They are widely discussed in press and television. In fact we are beset with forecasting. But this paper questions forecasting, and bearing in mind that in the 33 years since the last war, in this forecasting age, the United Kingdom has suffered a continual economic decline, then I for one give three cheers. In addition, the paper suggests the application of normal scientific experimental design to forecasting, with measures of unbiased testing of forecasts against subsequent reality, for success or failure. A long overdue reform.

With forecasting experts now seen to be fallible it seems to me that this must put a much greater dependence on our administrators—the decision makers—that they should be numerate and able to disentangle the various arguments put to them. This seems a vain hope. "Is there a scientist in the House?", asked MP Kenneth Warren recently in an article in *Computing Europe* (1978), and he gives the answer "But nobody, but nobody stoops to say he is a 'Scientist'". So what do they say they are? Just what are the numeracy or other qualifications for forecasting decision making of our present Westminster MP's?

Table D1 shows a simplified summary of the October 1974 MP's occupations, as given to *The Times* (1974). The largest group, with 108, are barristers and solicitors, and one would not include numeracy among their first qualifications; next follow teachers and lecturers, who are possibly numerate, but we do not know their subjects; the 33 engineers should know some mathematics, but the far greater number of journalists, farmers, trade union and party officials, mineworkers, and most of the others, seem very questionable as regards being highly numerate. Perhaps we should rely on the publican! For decision making, using complex, and now questionable forecasts, it is a singularly unimpressive list, and I believe we now need a far more highly qualified, and numerate, member, both at Westminster and at Strasbourg, if we are to be properly governed in the future.

Secondly, I think we should at once follow the example set in the paper and, in proper scientific fashion, test forecasts to see how they have turned out. To keep within the political spectrum, and as we are moving into Election Year, 1979, I have taken as an example a test of the main Public opinion polls for the October 1970 General Election. These were Business Decisions, Gallup, Louis Harris, Marplan, NOP, and ORC. Each of these organizations carried out at least two polls during the election period, and I have only time now to consider the final one, that nearest the actual poll, and these took place within the period October 2nd to 9th. I have again gone to *The Times* (1974) for my basic data.

TABLE D1

The professions of Westminster MP's elected in the General Election of October 1970

Profession	*Number of MP's*	*Profession*	*Number of MP's*
Barristers and solicitors	108	Clerical and technical workers	18
Teachers and lecturers	92	Mineworkers	17
Company directors	80	Underwriters and brokers	17
Journalists, publishers and Public relations officers	58	Accountants	13
Managers, executives and adminstrators	52	Railway and other manual workers	11
Other business workers	50	Doctors and surgeons	9
Engineers	33	Party officials	7
Farmers and landowners	26	Publicans	1
Trade Union officials	19		
		Total	611

Table D2 shows the Labour, Conservative and Liberal forecasts in terms of percentage votes for each of the six polls, the date of the fieldwork being given in the first column. These forecasts are compared with the General Election Actual Poll, which resulted in 39·3, 35·8 and 18·3 per cent for the three main parties respectively, the difference between the forecast and the actual percentage

TABLE D2

The final, or pre-election, set of Public opinion polls for the General Election of October 1970

The forecasting accuracy of Public opinion polls, October 1970 *General Election*

Date of fieldwork of poll	*Public opinion poll*	*Labour forecast*	*Conservative forecast*	*Liberal forecast*	*Inaccuracy total*	*Rank order of accuracy*
October 2nd	Business Decisions	40·0	35·5	20·0		
	—Inaccuracy	(+)0·7	(−)0·3	(+)1·7	2·7	1
October 2nd to 5th	NOP	45·5	31·0	19·5		
	—Inaccuracy	(+)6·2	(−)4·8	(+)1·2	12·2	6
October 7th to 8th	Gallup	41·5	36·0	19·0		
	—Inaccuracy	(+)2·2	(+)0·2	(+)0·7	3·1	2
October 8th	Marplan	43·0	33·3	19·5		
	—Inaccuracy	(+)3·7	(−)2·5	(+)1·2	7·4	5
October 8th to 9th	Louis Harris	43·0	34·6	19·3		
	—Inaccuracy	(+)3·7	(−)1·2	(+)1·0	5·9	4
October 9th	ORC	41·8	34·4	19·4		
	—Inaccuracy	(+)2·5	(−)1·4	(+)1·1	5·0	3
October 10th	General Election actual polls	39·3	35·8	18·3		
October 2nd to 9th	Opinion Polls average inaccuracy	3·17	1·73	1·15	6·05	

being the "Inaccuracy". The absolute average Inaccuracy of the Labour forecast was 3·17, ranging from 0·7 by Business Decisions to 6·2 by NOP. The Conservative forecasts were closer, with Average Inaccuracy 1·73, and range 0·2, by Gallup, to 4·8, by NOP. In contrast to Labour, and also Liberal, the Tory forecasts were underestimates, except for Gallup, as opposed to overestimates. The Liberal forecasts had Average Inaccuracy of 1·15, with the forecasts within a narrow range of 0·7 (Gallup) to 1·7 (Business Decisions).

So which was the best forecast? Column 6 shows the absolute Inaccuracy total for each, which can be compared with the average of 6·05. In first place—and the Rank order of accuracy is given in Column 7—we find Business Decisions, with an Inaccuracy total of only 2·7. Very close behind is

Gallup, with a 3·1 error, followed by ORC (5·0), Louis Harris (5·9), Marplan (7·4), and well behind, in sixth place, NOP, with 12·2 Inaccuracy total.

As a result of tonight's paper I hope that appropriate tests, similar to these, will be applied as normal routine to all forecasts, in whatever field, as checks on their accuracy. It would, I feel sure, prove a highly valuable exercise.

Dr D. J. Reid (Central Statistical Office and Bank of England): The authors have drawn attention to some apparent inconsistencies between their own results and some work done in the past by Granger and Newbold and by myself, so perhaps I may be allowed a brief comment.

The explanation that I was going to offer has been put far more eloquently than I could have done by Professor Priestley—simply that the authors are comparing different things on the basis of different criteria. It is hardly surprising, therefore, that the results differ.

Where I would differ from Professor Priestley is in his last remark, when he said that it was pointless to try and identify factors which account for forecast accuracy. It may be that by looking for such factors we are looking, albeit partially and inadequately, for evidence about whether the ARMA model fits the data. Insofar as that is what we are doing, perhaps these factors have some value.

I found it of considerable interest that despite the differences in approach of the authors' work and the work done at Nottingham, nevertheless we come up with the same factors as being important in determining which method should be used. I think we differ in the way in which those factors are interpreted, but the factors themselves are the same. In my own work, which I presented in the form of a decision tree—this has been reproduced rather more accessibly in Kendall (1973)—I listed the following:

1. How long is the series?
2. Are the data seasonal or non-seasonal?
3. Does the random component have high or low variance, compared with the other variability in the series?
4. Are there large peaks, non-stationarities and other discontinuities in the series?
5. Is one trying to predict long or short lead-times?

Almost exactly the same factors are mentioned by the authors in their paper tonight. So perhaps there is more common ground between these studies than appears superficially from the conclusions about which series does best on the respective criteria used.

Professor J. Durbin (London School of Economics): The discussion so far does not seem to me to have brought out what I feel to be the most interesting point in tonight's paper, namely the spectacular gains in forecasting seasonal series obtained by the authors by projecting seasonally-adjusted data and then re-applying the seasonal factor. I found this most unexpected—as I imagine did many others—because, as the authors say in their paper, they were able to find only one reference where this had been done previously in all their studies of comparative forecasting methods. I had rather taken for granted that what Box and Jenkins claimed was true, namely, that if there is a seasonal series, whether it is a changing seasonal or a constant seasonal, the most effective way of dealing with the seasonality is to take seasonal differences. Yet the authors have obtained much better performance from very simple, unweighted moving average—Method 15 in the Tables—than is given by the standard seasonal Box–Jenkins method.

I believe that is something which would be interesting to see followed up because, to me, it is counter-intuitive. For example, I would suggest to the authors that they make a non-seasonal Box–Jenkins analysis of the seasonally-adjusted series and then re-apply the seasonal factor. I find it hard to believe that their Method 15 could not be bettered by doing a Box–Jenkins analysis of the seasonally-adjusted series.

In reference to a point made by Mr Burman, my understanding of the paper was that the authors were assuming a constant seasonal. They state that they tried an X-11 type seasonal adjustment method and had found that it achieved no improvement over an ordinary ratio to moving average method, assuming a constant seasonal. This points to my main worry about this paper which is the extent to which the results are data-dependent. My experience of seasonal time series is that the main problems arise from changing seasonality. One reason for the spectacular success of the X-11 seasonal adjustment procedure, which is used all over the world for adjusting official series, is that it is very much concerned with making provision for changing seasonality.

The fact that the authors obtained no differences from the assumption of a constant seasonal, as compared with using X-11, suggests to me that perhaps their series had constant seasonals, and perhaps things would not be quite so good if their methods were tried on series with changing seasonals.

I should like, therefore, to suggest that in their reply to the discussion the authors might give us some more information about these series. I appreciate that it may not be possible to have a complete catalogue of the whole 111 series, but they could be divided into broad categories. For example, the X-11 program contains a test for seasonality; I should like to know how many of these 111 series are classified as non-seasonal, according to the so-called test for stable seasonality. Within each of the two classes, the seasonal and the non-seasonal, I should like to know what is the distribution of the number of observations in the series. This is very relevant for an assessment of the performance of the Box–Jenkins system. I should also like to see the distribution of the numbers of series that are monthly, quarterly, annual, etc., also a classification by the type of series—are they macroeconomic series, microeconomic, sales series and so on? Without some basic information of that kind, we cannot adequately assess the significance of the authors' results.

The following contributions were received in writing, after the meeting.

Mr O. D. ANDERSON (Nottingham): I do not wish to repeat the kind of remarks I made in Anderson (1977), especially as I hold the apparently heretical view (amongst academic statisticians, at any rate) that Spyros Makridakis does considerably more than most of us to improve the level of forecasting *in practice*. I therefore will not cavil with the present paper, except to say that I shall be surprised if it has any real impact on the statistical audience, towards which it appears so bravely aimed.

There are, however, three points on which I must comment, for fear that everyone else may feel they are too obvious to merit mention.

(1) Groff (1973) keeps on getting quoted as evidence that the Box–Jenkins approach is a trifle superfluous. It was mentioned during the Institute of Statisticians Forecasting Conference, in 1976, and again at the recent 1978 Cambridge Time Series Meeting. Unfortunately, Groff did not use the Box–Jenkins methodology—all he did was try out a specific selection of ARIMA models. This is rather like saying, "estimate some number (say, seven, in fact) by a sophisticated method which could give any integer, but only allow it to choose multiples of ten; and then compare its hamstrung performances with a less complicated approach, which can just estimate to the nearest 5". The simpler method evidently has a good chance of doing better than its hobbled competitor.

(2) Makridakis and Hibon should surely have been worried by the implications of their "Box–Jenkins" modelling. First time round, from 111 series, 83 gave residuals with none of their first 24 serial correlations significant at the 5 per cent level. Presumably, then, there were something like 4·8 significant residual correlation values, on average, for the remaining 28 series. This sort of effect was suggested in Anderson (1975, pp. 18–19) and is not too surprising for near perfect identification. However, eventually the authors ended up with 111 models, not one of which had a significant residual correlation amongst its first 24. This has to represent excessive global data mining, and perhaps helps explain why the forecasts were none too good. The moral is that, in the long run, even the perfect analyst should find about 5 per cent of his final model residual correlations are one-star significant.

(3) Ending on a more positive note, I believe that it is now easier to distinguish between pairs of similar "autoregressive" operators, such as $(1-B)$ and $(1-\phi B)$, with ϕ near to, but a little less than 1; and that there are usually extra clues about this in the serial correlations. Forthcoming publications, by myself and jointly with Jan de Gooijer (Amsterdam), establish the relevant theory and empirical evidence for such discrimination; and we hope also soon to have some firm indication of gains in forecasting performance, which can then result.

Dr R. T. BAILLIE (University of Aston Management Centre): The authors assert that quantitative forecasting methods can be classified as either "econometric (explanatory) or time series (mechanistic)". However, as has been noted by Prothero and Wallis (1976), these are not distinct alternative approaches, but merely different representations of the same structure which are inextricably linked via the notion of the final form of an econometric model. Similarly, multiple time series models may also be reduced to univariate ARIMA representations for each component series; see Chan and Wallis (1978). Many of the comparative papers referred to by the authors compare forecasts from ARIMA models with forecasts from econometric models with rather arbitrary dynamic specifications; hence the conclusions of Naylor *et al.* (1972) for example.

One of the main findings of the authors is that forecasts produced by exponentially smoothing an already seasonally adjusted series compare favourably with those from an ARIMA model. Unless the structure of the series changes between the sample and forecast periods, their finding must be due either to the failure of the ARIMA models to capture the seasonal effects as efficiently as the Census II method, or to their choice of a sub-optimal ARIMA model. As mentioned by the authors their findings differ from the substantial forecasting survey conducted by Newbold and Granger (1974). The authors' case is not helped by their failure to give details of length of series, characteristics of series and their approach in applying Box–Jenkins methodology. For example, did they implement the Box–Cox transform?

It is also unfortunate that the authors did not include in their survey at least some previously analysed series. Indeed, there remains the distinct possibility of another paper appearing to re-establish the superiority of ARIMA models on another data set, to be followed by yet another paper questioning this premise and based on yet another collection of series. Perhaps all such surveys should include some standard series. For example, how does exponential smoothing compare with ARIMA models when applied to the Wolfer sun spot series or the airline data in Box and Jenkins (1976)?

The authors introduce an interesting definition of forecasting accuracy. An alternative approach is to examine the mean squared error and to decompose it in terms of randomness in the model, prediction of exogenous variables and parameter estimation. For short series the latter component is likely to be important. Formal results for regression models with autoregressive errors are given in Baillie (1979) and with ARMA errors in Baillie (1977).

Dr M. BERTRAM (Scicon Computer Services Ltd, Milton Keynes): I should like to add my thanks to the authors for this interesting and provocative paper. With so many methods to choose from it is impossible to consider them all. But I should be interested in the authors' views on the Kalman filtering, mixed-model approach pioneered by Harrison and Stevens (1976).

We have practically applied this approach to a number of economic time series (mainly sales data) with reasonable success. The strength of this approach is that it allows for changes in the model structure of the kind one might expect in economic time series.

Professor D. R. Cox (Imperial College, London): The statistical approach, if there is such a thing, includes a large dose of empiricism, and the literature on time series being on the whole theoretical, a very empirical paper is to be welcomed. Yet, like several speakers, I have doubts about several points in the paper. The balance between 10 pages of tables of detailed conclusions, one short paragraph of vague description of the series analysed and nothing on their statistical characteristics makes discussion of the conclusions difficult; I hope the balance can be partly redressed in the authors' reply.

The authors' application of multiple regression in Section 3 seems especially uncritical. Theory and empirism alike suggest errors of forecasting behaving approximately like $a+b/n$, where n is the number of observations, and to fit a model linear in n is to fit an equation which inevitably has strange properties for large n. Also it seems odd not to express forecasting errors as a fraction of the intrinsic variability of the series.

Dr R. FILDES (Manchester Business School): In a discussion of time series methods of forecasting it is no surprise to find statistical arguments predominate and the "best" method is determined with respect to some convenient criterion such as expected quadratic loss. Unfortunately, the calculations are only valid as far as the statistical approximations in the argument permit and any declaration on relative model merit is only helpful insofar as the model on which the statistical arguments are based is correct.

Thus claims that the Holt–Winters variant of exponential smoothing as being a special case of Box–Jenkins are beside the point (as well as being technically incorrect). It is apparent from Newbold and Granger's study (1974), and supplemented now by the evidence presented in the paper under discussion, that the *ad hoc* approximations undertaken in initialization, together with the simpler structure of exponential smoothing, sometimes imply that in forecasting applications rather than within sample fitting, exponential smoothing can outperform its more sophisticated alternatives. The fact that simple exponential smoothing is a first-order moving average in first difference is to a large extent irrelevant. Thus we can conclude that the period of fit offers us only limited information, however processed, about forecasting effectiveness. If we adopt Priestley's (1974) categorization of the univariate forecasting problem we find that what is to be determined is "what class of random processes are we considering?" Only the data can tell us. Unfortunately economic series are not regularly describable as ARIMA processes, despite the apparent optimality of the appropriately identified estimated ARIMA model within that class. Paraphrasing Harrison and Stevens (1976) such well-behaved series seldom occur in practice.

Makridakis and Hibon are to be congratulated in addressing the right problem—establishing which methods work in practice and why. Of course their results do not show us which method to adopt any more than any other study; without a suitably defined population from which to make inferences this is impossible. What they do tell us is that large numbers of economic series are best forecast by simply, robust models, a conclusion my own work wholeheartedly supports (Fildes, 1979), and thoroughly in keeping with the approach of Harrison and Stevens (1976).

Theoretical analysis can only take us so far in the evaluation of a forecasting procedure. In the end what a forecaster needs to know is the likely performance of a particular method in the situation he faces. The authors have taken us some way towards a correct conceptualization of this important practical problem.

Professor ROBIN HOGARTH (INSEAD, France): Forecasting fascinates. The need people have to predict and control the environment guarantees both the interest in and importance of the paper by Makridakis and Hibon. The questions they address are crucial: What are the relative accuracies of different forecasting methods? Under what conditions are methods differentially accurate? The answers clearly depend on the *joint* characteristics of both the series predicted and the method used. Makridakis and Hibon mainly elaborate aspects of the former; my first point concerns the latter.

I believe one useful dimension for characterizing methods is degree of sensitivity of parameters to variations in the data. It used to be thought that sensitivity was desirable. However, recent work in the use of regression methods for repetitive predictions indicates the contrary. Specifically, in situations that Makridakis and Hibon would describe as exhibiting much randomness, simple unit- or equal-weighting procedures will outpredict models based on the "optimally" estimated least-squares weights. The empirical evidence and theoretical rationale for this statement are provided elsewhere (Dawes and Corrigan, 1974; Einhorn and Hogarth, 1975). However, an additional and, in applied situations, crucial point is that processes generating data rarely meet the assumptions implied by models. There is therefore a need for parameters to be robust against variations in the underlying data generating processes. Indeed, I strongly suspect that the relative success of simple-minded schemes often depends precisely on this factor. Furthermore, it would seem that simple-minded or "naive" models can, again relative to more complex schemes, simultaneously capitalize on two apparently contradictory features of time-series: first, unpredictable structural changes; and second, the sheer inertia of economic activity which induces high correlations between successive observations. Investigations of the robustness of simple methods are clearly required.

My second point is to support the authors in urging further research into methods for combining forecasts, including human intuition. Once again, simple-minded models, e.g. taking an arithmetic average, seem to be remarkably effective and the investigation of robustness is clearly also relevant here.

Finally, an issue that needs to be faced squarely is whether forecasting based on mechanistic methods is possible beyond short time horizons. As stated above, relative inertia of phenomena across time permits short-term forecasting. However, where are the limits? Interest in forecasting methodology attests to the fact that we need to believe we can make model-based forecasts. However, can we learn to live with the fact that often we cannot? The future is not necessarily "ours to see".

Professor COLIN LEWIS (University of Aston Management Centre): Whilst applauding the authors' comprehensive coverage of many of the forecasting models in everyday use, I would point out that the adaptive response rate exponential smoothing model adopted is in practice rarely used; the "delayed" version generally being preferred because of its lack of response to a single period impulse. In this delayed version the α_t of equation (A-11) is set equal to $|E_{t-1}/M_{t-1}|$ rather than $|E_t/M_t|$.

This could well explain the poor performance of this model in the authors' analyses.

The AUTHORS replied later, in writing, as follows.

We would first like to thank all discussants of our paper for their comments, the majority of which we have found extremely useful. We understand, of course, that the topics covered in our paper, and the conclusions reached, cannot fail to provoke disagreement from those who believed that their "favourite" method should have performed better. Furthermore, any empirical study of the sort undertaken cannot but raise many questions, since there is little guidance available to researchers as to how such studies should be done. We understand and accept, therefore, differences in opinions and we hope that further empirical studies will be conducted in a way such that a more standard methodology will become available for this type of research.

The main purpose of our research was not to compare various forecasting methods, but rather to find out the factors that affect *post-sample* forecasting accuracy, as opposed to being concerned with how well a forecasting model fits to a set of available data. Post-sample (i.e. beyond the available data on which the model has been developed) forecasting accuracy is of paramount importance, for practical purposes, since it is the only way to estimate forecasting errors and be able to evaluate when method A should take preference over method B, for a specific set of data. It is unfortunate that the discussion became concentrated on a comparison between simpler methods and the Box–Jenkins methodology to ARMA models, since this aspect of our paper was only secondary when the study was started. At that time, we believed, as most people involved with forecasting, that that particular question had been settled and that the Box–Jenkins methodology was the most accurate. To our surprise the results of our study showed otherwise; we could indeed have reacted in the same way as the Operational Research Society Forecasting Study Group (see Professor Gilchrist's contribution) and be so horrified by the results as to have "brushed them under the carpet, the carpet patted down, and no more said about them". However, we felt that this would be a disservice to the profession and the field of forecasting. Our main concern was that we must "shed light on the mystery of why, under certain circumstances, simpler methods do as well as or better than sophisticated ones"; similarly, we wanted to be able to know when one method should take preference over another, for a specific set of data, where post-sample accuracy is concerned.

In a recent paper entitled "Forecasting with econometric methods: folklore versus fact", Armstrong (1978) cites much evidence about 'the tendency of intelligent adults to avoid disconfirming evidence" (pp. 550–551). This is not something new, and Kuhn (1962) has shown its full implications for the evolution (or revolution as he argues) of science. We would like to point out that the empirical evidence which exists at present is extremely weak and conflicting. Most importantly, however, intelligent adults should not judge evidence by different criteria if it is discomforting to their own beliefs. It was suggested by several discussants that Groff's (1973) study should be disregarded as incorrect. This leaves the studies by Reid (1969) and Newbold and Granger (1974). Some say that these two studies should be counted as one, since both were conducted at Nottingham, at about the same time, without it ever having been made clear whether or not the data used was the same. More importantly, however, Reid's studies compare Box–Jenkins with *Generalized Exponential Smoothing*, which is no longer a method used to any great extent (if at all). Thus the work of Newbold and Granger remains as the only major study which compares Box–Jenkins with the Holt–Winters method. This is very little empirical evidence indeed. But even a close look at the Newbold and Granger study raises the same type of question that we have been asking in our paper. For instance, the percentage of occasions that Box–Jenkins outperforms the Holt–Winters model is 73, 64, 58, 58, 57, 58 and 58 per cent, for lead times of 1, 2, 3, 4, 5, 6, 7 and 8 periods. After a lead time of three periods, Box–Jenkins is no better than Holt–Winters but for 8 per cent of the series. How can this be explained? Is the extra accuracy worth the much higher costs involved (in terms of both personal effort and computer cost)? Which is the 42 per cent of the series where Holt–Winters outperforms Box–Jenkins? Newbold and Granger do not provide detailed information on accuracy measures other than the percentage of time that method A did better than B, and as Carbone (1978) has shown, this type of accuracy measure can be misleading. Thus, even the Newbold–Granger study does not answer the question as to why, in such a high percentage of series (i.e. 42 per cent), exponential smoothing (supposedly a special case of Box–Jenkins) does better than Box–Jenkins.

Armstrong concluded that "a review of the published empirical evidence yielded little support for . . . two important questions. First, do econometric methods provide the most accurate way to obtain short-range forecasts? Second, do complex econometric methods provide more accurate forecasts than simple econometric methods"? Armstrong's conclusions are particularly relevant for time series forecasting because he compared econometric methods with, principally, ARMA models. Secondly, his conclusion that higher complexity does not necessarily lead to greater accuracy is a lesson that time-series forecasters should keep in mind. Even though it is difficult to accept Armstrong's flamboyant style, we can but accept his main conclusions. Econometricians, obviously, cannot be pleased with the implications drawn by Armstrong. Our belief is that time-series forecasting must avoid the mistakes of econometricians: firstly, empirical evidence should not be ignored, but instead used to guide the development of theory; and secondly, it is not necessarily true that more complexity will result in better post-sample forecasting accuracy.

Dr Gilchrist's thoughtful comments have captured the essence of our paper. He raises several important questions and we will try to answer them as best we can. First, we are thankful for his bringing to the attention of forecasters what has been "under the carpet" concerning the study undertaken by the Operational Research Society Forecasting Study Group. This is obviously a reconfirmation of our point that for *some* series (in particular those which are unstable) random walk (or Naive methods) do as well as, or better than, complex methods. Secondly, he argues, and we agree, that a distinction between seasonal and non-seasonal methods should have been made. We simply refrained from doing so because of the extra tables that would have been required, which would have further increased the size of an already large paper. However, there is a more fundamental reason. Once the data for seasonal series have been deseasonalized (lower part of all tables), then non-seasonal methods can be compared, on an equal basis, with seasonal ones. This provides us with all the comparisons we need. In this respect, methods 1 to 10 in the tables (Original Data: Nonseasonal Methods) are not needed other than to provide us with a yardstick for how much improvement has been achieved by deseasonalization of the data. This information can tell a user, who might not want to go through the trouble of deseasonalizing the data, how much he would lose by not doing so, and was provided only for comparative purposes—no more. Thirdly, Professor Gilchrist is right when he assumes that we only used one-step-ahead forecasts. Unfortunately, the time required to re-run the models for 2, 3, 4, ..., 12 lead times is enormous: new models must be identified, the differencing might change, the parameters must be re-estimated, etc. The amount of work involved would be multiplied by twelve, and in a huge study of the type we undertook, this is very difficult on practical grounds. However, this deficiency should be corrected in future studies. Fourthly, concerning the accuracy measures, we agree with him that it might have been better if we had not averaged the various errors for forecasting horizons larger than one period ahead. We in fact did this because practitioners care much more about the average errors (e.g. in budgeting) up to a certain month of the year than single error values. We also share his view that "the dice are very much loaded in favor of ARMA models", and as Chatfield (1978a) has shown, for instance, some subjective modifications to the Holt–Winters automatic version could substantially improve the results. What the effect of this is on the comparisons, we do not know. Finally, we can state that backwards forecasting has little effect on post-sample forecasting accuracy (see Makridakis, 1978). This was one of the first avenues we explored in order to explain the poor performance of ARMA models.

Professor Priestley raises a fundamental question which we believe has extremely important implications for the field of forecasting and statistics in general. To quote him, "*If the series conforms to an* ARMA *model, and the model has been fitted correctly, then the forecast based on this* ARMA *model must, by definition, be optimal.*" This is, obviously, a correct statement, of which every student of statistics is well aware. The statement, however, is based on an assumption, namely the *assumption of constancy*. The least-squares approach—widely used, not only in forecasting, but in the whole field of statistics—fits a model to a set of data. Consequently, *if* the pattern, or the structure, of these data *does not* change, *then* the same model can be used to estimate or forecast beyond the data used (i.e. the post-sample period). Many data series are structurally stable, in which case the assumption of constancy holds. This means that minimizing the mean square error of the sample data also minimizes the mean square error of the post-sample predictions. Unfortunately, however, most of the structurally stable series are in the fields of physical sciences and engineering, and very few in the field of social sciences, business or economic data. It might, therefore, be that ARMA models are more appropriate for engineering applications, where the great majority of series are structurally stable, while exponential smoothing models are better suited to business and economic applications where most series show discontinuities and changes in patterns, and are, therefore, structurally unstable. Obviously, this is only a conjecture at this point, which must be tested in the future, but is, we believe, worth considering. There is a fact that Professor Priestley must accept: empirical evidence is in *disagreement* with his theoretical arguments. Our study is not the only one which has shown that simpler methods do as well as, or better than, sophisticated ones. For instance, Chatfield and Prothero (1973), Groff (1973), Dawes and Corrigan (1974), Geurts and Ibrahim (1975), McWhorter (1975), Narasimham *et al.* (1975) and McCoubrey and McKenzie (1975) have reached similar conclusions; furthermore, our main paper, Armstrong (1978) and Hogarth and Makridakis (1979) cite many more studies which have found that more complex models are no more accurate than simpler ones. Finally, there are also other areas where simple models have been shown to be more robust, and at least as accurate as complex ones (for a review, see Leung, 1978). In theory, this should never have happened, but in practice it does. The reasons are obvious to us: minimizing mean square errors during the model fitting phase does *not* guarantee minimum mean square errors in the post-sample phase simply because real-life data are often structurally unstable. The need, therefore, is for methods which can minimize the post-sample forecasting accuracy (see Fildes and Howell, 1977) and can deal with other types of data than those currently used (e.g. the Airline data) to test statistical methods (Harrison and Stevens, 1976; Fildes, 1979).

We do not understand Professor Priestley's comments saying that we have not used a quadratic loss function. We indeed used Theil's U-coefficient (see Tables 7 and 8). As a matter of fact, ARMA models do even worse than simpler methods when the quadratic loss function is involved. Interestingly enough, however, the mean square error (as measured by the U-statistic) is at its smallest during the model fitting phase, whilst being very large in the post-sample period. Moreover, we did not use a straightforward mean square error measure because the results would not have been comparable with those of *other* studies, since mean square error measures are absolute, depending upon the the specific series used.

Professor Priestley makes several comments on our construction of the regression equations. We believe that the procedure used in developing these regression equations is standard in the social sciences. We did not start with more than ten variables and we had close to 1000 observations for each equation of Table 10. Regression results with such a large number of observations cannot be questioned. Furthermore, there is a remarkable consistency in the variables included in the equations—see Tables 9 and 10—and their values; the chances that such results can be attributed to chance are astronomically low. Concerning the variable of "number of data points", perhaps it should have been included, and we regret we did not do so. But even if we had included the variable, nothing would have changed.

Mr Burman suggests that we did not present enough information about how the various models were developed and used. This might be so, but we have still provided more details (see Appendix A) than any previous study comparing forecasting accuracies. Furthermore, the book *Interactive Forecasting: Univariate and Multivariate Methods*, by Makridakis and Wheelwright (1978), provides additional information about the methods used and the algorithms employed. Unfortunately, space constraints made it impossible to provide more information about the methods used. But we can assure Mr Burman that we devoted ample time to identifying an appropriate ARMA model (see Appendix A) while using the Box–Jenkins methodology. Concerning number of data points, Table D3 below gives their distribution.

We would like to thank Dr Chatfield for accepting our study as a corroboration of the view that he has advanced for some time; namely that "simple forecasting methods are often to be preferred to sophisticated procedures". We can only apologize because our "results go too far" in the opposite direction. Unfortunately, we had no control over the findings. Furthermore, he argues that some of the results obtained are due to not knowing how to apply the Box–Jenkins methodology properly. This is an easy accusation to make, as Dr Chatfield well knows, after the discussion of his paper with Prothero (1973). Obviously, there is no way of proving or disproving that we properly applied the Box–Jenkins methodology; on the other hand, we have had experience in using it with hundreds of real-life series; moreover, we can say that we did the best we could to apply it as correctly as possible, spending an enormous amount of time in doing so. Dr Chatfield argues that we might not have used the right starting values, for what he calls the Holt–Winters procedure. Had we used more appropriate starting values, however, the Holt–Winters method would have done even better than the Box

TABLE D3

Distribution of number of data in the 111 *series*

Number of data points	*Frequency*
less than 40	3
40–49	7
50–59	5
60–69	1
70–79	9
80–89	30
90–99	10
100–109	8
110–119	4
120–129	7
130–139	13
140–149	13
more than 150	1
	111

Jenkins methodology. We do not, therefore, understand the reason for this type of argument except that it points out that Chatfield (1978) has done some work on the Holt–Winters method.

Finally, Dr Chatfield expresses some personal views about the first author and Professor Wheelwright. (Mrs Hibon has difficulty seeing the relevance of these points to the present paper.) He talks about "prior probabilities". It might be useful for Dr Chatfield to read some of the psychological literature quoted in the main paper, and he can then learn a little more about biases and how they affect prior probabilities. Specifically, he mentions adaptive filtering. Even though Professors Makridakis and Wheelwright are flattered about the repeated interest of Dr Chatfield (Chatfield and Newbold, 1974; Chatfield, 1978) in this technique, they fail to see its relevance to the present paper. In any event, the answers to the questions he raises on adaptive filtering are as follows: The comparisons made with the Box–Jenkins methodology are based on a single period because this is how the Airline data were forecast by Box–Jenkins (1976, pp. 307–308). If other base periods were taken, no comparisons would have been possible, without the risk of being told that the Box–Jenkins methodology was incorrectly used. (Needless to say, his comment that the comparisons, numerically correct, which show that adaptive filtering—for the base period used—is more accurate than the Box–Jenkins methodology is appreciated). Dr Chatfield also expresses certain further views about adaptive filtering. There is some additional evidence that he might wish to consult. Firstly, there was a "forecasting tournament" organized by Professor Dave Pack for the Los Angeles ORSA/TIMS meeting of November 1978. There were four series used and adaptive filtering came second, in most measures of accuracy, behind state space forecasting. Box–Jenkins came last (see Carbone, 1978; Granger and McCollister, 1978). Secondly, there have been several papers (see Bretschneider *et al.*, 1978; De Lange and Wojchiechowicz, 1974; Nau and Oliver, 1978) that have established the theoretical correctness of adaptive filtering and its practical usefulness; papers of which he should be aware, due to the great, and repeated, interest he has shown in adaptive filtering.

Another piece of evidence cited by Dr Chatfield (1978) is his review of *Forecasting Methods for Management*. Dr Chatfield should bear in mind that this book was written in 1971—the part dealing with the Box–Jenkins methodology was not revised when the second edition was published in 1977. A more correct illustration of how at least Professor Makridakis can apply the Box–Jenkins methodology is to be found in Makridakis and Wheelwright (1978a, b), or in the forthcoming third edition of *Forecasting Methods for Management*, in which a revised chapter for the Box–Jenkins method has been written. Finally, he might want to see some other reviews of *Forecasting Methods for Management* (McR, 1974; Benson, 1975; Brown, 1977) and then he will perhaps understand how prior probabilities can be biased.

We appreciate Mr Stern's suggestions for how to go about replicating our present study with the 1001 series we have already collected. Furthermore, we will propose to him, and to anyone else who would be willing to participate, to use our data. We would also like to see how the improved version of the Holt–Winters method would do, and we hope that Dr Chatfield will also participate. It would indeed be desirable to have access to more information than is available, but unfortunately empirical studies are extremely difficult to conduct and they require countless days of hard work. Moreover, since many difficult decisions have to be made, on many unresolved problems, criticism is easy. Mr Stern criticizes our study for being "applied blindly" without subjective adjustments. First of all, we never pretended to have done otherwise. Secondly, it is not at all clear from the psychological literature dealing with human judgement that subjective interventions will improve the results. As unlikely and unintuitive this statement may seem, it is accepted as a well-established fact amongst psychologists. Reading Slovic (1972), Dawes and Corrigan (1974) or Slovic *et al.* (1977) will easily establish the belief of psychologists concerning human judgement, as compared to quantitative models.

We cannot disagree with Dr Hansford-Miller, and we would like to thank him for enlarging the scope of our study by relating it to many additional implications.

We greatly appreciate Dr Reid's comment "perhaps there is more common ground between these studies [ours and previous ones] than appears superficially from the conclusions about which series does best on the respective criteria used". Our main purpose was not that of comparing various forecasting methods and we therefore hope that some logical explanations will be found. Our results have been checked and rechecked, and leave us in no doubt as to their correctness. Thus, we might not have used the best starting values for the Holt–Winters model, or applied the best procedure for other methods, but given the way the various methods were used (see Appendix A), the results are correct and in close agreement with programs other than ours, with which we have checked the numerical answers. As a most critical point, we believe that the forecasting accuracy of various methods differs significantly depending upon the characteristics of the series (point No. 3 in Dr Reid's reply). Unfortunately, we know very little about this particular factor, as well as the others mentioned by Dr Reid. In our opinion, the challenge is to quantify and measure the influence of these factors; hence our present study, which is unique in trying to identify and measure such criteria as those mentioned by Dr Reid.

We would like to thank Professor Durbin for the insight of his comments, and for re-emphasizing that no tests concerning seasonally-adjusted data (except involving a single series) have been reported in the literature. We think it is a good idea to apply Box–Jenkins to seasonally-adjusted data and we will try to follow his advice. We indeed hope that the results of our study are data-dependent, but still maintain that we must know which types of data enable some methods to perform better than others. Table D3 gives the distribution of the number of data points. About 80 per cent of the series were monthly, 12 per cent quarterly and 8 per cent annual. About one-third of the series were macroeconomic and two-thirds were taken from business sources—some of them originating from industry. About two-thirds of the series were seasonal. It is difficult to assess the ending date of the series (we did not record this information), but we would estimate that about half of them ended between 1974 and 1976, thus being influenced by the 1974–75 recession.

Most of the points made by Dr Baillie have already been dealt with—we will not therefore repeat them. His comment about the structure of the series changing is correct. Many of our series were not structurally stable; but this should not surprise anyone, since real economic and business series are often structurally unstable. Dr Baillie regrets that we did not include in our survey "at least some previously analysed series". We regret it too, since we wrote to all those who had carried out previous surveys, but we received no series—not even a single reply. We indeed, however, included the famous Airline data whose m.a.p.e., for a number of methods, is shown in Table D4. It should be noted that 90 data points were used for model fitting. Thus the 12 forecasts refer to periods 91 to 102.

Mr Anderson is very generous in the opinion he expresses about Spyros Makridakis. Concerning his remarks made in his 1977 paper, there is no point in arguing here since there has been a reply (Makridakis, 1978). However, we would like to point out to Mr Anderson that it is a little premature to talk now about the "real impact" that our study will have. Any good forecaster must know that prejudgements of this type have little value, and add nothing to the discussion. Concerning his specific points, our answers are as follows: (1) We do not want to make a value judgement of whether or not Groff's study was correctly conducted; in the same way that we do not do so for Reid's (1969) or Newbold and Granger's (1974) studies. What we fail to see, however, is why nobody questions Newbold and Granger, for instance, on whether or not they applied the various methods correctly (or is it, as we have already argued, because their conclusions are in agreement

TABLE D4

Mean average percentage error (m.a.p.e.) of the Airline data
(errors refer to periods 91–102)

Forecasting method	*Model fitting*	*Forecasting horizons* 1	2	3	4	5	6	9	12
Naive 1	8·56	9·44	8·55	7·48	11·17	16·54	17·48	16·99	14·76
Single exponential smoothing	8·56	9·94	8·55	7·48	11·17	16·54	17·48	16·99	14·76
Brown's linear exponential smoothing	9·59	3·13	4·72	13·09	24·97	38·49	45·49	60·28	67·41
Holt's linear exponential smoothing	8·60	8·94	7·72	7·62	12·07	18·18	19·68	20·54	18·09
Linear trend	10·72	25·77	24·76	20·63	16·07	16·14	14·09	11·34	11·42
Harrison's harmonic smoothing	3·64	3·46	2·67	2·01	2·11	2·07	2·26	3·13	3·03
Winter's exponential smoothing	3·72	1·26	·85	2·15	3·81	4·75	5·75	7·59	7·15
Adaptive filtering	3·60	2·36	2·48	1·76	1·50	1·48	1·49	1·64	1·87
Box–Jenkins	3·58	1·32	1·10	1·60	2·45	2·64	3·11	4·05	4·16
Naive 2	3·02	1·75	·93	1·00	1·35	1·24	1·14	1·34	2·66
Exponential smoothing (seasonal adjustment)	3·00	2·30	1·18	1·17	1·47	1·34	1·22	1·40	2·70
Brown's exponential smoothing (seasonal adjustment)	3·00	2·82	1·65	2·10	3·02	3·53	4·09	5·94	5·95
Holt's exponential smoothing (seasonal adjustment)	3·10	1·86	1·46	2·03	2·91	3·31	3·74	5·08	4·78
Linear trend (seasonal adjustment)	4·19	11·30	10·14	9·21	8·29	7·90	7·56	6·61	7·35

with preconceived notions?). They gave no information as to how the different methods were used, the algorithms employed, the starting values, etc. How can it be known, then, that what they did was correct? (This is not to suggest that it was not.) (2) We believe that Mr Anderson did not read carefully the description of how we used the Box–Jenkins methodology. Thus, when he says that "The moral is that, in the long run, even the perfect analyst should find about 5 per cent of his final model residual correlations are one-star significant", he is testifying exactly to what we describe in Appendix A, where we state: "There were rarely autocorrelations [of the residuals] outside the 95 per cent confidence limits. . . ." If this is not what is represented in the above statement by Anderson, then perhaps we have been let down by our knowledge of English, which after all is not our native language. (3) We fail to see the connection of this point with our paper, and as before we will let the acid test of time judge the importance of Anderson's positive note.

We thank Dr Bertram for his suggestion, and we do hope that Bayesian Forecasting will be included in the next comparative study. Unfortunately, when the present study was conducted, no computer programs were available to us, hence we were unable to include the approach of Harrison and Stevens.

We hope that our reply has already answered the several questions raised by Professor Cox; there is no point in repeating them here. Concerning forecasting errors and how they behave, we can only say that n—for the data we used—did not have a statistically significant effect on forecasting accuracy, but for a very slight number of exceptions. Yes, indeed we do "express forecasting errors as a fraction of the intrinsic variability of the series"; this is precisely what most of the independent variables are (e.g. the mean absolute percentage change in randomness).

We are in complete agreement with Professor Fildes' well thought-out comments. His point that "the period of fit offers us only limited information, however processed, about forecasting effectiveness" needs to be framed and put before anyone working in forecasting because, somehow, training and habit lead all of us to ignore this fact; namely, that with real-life data model fitting and post-sample forecasting are not necessarily the same.

TABLE D5

Time interval between successive observations	*Types of time series data*								
	Micro-data				*Macro-data*				
	Total firm I	*Major divisions II*	*Below major divisions III*	*Industry*	*GNP or its major components I*	*Below GNP or its major components II*	*Demographic*	*Total*	
Yearly	17	30	11	29	28	30	30		175
Quarterly									
Ending date									
July 1969–December 1970	3	3	3	3	15	23		70	
July 1974–December 1975									
Ending dates other than							40		200
July 1969–December 1970	4	15	13	13	34	31		130	
July 1974–December 1975									
Quarterly									
Ending date									
July 1969–December 1970	4	23	33	15	20	30		163	
July 1974–December 1975									
Ending dates other than							76		626
July 1969–December 1970	12	49	90	165	48	61		463	
July 1974–December 1975									
Total	40	120	150	225	145	175	146		1001

We would like to thank our colleague, Robin Hogarth, for his comments, with which we wholeheartedly agree. Most importantly, however, we would like to thank him for making us aware of the huge psychological literature dealing with human judgement and its comparison with quantitative models. Somehow, in getting to know this literature, as well as looking at additional comparisons between time-series and other methods, one cannot but summarize the findings as "there is some bad news and some good news for time series forecasters": Well, the bad news is that simpler time-series methods do as well as or better than more sophisticated ones, whilst the good news is that time-series methods (for the short term) do at least as well as all alternatives (including human judgement) for which comparisons are available.

Table D5 shows 1001 real-life time series, collected on a quota sampling. They come from countries all over the world, and have various ending dates, as well as the different characteristics listed as the headings of Table D5. We believe it is imperative to put an end to the argument of what method(s) is (are) better than others and under which circumstances. We have, therefore, decided to make these data available to all researchers. (A tape containing these data can be obtained, at cost, from Tim Davidson at Applied Decision Systems, Temple, Barker & Sloane, Inc., 33 Hayden Avenue, Lexington, MA 02163, USA.) We hope others will contribute to enlarge this data bank and provide a realistic basis on which time-series forecasting methods can be tested. In addition, we have asked recognized experts in the various forecasting methods to forecast each of the 1001 series. Doing so will introduce an indisputable level of objectivity, and eliminate any accusations that the different methods have not been properly used. Up to now, we have received positive replies from:

1. Cameron–Mehra for State space forecasting;
2. Carbone for Adaptive estimation procedure;
3. Fildes–Stevens for Bayesian forecasting;
4. Johnson–Montgomery for the various Exponential smoothing methods;
5. Reilly for the automatic Box–Jenkins methodology;
6. Wheelwright for Adaptive filtering.

We hope that the remaining invited persons will join [eight more persons have joined at the time of correcting the proofs] this "forecasting competition", and that the final results will be available to the profession not later than the end of 1980.

References in the Discussion

Abdrabbo, N. A. and Priestley, M. B. (1967). On the prediction of non-stationary processes. *J. R. Statist. Soc.* B, **29**, 570–585.

Anderson, O. D. (1975). *Time Series Analysis and Forecasting: The Box–Jenkins Approach.* London: Butterworth.

—— (1977). A commentary on "*A Survey of Time Series*". *Int. Statist. Rev.*, **45**, 273–297.

Armstrong, J. C. (1978). Forecasting with econometric methods: folklore versus fact. *J. Business*, **S1**, 549–600.

Baillie, R. T. (1977). Asymptotic prediction mean square error for dynamic single equation models. Working paper No. 107, Aston Management Centre.

—— (1979). The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors. *J. Amer. Statist. Ass.*, **74**, 175–184.

Benson, P. H. (1975). Book review of *Forecasting Methods for Management. Interfaces, J. Marketing Res.*, **XII**, 492.

Box, G. E. P. and Jenkins, G. M. (1976). *Times Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

Bretschneider, S. *et al.* (1979). An adaptive approach to time series forecasting. *Decision Sci* (in press).

Brown, W. J. (1977). Book review of *Forecasting Methods for Management. Business Econ.*, May 1977, 62–63.

Carbone, R. (1978). On the performance of the adaptive methods in the L.A. competition. Paper presented at the TIMS/ORSA National Conference, Los Angeles.

Chan, W.-Y. T. and Wallis, K. F. (1978). Multiple time series modelling: another look at the mink–muskrat interaction. *Appl. Statist.*, **27**, 168–175.

Chatfield, C. (1978a). The Holt–Winters forecasting procedure. *Appl. Statist.*, **27**, 264–279.

—— (1978b). Adaptive filtering: a critical assessment. *J. Oper. Res. Soc.*, **29**, 891–896.

—— (1978c). Book review of *Forecasting Methods for Management. J. R. Statist. Soc.* A, **141**, 113–114.

Computing Europe (1978). Article entitled "In Parliament—is there a scientist in the House?" **6**, No. 23 (issue of June 8th), p. 14.

De Lange, P. W. and Wojchiechowicz, K. (1974). Application of adaptive filtering forecasting to the gold bullion price. *Business Manag.*, August 1974, 3–6.
Einhorn, H. J. and Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organiz. Behav. and Human Perform.*, **13**, 171–192.
Ekern, S. (1976). Forecasting with adaptive filtering. *Oper. Res. Q.*, **27**, 705–715.
Fildes, R. (1979). Quantitative forecasting—the state of the art. *J. Oper. Res. Soc.*, **30** (in press).
Fildes, R. A. and Howell, S. (1979). On selecting a forecasting model. *TIMS Studies in Manag.*, **12**, (in press).
Golder, E. R. and Settle, J. G. (1976). On adaptive filtering. *Oper. Res. Q.*, **27**, 857–868.
Granger, C. W. J. and McCollister, G. (1978). Comparison of forecasts of selected series by adaptive, Box–Jenkins and state space methods. Paper presented in the TIMS/ORSA National Conference, Los Angeles.
Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with Discussion). *J. R. Statist. Soc.* B, **38** 205 247
Hogarth, R. M. and Makridakis, S. (1979). Forecasting and planning: an assessment and some suggestions INSEAD Working paper 185.
Kendall, M. G. (1973). *Time Series.* London: Griffin.
Kuhn, T. S. (1962). *The Structure of Scientific Revolution.* Chicago: University of Chicago Press.
Leung, P. (1978). Sensitivity analysis of the effect of variations in the form and parameters of a multivariate utility model: a survey. *Behav. Sci.*, **23**, 478–485.
McR, M. J.(1974). Book review of *Forecasting Methods for Management. Interfaces, J. Marketing Res.*, **XI**, 67–69.
Makridakis, S. (1978). Time series analysis and forecasting: an update and evaluation. *Int. Statist. Rev.*, **46**, 255–278.
Makridakis, S. and Wheelwright, S. C. (1977). Adaptive filtering: an integrated autoregressive-moving average filter for time series forecasting. *Oper. Res. Q.*, **28**, 425–437.
—— (1978a). *Interactive Forecasting: Univariate and Multivariate Methods.* San Francisco: Holden Day.
—— (1978b). *Forecasting: Methods and Applications.* New York: Wiley.
Montgomery, D. C. and Contreras, L. E. (1977). A note on forecasting with adaptive filtering. *Oper. Res. Q.*, **28**, 87–91.
Nau, R. F. and Oliver, R. M. (1978). Adaptive filtering revisited. University of California Working paper.
Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts (with Discussion). *J. R. Statist. Soc.* A, **137**, 131–165.
Priestley, M. B. (1974). Discussion contribution to the paper by Newbold and Granger. *J. R. Statist. Soc.* A, **137**, 152–153.
Prothero, D. L. and Wallis, K. F. (1976). Modelling macroeconomic time series (with Discussion). *J. R. Statist. Soc.* A, **139**, 468–500.
Reid, D. J. (1975). A review of short-term projection techniques. In *Practical Aspects of Forecasting* (H. A. Gordon, ed.), pp. 8–25. London: Operational Research Society.
Slovic, P., Fischhoff, B. and Lichtenstein, S. (1977). Behavioral decision theory. *Ann. Rev. Psychol.*, **28**, 1–39.
Surrey, M. J. C. (1971). *The Analysis and Forecasting of the British Economy.* Cambridge: Cambridge University Press. (N.I.E.S.R. Occasional papers XXV.) *The Times* (1974). *The Times Guide to the House of Commons, October*, 1974. London: Times Books.
Wheelwright, S. C. and Makridakis, S. (1977). *Forecasting Methods for Management.* New York: Wiley.

As a result of the ballot held during the meeting, the following were elected Fellows of the Society.

Affleck-Graves, John F.
Agbeyegbe, Terence D.
Amphlett, Gillian E.
Beaumont, Christopher D.
Bulmer, Martin
Burdon, Michael W.
Cave, Christopher J.
Connolly, Kevin B.
Davies, Leslie A.
Drury, Michael R.
Evans, Michael
Farrow, Malcolm
Forster, Janet E.
Ghosh, Dhirendra N.
Godward, Ernest W.
Helks, Christopher J.
Holliday, Frederick M.
Hope, James
Humble, Stephen
Imeson, John D.
Jones, David R.
Jones, Keith F.
Kurian, Thomas
Lynch, Robert M.
McBean, Patricia A.
Martin, George R.
Morley, Kenneth R.
Ogston, Simon A.
Oyejola, Benjamin A.
Pillai, R. Krishna
Simpson, Stephen N.
Thompson, Simon G.
Wallis, Walter D.
Wilson, Arthur J. C.
Wright, Peter A.
Wu, Chien-Fu