AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

MONASH

WP# 3/92

MONASH
UNIVERSITY

AUSTRALIA

CHOICE OF TIME-SERIES FORECASTING METHOD USING DISCRIMINANT SCORES

C. Shah and M.L. King

Working Paper No. 3/92

February 1992

GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS
LIBRARY
WITHDRAWN
MAY 06 1992

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0353 6

CHOICE OF TIME-SERIES FORECASTING METHOD USING DISCRIMINANT SCORES

C. Shah and M.L. King

Working Paper No. 3/92

February 1992

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS COMMERCE & MANAGEMENT

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

CHOICE OF TIME SERIES FORECASTING METHOD USING DISCRIMINANT SCORES

by

C. Shah and M.L. King
Department of Econometrics
Monash University
Wellington Road, Clayton, Vic. 3168

ABSTRACT

A lot of research has been done on comparing the forecasting accuracy of different univariate time series forecasting methods. The biggest such study, using empirical data, was undertaken by Makridakis *et al.* (1982). The evidence from such comparative studies indicate that there is not one "best" method for all kinds of data. Furthermore, there also seems to be evidence to suggest that the simpler methods, such as exponential smoothing, often perform as well as or even better than the more complex methods. This is particularly true for short term forecasting. Unfortunately, there has been limited success in identifying the factors that contribute to the relative advantage of one method over another. Consequently a practitioner is still faced with the problem of objectively choosing one out of several methods available to use in forecasting a given time series. In this paper we address this problem by considering certain characteristics of a time series in order to calculate its discriminant score. This score is then used to calculate the probability of a particular method being "best" in forecasting that series. Three forecasting methods, simple exponential smoothing, Holt-Winters method and basic structural time series model using the STAMP package, are considered. Quarterly time series from Makridakis *et al.* (1982) "M-Competition" are used as data.

1. INTRODUCTION

Forecasting economic time series can be and is often a hazardous task. There are numerous methods to choose from, ranging from the sophisticated and complex econometric methods to the simple heuristic methods based on exponential smoothing. Over the years a lot of research effort has gone into improving these methods, developing variations of them and comparing their relative forecasting performance. Some of the major studies comparing the accuracy of different univariate forecasting methods have been done by Reid (1975), Newbold and Granger (1974), Makridakis and Hibon (1979), and Makridakis *et al.* (1982, 1984). The results from these studies have not been consistent. For example, Newbold and Granger (1974) found that forecasts using the Box-Jenkins method seem to be generally better than forecasts using the Holt-Winters method, while Makridakis *et al.* (1982) found that simpler methods, such as the class of exponential smoothing, compare rather well with the more statistically sophisticated methods when there is a large degree of 'randomness' in a series. As Chatfield (1988) points out, the lack of consistency in the results is not surprising given different analysts and different data sets.

The question that begs an answer is: what features of a given time series determine what method is most likely to produce the most accurate forecasts? Most studies seem to give general rules of thumb, that are based on frequency counts, to answer the above question. As D.J. Reid pointed out in the discussion following the Newbold and Granger (1974) paper that compared forecasting performance of three univariate methods: "Frequencies alone are of limited usefulness in this context and it would have been more interesting to have seen various multivariate tools, such as cluster analysis or factor analysis, applied to the

results of the study to try and find reasons for differences in performance on different series."

In this paper we attempt to go further than merely giving general rules of thumb to somebody faced with the decision of choosing the most accurate forecasting method for a particular univariate time series. We propose to use discriminant analysis to generate the probability of a given method producing the most accurate post sample forecasts. Section 2 describes the data set that is used in this study. We have considered only three forecasting methods in this preliminary study and these are briefly described in section 3. Section 4 describes the discriminant analysis methods and section 5 includes the variables that we believe influence discrimination between the different methods. In section 6 we consider the criteria for evaluation of the discriminant functions. Then in section 7 we present the results and finally the conclusions in section 8.

2. DATA

All 203 quarterly series from the 'M-competition' (see Makridakis *et al.* (1982)) are used in this study. As was stated in the above paper, this set is not random in a strictly statistical sense, but it does cover a wide spectrum of possibilities. The series are obtained from different sources, with different starting/ending dates, covering micro, macro, industry and demographic data. The average number of observations per series is around 41, with a minimum of 10 and a maximum of 106 observations. An additional 8 observations at the end of the series are used for evaluating the forecasting performance of the three forecasting methods. The series were appropriately scaled so that the mean value of each of them was between 1 and 10.

3. FORECASTING METHODS

Three forecasting methods are considered in this study. Each of the methods is applied to all 203 series. Only the first part of the series is used to estimate the parameters of the method and only the last 8 observations are used to calculate the mean squared error of the one-step-ahead forecasts. It should be noted that a one-step-ahead forecast made in period T uses all the observations up to and including period T. The three methods are briefly described below.

3.1 Simple exponential smoothing

Simple exponential smoothing is quite popular in large inventory management applications, where short term forecasts of thousands of items are required on a routine basis. The method can be automated on a computer and requires little human interference. Adopting the notation from Gardner (1985), the method can be embodied in the following equations:

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1} \tag{1a}$$

$$\hat{X}_t(1) = S_t \tag{1b}$$

where S_t is the smoothed level of the series and X_t is the observed value of the series in period t. $\hat{X}_t(1)$ is the one-step-ahead forecast for period t + 1 made in period t and α is the smoothing parameter.

The parameters were selected using a two-stage grid search algorithm. The first stage involved a coarse grid of size .1 with $0 < \alpha < 2$. During the second stage, a finer grid of size .01 around the optimum identified in the first stage is used. The wider range for α was chosen because simple exponential smoothing is equivalent to a difference equation which is stable in the range $0 < \alpha < 2$ (see Brenner

et al. (1968)). An alternative justification is that simple exponential smoothing is optimal for the ARIMA (0, 1, 1) process which is invertible for α in the range 0 to 2 (see Muth (1960)). The value of α which gave the minimum mean squared error for the one-step-ahead forecasts within the sample period was considered to be its 'optimal' value. Initial value was obtained by backcasting. Since backcasting itself requires a starting value, the average of the last 4 observations was taken as the starting value for the level.

3.2 Holt-Winters Method

This method generalizes simple exponential smoothing to cope with trend and seasonality in the data. Holt (1957) developed an algorithm that allowed for a local adjustment of the trend in a time series. This was extended by Winters (1960) to include an evolving seasonal pattern. A multiplicative seasonal factor is assumed in this study. The equations describing the algorithm are:

$$S_t = \alpha X_t / I_{t-p} + (1 - \alpha)(S_{t-1} + T_{t-1}) \tag{2a}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \tag{2b}$$

$$I_t = \delta \frac{X_t}{S_t} + (1 - \delta)I_{t-p} \tag{2c}$$

$$\hat{X}_t(1) = (S_t + T_t)I_{t-p+1} \tag{2d}$$

where T_t is the smoothed trend and I_t is the smoothed seasonal index. γ and δ are the smoothing parameters for the trend and seasonal index respectively. Once again a two stage grid search was employed to find the optimal values of the smoothing parameters. The range of values considered for α and γ was between 0 and Z and that for δ between 0 and 1. Backcasting was once again used to obtain the starting values.

3.3 Structural time series models

We have considered the basic structural model as described in Harvey and Peters (1990). It is defined by

$$X_t = \mu_t + \gamma_t + \varepsilon_t , \qquad t = 1, \ldots, T \tag{3a}$$

where μ_t, γ_t and ε_t are the trend, seasonal and irregular components, respectively. We assume a local approximation to the linear trend such that

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad t = 1, \ldots, T \tag{3b}$$

$$\beta_t = \beta_{t-1} + \xi_t, \qquad t = 1, \ldots, T \tag{3c}$$

where η_t and ξ_t are distributed normally and independently of each other and over time, with mean zero and variances σ^2_η and σ^2_ξ, respectively. The process generating the seasonal component is

$$\gamma_t = -\sum_{j=1}^{p-1} \gamma_{t-j} + \omega_t, \quad t = 1, \ldots, T \tag{3d}$$

where $\omega_t \sim NID(0, \sigma^2_\omega)$. The irregular component is also assumed to be Gaussian white noise with mean zero and variance σ^2_ε .

The state space form of the model consists of a transition equation

$$\underline{\alpha}_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} = \left[\begin{array}{cc:ccc} 1 & 1 & & 0 & \\ 0 & 1 & & & \\ \hdashline & & -1 & -1 & -1 \\ & 0 & 1 & 0 & 0 \\ & & 0 & 1 & 0 \end{array}\right] \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \xi_t \\ \omega_t \\ 0 \\ 0 \end{bmatrix} \tag{3e}$$

and a measurement equation

$$x_t = [1 \quad 0 \quad 1 \quad 0 \quad 0]\, \underline{\alpha}_t + \varepsilon_t \quad , \quad t = 1, \ldots, T \qquad (3f)$$

The covariance of the disturbance term in equation (3e) is $\underline{Q}$ where $\underline{Q} = \text{diag}\ \{\sigma^2_\eta, \sigma^2_\xi, \sigma^2_\omega, 0, 0\}$. Let $\underline{a}_t$ be the minimum mean square estimator (MMSE) of $\underline{\alpha}_{t-1}$ at time $t-1$ and $\underline{P}_{t-1}$ the covariance matrix of the estimation error $\underline{a}_{t-1} - \underline{\alpha}_{t-1}$.

Once a new observation, x_t, is available, both $\underline{a}_{t-1}$ and $\underline{P}_{t-1}$ can be updated using the Kalman filter. Under the assumption that $\underline{\alpha}_0 \sim N(\underline{a}_0, \underline{P}_0)$ and known values of $\underline{a}_0$ and $\underline{P}_0$, one can now obtain the T one-step-ahead prediction errors and their variances. Prediction of future observations can be made using the Kalman filter. The maximum likelihood estimates of the hyperparameters, σ^2_ε, σ^2_η, σ^2_ξ and σ^2_ω, in the time domain are used in the Kalman filter. A diffuse prior for $\underline{P}_0$ and $\underline{a}_0 = \underline{0}$ are used to initialise the process. Details of the operation of the Kalman filter, prediction error decomposition and the maximum likelihood estimation procedure can be found in Harvey (1989) and Harvey and Peters (1990). Alternative methods of computing maximum likelihood estimators and using various starting values are discussed in the above two references.

The STAMP package was used to estimate the paramaters and produce forecasts for the basic structural model. The default settings in the package were generally used. The only exception was that when estimating the hyperparameters in the Kalman filter, the time domain method was used for 90% of the series. For the 10% of the series the time domain estimation seemed to crash the program and hence for these series the method of scoring was employed.

4. DISCRIMINANT ANALYSIS

An analysis which leads to the best way in which two or more populations may be distinguished is known as discriminant analysis. A function, which is some combination of observed variables, that maximizes the separation between the populations is a discriminant function. The discriminant function may be used to describe and interpret the differences between the populations and may also be used to predict to which population an unknown observation is most likely to belong. Applications of discriminant analysis can be found in a wide variety of fields, from auditing to medicine (see Goudie (1987), Titterington *et al.* (1981), Stoodley *et al.* (1982)). The pioneering work in this area was done by Fisher (1936) and Mahalanobis (1936).

In the context of this study we have three populations, simple exponential smoothing (SE), Holt-Winters (HWM) and structural time series model (STAMP). A series i belongs to population j if the series' post sample mean squared error of the one-step-ahead forecasts is least when method j was used to forecast. An observation $\underline{x}$ contains information about the characteristics of a particular series. We have defined a series in terms of 20 variables so that $\underline{x}$ is a 20 dimensional vector. The variables are defined in section 5.

Classical discriminant analysis assumes that the M populations have multivariate normal distribution. Furthermore, if the M populations can be assumed to have equal covariance matrices then a much simpler linear discriminant procedure can be used, otherwise quadratic discriminant analysis should be used. The linear discriminant function has the advantage that, since the function is linear, the coefficients are easy to interpret. A multivariate generalisation of Bartlett's test for homogeneity of M population variances can be used to help make a

decision as to whether linear or quadratic discriminant analysis is appropriate.

Although non-parametric methods of discriminant analysis exist (see Silverman (1986)), they have not been considered due to the non-availability of reliable computer software.

4.1 Linear discriminant function

We use the generalised squared distance to separate the M populations. The generalised squared distance of a p-variate observation, $\underline{x}$, from population π_g is defined as

$$D_L(\underline{x}, \pi_g) = (\underline{x} - \underline{x}_g)'\underline{s}^{-1}(\underline{x} - \underline{x}_g) \tag{4}$$

where $\underline{x}_g$ is a vector containing the sample means of the p variables in population π_g and $\underline{s}$ is the sample pooled covariance matrix of a p-variate population. The elements of $\underline{s}$ are given by

$$s_{ij} = \sum_{g=1}^{M} \sum_{k=1}^{n_g} \frac{(x_{ikg} - \bar{x}_{ig})(x_{jkg} - \bar{x}_{jg})}{\sum_{i=1}^{M} n_i - M} \tag{5}$$

where $\bar{x}_{ig}$ is the sample mean of the ith variable of population π_g, n_i is the number of observations in the sample from population π_i and x_{ikg} represents the ith variate of observation k from population g. Hence, an unclassified observation $\underline{x}_0$ is classified as belonging to population g if

$$D_L(\underline{x}_0, \pi_g) = \min_j \{D_L(\underline{x}_0, \pi_j)\} \tag{6}$$

This rule has the property of minimizing the probability of misclassification. The centroids of each of the populations and the sample pooled covariance matrix, s, are obtained from a training sample which consists of observations whose population classification is known.

4.2 Quadratic Discriminant function

In case the M populations cannot be assumed to have equal covariance matrices, the definition of the generalised squared distance has to be modified to

$$D_Q(\underline{x}, \pi_g) = (\underline{x} - \underline{x}_g)'\underline{s}_g^{-1}(\underline{x} - \underline{x}_g) + \log_e|\underline{s}_g| \tag{7}$$

where $\underline{s}_g$ is the sample covariance matrix of population g. The elements of $\underline{s}_g$ are given by

$$s_{ij} = \sum_{k=1}^{n_g} (x_{ikg} - \bar{x}_{ig})(x_{jkg} - \bar{x}_{jg})/(n_g - 1). \tag{8}$$

The classification rule for an unclassified observation $\underline{x}_0$ is similar to (6) and once again it has the property of minimizing the probability of misclassification.

4.3 Probability of population membership

In both the linear and quadratic classification rules, the property of minimum probability of misclassification holds only if the prior probabilities of population membership are equal, otherwise a modification to D_L and D_Q must be made. Suppose the prior probability of an observation $\underline{x}$ belonging to population π_g is p_g, then the modified distance measure for linear discrimination is

$$D_L'(\underline{x}, \pi_g) = D_L(\underline{x}, \pi_g) - 2 \log_e p_g \tag{9}$$

and for quadratic discrimination it is

$$D_Q'(\underline{x}, \pi_g) = D_Q(\underline{x}, \pi_g) - 2 \log_e p_g. \tag{10}$$

Now if $p(\pi_g|\underline{x})$ is the probability of an observation belonging to population π_g given its p-variate vector is $\underline{x}$, then

$$p(\pi_g|\underline{x}) = \frac{p_g p(\underline{x}|\pi_g)}{\sum_{j=1}^{M} p_j p(\underline{x}|\pi_j)} \tag{11}$$

is easily obtained using Bayes' theorem. The assumption of multivariate normality allows us to simplify (11) to

$$p(\pi_g|\underline{x}) = \frac{\exp(-D_L'(\underline{x}, \pi_g)/2)}{\sum_{j=1}^{M} \exp(- D_L'(\underline{x}, \pi_j)/2)} . \tag{12}$$

D_L' is replaced by D_Q' in (12) for quadratic discrimination.

5. FEATURE VARIABLES

In order to use discriminant analysis to predict the most likely forecasting method that will give the 'best' post sample forecasts of a given time series, some structural characteristics of the time series that may help in distinguishing between the forecasting methods need to be defined and measured. We conjecture that all or some of the following feature variables of the time series may help in this discrimination:

1. N_OBS - the number of observations in the sample period.

2. N_TPS - Non-parametric turning point test, $\frac{u-2(n-2)/3}{\sqrt{\frac{16n-29}{90}}}$, where u is the number of turning points.

3. N_STEPS - ratio of number of step changes to the number of observations. A step change occurs at time t if $|x_t - \bar{x}_{t-1}| > 2\, s_{t-1}$ where $\bar{x}_{t-1}$ and s_{t-1} are the mean and standard deviation of $x_1 \ldots x_{t-1}$.

4. SKEWNS - coefficient of skewness, $\dfrac{\Sigma(x_i - \bar{x})^3/n}{[\Sigma(x_i - \bar{x})^2/n]^{3/2}}$

5. KURTOSIS - kurtosis, $\dfrac{\Sigma(x_i - \bar{x})^4/n}{[\Sigma(x_i - \bar{x})^2/n]^2} - 3$

6. CV - coefficient of variation.

7. T_STAT1 - the value of the t-statistic for the test that coefficient of time, t, is zero when data for the series are fitted to a cubic polynomial.

8. T_STAT2 - as in 7 for the coefficient of t^2.

9. T_STAT3 - as in 7 for the coefficient of t^3.

10. AC1 - autocorrelation at lag 1 of the residuals after a third degree polynomial in time has been fitted to the series.

11. AC2 - as in 10 for lag 2.

12. AC3 - as in 10 for lag 3.

13. AC4 - as in 10 for lag 4.

14. PAC2 - as in 10 but now it is the partial autocorrelation at lag 2.

15. PAC3 - as in 14 for lag 3.

16. PAC4 - as in 14 for lag 4.

17. R1 - ratio of the variance of the first differenced series to the variance of the series.

18. R2 - ratio of the variance of the second differenced series to the variance of the first differenced series.

19. RS1 - as in 17 but the series and the first differenced series are first seasonally differenced.

20. RS2 - as in 18 but the first and second differenced series are first seasonally differenced.

6. EVALUATION CRITERION

Two error rates, the apparent and actual, are calculated. The apparent error rate is the one obtained by resubstituting the training sample and determining the misclassification. This rate is typically over optimistic and can badly mislead the user if the sample size is not much larger than the number of variables in the discriminant rule. The actual error rate is obtained when the rule under consideration is used to classify all possible future samples.

Out of 203 series, 50 were reserved in a test set to evaluate the actual error rate. They were selected on a random basis in such a way that the proportion from each population was the same for the training and test sets.

7. THE RESULTS

In order to calculate the post-sample mean squared error of the one-step-ahead forecasts, Fortran programs were written for the simple exponential smoothing and Holt-Winters methods. The STAMP package, developed by Harvey and Peters (1989), was used for structural time series model. Table 1 shows the number of series in the 3 populations for the training sample and the test sample.

Table 1

	SE	HWM	STAMP	TOTAL
TRAINING	51	37	65	153
TEST	16	12	22	50
TOTAL	67	49	87	203

Another Fortran program incorporating IMSL subroutines was developed for obtaining the feature variables of each series. Using the DISCRM procedure in the SAS computer package, the test for homogeneity of the covariance matrices was found to be significant at the 10% level. Hence the quadratic discriminant function was adopted as the discriminator. Proportions of each population in the total sample were used as prior probabilities.

Table 2 shows the classification of the training sample using a quadratic discriminant function. The figures in brackets are percentages rounded to the nearest unit. For example, of the 51 series in the population SE, 48 or 94% were classified into SE, 2 or 4% were classified into HWM and 1 or 2% was classified into STAMP. The total number of series from population SES which were misclassified was about 6%.

Table 2

FROM METHOD	SE	HWM	STAMP	TOTAL
SE	48(94)	2(4)	1(2)	51(100)
HWM	3(8)	34(92)	0(0)	37(100)
STAMP	6(9)	7(11)	52(80)	65(100)
TOTAL	57(34)	43(28)	53(35)	153(100)
PRIORS	.33	.24	.42	
ERROR RATE	.06	.08	.2	.12

A similar classification for the test set is shown in Table 3.

Table 3

FROM METHOD	SE	HWM	STAMP	TOTAL
SE	9(56)	3(19)	4(25)	16(100)
HWM	4(33)	1(8)	7(58)	12(100)
STAMP	8(36)	2(9)	12(55)	21(100)
TOTAL	21(42)	6(12)	23(48)	50(100)
PRIORS	.33	.24	.42	
ERROR RATE	.44	.92	.45	.56

8. CONCLUSION

The apparent error rate of only about 12% is encouraging. As for the actual error rate that is obtained using the test set, the error rate of 56% seems, on first reflection, disappointing. But, if one had a series which belonged to the population SE, then its chances of being classified correctly using just a random allocation rule based on population proportions are .33, while using the discriminant rule its chances are .56. Similarly for a series from HWM, the corresponding chances are .24 and .08, and for a series from STAMP they are .42 and .55. Except for HWM, the discriminant rule does a better job at classification of a series into its correct grouping than a random allocation according to population proportions.

One possible reason for the test set having an overall success rate of 44%, may be due to possible violation of the assumption of multivariate normality of each of the populations. A more careful transformation of those variables which fail a normality test may yield better results. The Box-Cox transform could be a useful tool here.

Another problem may be due to the choice of variables used to characterise a series. Stepwise selection procedures, similar to those employed in regression, may be employed to reduce the set of variables to that set which separates the populations most and does it most 'efficiently'. Alternative variables to characterise a series may yield different results too.

This is a preliminary report and further work is continuing.

REFERENCES

Brenner, J.L. *et al.* (1968) 'Difference equations in forecasting formulas', *Management Science*, 14, 141-159.

Chatfield, C. (1988) 'The future of time-series forecasting', *International Journal of Forecasting*, 4, 411-419.

Chatfield, C. and Yar, M. (1988) 'Holt-Winters forecasting: some practical issues', *The Statistician*, 37, 129-140.

Fisher, R.A. (1936) 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, 7, 179-184.

Gardner, E.S. (1985) 'Exponential smoothing: The state of the art,' *Journal of Forecasting*, Vol. 4, 1-28.

Goudie, A.W. (1987) 'Forecasting corporate failure: The use of discriminant analysis within a disaggregated model of the corporate sector', *Journal of the Royal Statistical Society (A)*, 150, 1, 69-81.

Harvey, A.C. (1989) 'Forecasting structural time-series models and the Kalman filter', Cambridge University Press, Cambridge.

Harvey A.C. and Peters, S. (1990) 'Estimation procedures for structural time-series models,' *Journal of Forecasting*, Vol. 9, 2, 89-108.

Holt, C.C. (1957) 'Forecasting seasonal and trends by exponentially weighted moving averages', Office of Naval Research, Memorandum No. 52.

Mahalanobis, P.C. (1936) 'On the generalised distance in statistics', *Proc. Nat. Inst. Sci. India*, 2, 49-55.

Makridakis, S.G. and Hibon, M. (1979) 'Accuracy of forecasting: an empirical investigation (with discussion)' *Journal of Royal Statistical Society (A)*, 142, 97-145.

Makridakis, S.G., Anderson, A., Carbone, R., Fidles, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R., (1982) 'The accuracy of extrapolation (time-series) methods: Results of a forecasting competition', *Journal of Forecasting*, Vol. 1, No.3, 111-153.

Makridakis, S.G., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1984) 'The forecasting accuracy of major time-series methods', Wiley, New York.

Muth, J. (1960) 'Optimal properties of exponentially weighted forecasts', *Journal of the American Statistical Association*, 55, 299-306.

Newbold, P. and Granger, C.W.J. (1974) 'Experience with forecasting univariate time-series and the combination of forecasts (with discussion)', *Journal of the Royal Statistical Society (A)*, 137, 131-165.

Reid, D.J. (1975) 'A review of short term projection techniques' in 'Practical aspects of forecasting' (H.A. Gordon, ed.) London: *Operational Research Society*, 8-25.

Silverman, B.W. (1986) 'Density estimation for statistics and data analysis', Chapman and Hall, London.

Stoodley, K.D.C., Naghdy, F. and Crew, A.D. (1982) 'Monitoring of cardiovascular variables in a post-surgical cardiac intensive care unit', In Time-Series analysis: theory and practice 1, ed. O.D. Anderson, North-Holland Publishing Company, Amsterdam.

SAS Institute, Inc. (1985). SAS user's guide: Statistics, Version 5 edition. SAS Institute, Inc., Cary, N.C.

Titterington, D.M., Murray, G.D., Murray, L.S., Spiegelhalter, D.J., Skene, A.M., Habbema, J. and Gelpke, G. (1981) 'Comparison of discrimination techniques applied to a complex data set of head injured patients', *Journal of the Royal Statistical Society (A)*, 144, 2, 145-174.

Winters, P.R. (1960) 'Forecasting sales by exponentially weighted moving averages', *Management Science*, 6, 324-342.