MONASH

WP 5/88

ISSN 0729-0683

ISBN 0 86746 687 1

ANCORA IMPARO

MONASH UNIVERSITY

STATISTICAL FOUNDATIONS OF EXPONENTIAL SMOOTHING

Ralph D. Snyder

Working Paper No. 5/88

February 1988

DEPARTMENT OF ECONOMETRICS

ISSN 0729-0683

ISBN 0 86746 687 1

STATISTICAL FOUNDATIONS OF EXPONENTIAL SMOOTHING

Ralph D. Snyder

Working Paper No. 5/88

February, 1988

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS AND POLITICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

Statistical Foundations of Exponential Smoothing

Ralph D. Snyder

Department of Econometrics

Monash University

February, 1988

Abstract:

In this paper the exponential smoothing methods of forecasting are rationalized in terms of a statistical state space model with only one primary source of randomness. Their link, *in general terms*, with the ARMA class of models (both stationary and nonstationary cases) is also explored.

Keywords: Exponential smoothing, Forecasting, Estimation, Kalman filtering, ARMA models, ARIMA models.

1. INTRODUCTION

When first developed, the exponential smoothing methods of forecasting (Brown, 1959; Holt, 1957; Winters, 1960) were presented as heuristics. Their early dissemination concided with the computer revolution in business where the emphasis on recursion to reduce computational loads and on-line data requirements provoked considerable interest. In most applications these techniques proved to be satisfactory vis-a-vis the alternatives, a finding supported in more recent times by the forecasting competition organized by Makridakis (1981). Moreover, their relative simplicity meant that they could be introduced into business organizations and accepted by staff with little training in the formal methods of statistics.

The statistical foundations of the techniques proved to be more elusive but one can discern in Gardner's (1985) comprehensive survey of exponential smoothing the evolution of two distinct explanations. One had its origins in the work of Box and Jenkins (1976) with special cases of their integrated autoregressive moving average processes. The other was tied to special state space models (Muth, 1960; Theil and Wage, 1964, Nerlove and Wage, 1964; Harrison and Stevens, 1976; Harvey, 1984) where exponential smoothing is optimal in large samples.

Curiously, these statistical frameworks are not equivalent, as exemplified by the simplest case of the exponentially weighted average. In the Box-Jenkin's approach this type of average is rationalized in terms of an ARIMA(0,1,1) model where stability is achieved while the smoothing parameter α lies in the range $0 \le \alpha \le 2$. However, the Kalman filter associated with the traditional state space explanation only converges to the error correction equations for the exponentially weighted average in large samples with an α in the smaller range

$0 \leq \alpha \leq 1$. Gardner (1985) indicates that the latter range has dominated most research on exponentially weighted averages but suggests that there is little evidence to support it. Brenner (1968), who recognized the possibility of the larger range, still rejected values above 1 on the grounds that randomness in the data is then amplified. His argument, however, has little credence when the data are highly correlated as occurs in economic time series with a pronounced business cycle effect. Then a value for α in excess of 1 is needed to ensure that the forecasts do not lag behind the data. It may be counter argued that these medium term cycles should be modelled explicitly rather than employing a large α. But in the context of short term business forecasting the available data is usually too short to permit reliable estimates of such cycles so that there is little choice in the matter. This illustrates that the ARIMA framework provides greater flexibility which, under appropriate conditions, translates into better predictions than those from the traditional state space approach.

Needless to say, the relationship between the various forms of exponential smoothing and their ARIMA foundations is rather obstruse and this has undesirable pedagogic consequences. In this paper it is shown that most of these methods can be rationalized in terms of an alternative form of the state space framework reminiscent of the error correction formulae of exponential smoothing but which is equivalent to the entire class of ARMA models in its most general sense (ie. incorporating the nonstationary ARIMA models as a special case). As will be seen the relationship with exponential smoothing is then clearer and it enables users to develop models in more meaningful terms with concepts such as mean level, mean growth and seasonal indexes rather than the more obscure difference equations of the ARMA framework.

2. THE STATE SPACE FRAMEWORK

The state space framework is based on the idea that all the past information contained in a time series $y(t)$, $y(t-1)$, ... can be condensed into the so called state vector $\beta(t)$ with small dimension r and that this, in turn, can be utilized to provide information about future values of a series. Its most general time invariant form can be written as:

$$y(t) = x' \beta(t-1) + \varepsilon(t) \tag{2.1a}$$

$$\beta(t) = T \beta(t-1) + \eta(t) \tag{2.1b}$$

where x is a fixed r-vector, T is a fixed, square matrix, the $\varepsilon(t)$ are independent $N(0,\sigma^2)$ random variables, the $\eta(t)$ are independent $N(0,\sigma^2 Q)$ random r-vectors, and the $\varepsilon(t)$ and $\eta(t)$ are contemporaneously correlated with covariance vector $\sigma^2 q$. It is further assumed that the initial state vector $\beta(0)$ is an $N(0,\sigma^2 C)$ random r-vector, and that $\eta(t)$ is independent of $\beta(t-1)$. The conventional state space explanations of exponential smoothing have traditionally been based on the special case where $q = 0$ and C is diagonal to reduce the parameters to a manageable level eg. see Harvey (1985). In this paper we consider another possibility where it is assumed that $\eta(t) = \alpha \varepsilon(t)$ where α is a fixed r-vector. The framework then only relies on one primary source of randomness (ie. the $\varepsilon(t)$) and takes the inherently simpler form:

$$y(t) = x' \beta(t-1) + \varepsilon(t) \tag{2.2a}$$

$$\beta(t) = T \beta(t-1) + \alpha \varepsilon(t). \tag{2.2b}$$

The α-vector determines the extent to which there is structural change (ie. non-deterministic change) in the model's coefficients and, as will be seen later, its elements play the same role as the *smoothing parameters* from exponential smoothing.

Because of its invariant nature (ie. x, T and α are independent of time), this version of the state space framework has generally been reserved for modelling stationary time series (Koehler and Murphree,

1988) with a particular emphasis on stationary ARMA models. However, invariance does not imply stationarity and recent work (Harvey and Pierse, 1984; Kohn and Ansley, 1986) utilizing a slightly different form of the above framework without the $\varepsilon(t)$ term in the measurement equation (2.2a) suggests that it can also be used to represent nonstationary ARIMA processes. Surprisingly, the likeness of (2.2b) to the error correction forms of exponential smoothing has largely gone unnoticed. It is this aspect of (2.2) which is explored in this paper.

The close link between the ARMA models and a state space framework with only one primary source of randomness was originally established by Akaike (1974). His results can be adapted to the version of the state space framework of this paper also without recourse to his stationarity assumption. More specifically, it can be established, as shown in the appendix, that any model conforming to (2.2) can be converted to the ARMA form

$$\phi(B)\, y(t) = \theta(B)\, \varepsilon(t) \qquad (2.3a)$$

where B is the usual backward shift operator and

$$\phi(B) = 1 + \phi_1 B + \ldots + \phi_p B^p \qquad (2.3b)$$

$$\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q. \qquad (2.4c)$$

In the process, it is demonstrated that the parameters of both frameworks are related by the formulae:

$$\phi_j = a_j \qquad j = 1, \ldots r \qquad (2.4a)$$

$$\theta_j = x' \sum_{i=0}^{j-1} a_i T^{j-i-1} \alpha + a_j \qquad j = 1, \ldots r \qquad (2.4b)$$

where the a_i are the coefficients of the characteristic equation

$$a_0 \lambda^r + a_1 \lambda^{r-1} + \ldots + a_r \lambda^0 = 0 \qquad (2.4c)$$

of the transition matrix T. Pearlman (1980) has established the converse result. Taken together, they mean that both frameworks are alternative but equivalent ways of representing the same processes.

Although the number of autoregressive and moving average parameters are both equal to r in (2.4), the result still effectively refers to the general ARMA(p,q) framework where p and q need not be equal. To illustrate, when it is a singular matrix, T has at least one characteristic root equal to zero. In these circumstances $\phi_r = 0$ without necessarily having $\theta_r = 0$.

This result is very important in the context of exponential smoothing. As stated earlier, it will be shown that most of the exponential smoothing methods can be rationalized in terms of the framework (2.2). The formulae (2.4) therefore provide the effective link between exponential smoothing and the general ARMA framework and thus unify the work of Muth (1960), Harrison (1967), Nerlove and Wage (1964), Theil and Wage (1964), Cogger (1974), Roberts (1982), Godolphin and Harrison (1975), Goodman (1974), Ledolter and Box (1978), and McKenzie (1974, 1976) which all consider special cases of this relationship. It also confirms Cogger's (1985) assertion in his comments on Gardner's(1985) paper "that every ARIMA model can probably be described in exponential smoothing terminology".

3. EXPONENTIAL SMOOTHING

Given its diverse nature (eg. see Gardner(1985)) it is difficult to give a completely comprehensive definition for exponential smoothing. However, the following criteria encompass its linear forms.

Criterion 1

A forecast of the series in period t, denoted by $\tilde{y}(t)$, is produced from estimates b(t-1) of the state vector with the linear relationship

$$\tilde{y}(t) = x' b(t-1) \qquad (3.1)$$

where x is a fixed r-vector.

Criterion 2

The estimates of the state vector are revised, explicitly or implicitly, by a linear error correction formula

$$b(t) = T\ b(t-1) + \alpha\ e(t) \tag{3.2}$$

where T and α are given, and e(t) is the one step ahead forecast error

$$e(t) = y(t) - x'\ b(t-1) \tag{3.3}$$

Criterion 3

For given T, x, and α the associated *discount matrix* D, defined by

$$D = T - \alpha\ x' \tag{3.4}$$

must be stable, ie it must possess characteristic roots all lying within the unit circle in the complex plane.

The criterion 2 is based on the observation of Harrison (1967) and Gardner (1985) that most of the exponential smoothing methods can be converted to equivalent error correction forms involving only the one step ahead forecast error. Criterion 3 is required to ensure stable forecasts. More specifically, substitute (3.3) into (3.2) to give the recurrence relationship

$$b(t) = D\ b(t-1) + \alpha\ y(t) \tag{3.5}$$

with the closed form solution

$$b(t) = D^t\ b(0) + \sum_{j=0}^{t-1} D^j\ \alpha\ y(t-j). \tag{3.6}$$

The solution is dependent on the seed estimate b(0) and the observations y(1) ... y(t). Criterion 3 ensures that D^t tends to a null matrix as t increases, so that:

(a) the influence of the seed estimate b(0), which may be quite awry given that it is normally made without access to any data, disappears in large samples;

(b) older observations eventually have less weight than more recent ones in the forecasting process.

Interestingly, (3.6) is reminescent of the formula for an exponentially weighted average where the usual scalar discount factor is replaced by the discount matrix D and where similar conclusions apply.

Exponential smoothing, as presented here, relies on the error correction formulae (3.2) which has a remarkable likeness to (2.2b) in the state space framework. Assuming that the latter represents the generating process of the data, it is pertinent to examine the performance of the b(t) as estimates of the state vectors $\beta(t)$. It is readily established that the estimation errors satisfy the recursion

$$\beta(t) - b(t) = D\ (\ \beta(t-1) - b(t-1)\ ) \tag{3.7}$$

with the closed form solution

$$\beta(t) - b(t) = D^{t}\ (\ \beta(0) - b(0)\ ). \tag{3.8}$$

Therefore the estimation errors are dependent on the initial estimation error but they converge to zero if D is a stable matrix. Thus, according to any well defined error criterion, exponential smoothing is optimal for the state space framework in large samples provided that D is stable.

4. OPTIMAL ESTIMATION IN SMALL SAMPLES

In small samples exponential smoothing is not an optimal estimation procedure except in the special and rather unlikely case where b(0) is a perfectly accurate estimate of $\beta(0)$. It can be established that the minimum variance estimates of $\beta(t)$, denoted by $\hat{\beta}(t)$, can be computed with the version of the Kalman filter in Snyder (1985). Here the estimates are also obtained with an error corrrection formula where the smoothing vector α is replaced by a time dependent r-vector a(t) called the gain ie.

$$\hat{\beta}(t) = T\ \hat{\beta}(t-1) + a(t)\ (\ y(t) - x'\ \hat{\beta}(t-1)\ ). \tag{4.1}$$

The distinctive feature of the Kalman filter is that the gain vector is selected at each stage so as to minimize the mean squared one step ahead forecast error for given α and so the estimates obtained this way consequently differ from those of exponential smoothing. However, it can be established that if D is stable then a(t) tends to α as t increases, which means that the Kalman filter collapses to the simpler form of exponential smoothing in large samples.

5. SPECIAL CASES

The following cases illustrate but do not exhaust some of the applications of the proposed state space framework. In each example it is assumed that the reader can discern immediately the exponential smoothing error correction equations without their being presented explicitly.

5.1 Exponentially Weighted Averages

The exponentially weighted average can be rationalized in terms of a model where it is only structural change which induces change in the mean of the series ie.

$$y(t) = \mu(t-1) + \varepsilon(t) \qquad (5.1a)$$

$$\mu(t) = \mu(t-1) + \alpha\, \varepsilon(t) \qquad (5.1b)$$

where $\mu(t)$ is the current mean. Here x, T and α are scalars with x = T = 1. Furthermore, the characteristic equation of T is $\lambda - 1 = 0$ so that $a_1 = -1$. An application of (2.4) indicates that this simple state space model is equivalent to the ARIMA model

$$(1 - B)\, y(t) = (1 + \theta B)\, \varepsilon(t) \qquad (5.1c)$$

where $\theta = \alpha - 1$. The latter, or at least its closed form equivalent, was originally proposed by Muth (1960) as a possible statistical framework for the exponentially weighted average. Note, however, that (5.1a) and (5.1b) is a simpler and more direct representation for pedagogic purposes.

5.2 Trend Corrected Exponential Smoothing

Trend corrected exponential smoothing is based on the notion of a local trend line which adapts over time to structural change. Letting $\delta(t)$ represent the current rate of growth, the model takes the form:

$$y(t) = \mu(t-1) + \varepsilon(t) \quad (5.2a)$$

$$\mu(t) = \mu(t-1) + \beta(t-1) + \alpha_1 \varepsilon(t) \quad (5.2b)$$

$$\beta(t) = \beta(t-1) + \alpha_2 \varepsilon(t) \quad (5.2c)$$

Hence

$$\beta(t) = \begin{bmatrix} \mu(t) \\ \delta(t) \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

The transition matrix is upper triangular with characteristic roots equal to its diagonal elements of 1. The characteristic equation is $(\lambda - 1)^2 = (\lambda^2 - 2\lambda + 1) = 0$. Hence, (2.4) implies that the above model is equivalent to the ARIMA model

$$(1 - B)^2 y(t) = (1 + \theta_1 B + \theta_2 B^2) \varepsilon(t) \quad (5.2d)$$

where $\theta_1 = \alpha_1 - 2$ and $\theta_2 = 1 - \alpha_1 + \alpha_2$. This ARIMA model was originally proposed by Harrison (1967) as the rationale for an error correction form, giving similar results to Holt's (1957) trend corrected exponential smoothing.

5.3 Damped Trend Corrected Exponential Smoothing

A variation of trend corrected exponential smoothing entails dampening the mean growth rate by a factor ϕ in the range $0 < \phi < 1$. The model in the previous section is amended to

$$y(t) = \mu(t-1) + \varepsilon(t) \quad (5.3a)$$

$$\mu(t) = \mu(t-1) + \delta(t-1) + \alpha_1 \varepsilon(t) \quad (5.3b)$$

$$\delta(t) = \phi\, \delta(t-1) + \alpha_2 \varepsilon(t) \quad (5.3c)$$

The x, $\beta(t)$ and α vectors are the same as in section 5.2 . The transition matrix is given by

$$T = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}$$

The characteristic roots of T are 1 and ϕ so that its characteristic equation is $(\lambda - 1)(\lambda - \phi) = \lambda^2 - (1 + \phi)\lambda + \phi = 0$. The application of (2.4) indicates an equivalence with the ARIMA model

$$(1 - B) (1 - \phi B) y(t) = (1 + \theta_1 B + \theta_2 B^2) \varepsilon(t) \tag{5.3d}$$

where $\theta_1 = \alpha_1 - \phi - 1$ and $\theta_2 = \alpha_2 + \phi - \phi\,\alpha_1$. The model is similar, but not identical to, a proposal by Roberts (1982) to dampen the growth rate and improve the forecasts given by trend corrected exponential smoothing. Again the state space model has a more direct link with its exponential smoothing analogue than the ARIMA model (5.3d).

5.4 Seasonal Models

The method outlined in Holt (1957) and Winters (1960) for seasonal data posits a separate coefficient for each season which is untenable from a computational and statistical point of view in applications involving weekly data. Although the additive forms of such models can be accomodated by the framework (2.2) we adopt the more satisfactory strategy of Brown (1963) and Harrison and Stevens (1965) where trigonometric functions are used to model the seasonal component. Although models with mixtures of linear trends and seasonal components are easily developed we shall expedite matters by considering only the simplest possible case involving the harmonic terms cos(ωt) and sin(ωt) where $\omega = 2\pi/m$ with m representing the number of periods per cycle. In this context the model involves the following components:

$$\beta(t) = \begin{bmatrix} \beta_1(t) \\ \beta_2(t) \end{bmatrix} \quad x = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad T = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_1 \end{bmatrix}$$

where both elements of α are the same. Note that $T^m = I$ so that the characteristic roots of T all lie on the unit circle in the complex plane and the characteristic equation is $(\lambda^m - 1) = 0$. The equivalent ARIMA model is therefore

$$(1 - B^m) y(t) = (1 + \sum_{i=1}^{m} \theta_i B^i) \varepsilon(t)$$

where

$$\theta_i = \alpha_1 (\cos(i-1)\omega - \sin(i-1)\omega) \qquad i = 1, \ldots , m-1$$

$$\theta_m = \alpha_1 (\cos(m-1)\omega - \sin(m-1)\omega) - 1$$

6. INITIAL CONDITIONS

Before using the Kalman filter approach it is necessary to specify the the prior distribution $N(\beta,C)$ for the model under consideration. To this end, we make the assumption that data has only been collected from period 1 onwards, but that the process has operated for m periods prior to this. Backsolving (2.2a) gives

$$\beta(0) = T^m \beta(-m) + \sum_{i=0}^{m-1} T^i \alpha \varepsilon(-i) \tag{6.1}$$

which in turn has a mean and variance given by

$$\beta = T^m E(\beta(-m)) \tag{6.2a}$$

$$C = T^m VAR(\beta(-m)) T'^m + \sigma^2 \sum_{i=0}^{m-1} T^i \alpha \alpha' T'^i \tag{6.2b}$$

Often m is unknown so that it is conventional to make it arbitrarily large ie assume that the underlying process has operated from the infinite past. When the series under consideration is stationary, all the characteristic roots of T lie within the unit circle and the quantities in question converge to well defined limiting values when this is done. It then emerges that $\beta = 0$ and that C can be obtained by solving the linear equations (eg. see Gardner et. al. 1980)

$$C = T C T' + \alpha \alpha'. \tag{6.3}$$

However, as indicated by the previous examples, the characteristic roots of T usually all lie on the unit circle and so the required convergence is not obtained. In these circumstances the processes involve a diffuse prior distribution $N(0,kI)$ where k is an arbitrarily large number.

It is particularly interesting to consider the special case of damped trend corrected exponential smoothing from section 5.3. As can be seen from (5.3c), the distribution of the growth rate considered by itself has a limiting form , yet (5.3b) indicates that the same cannot be said of the mean level. It is readily established in these circumstances that C has the general form

$$C = k\,C_1 + C_0 \qquad (6.4)$$

where

$$C_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad C_0 = \begin{bmatrix} 0 & (\alpha_1+\alpha_2\phi/(1-\phi^2))/(1-\phi) \\ (\alpha_1+\alpha_2\phi/(1-\phi)^2)/(1-\phi) & \alpha_2^2/(1-\phi^2) \end{bmatrix}$$

and k is arbitrarily large. Here we have an example of what is referred to as a partially diffuse prior distribution.

When applying exponential smoothing, knowledge of the prior distribution is not required. However, when reliable small sample estimates are sought, then the prior distribution is needed to seed the associated Kalman filter. The issue of initializing with partially or fully diffuse prior distributions without numerical instabilities arising from the use of a big k has been examined by Ansley and Kohn (1985) who developed a modified Kalman filter for this purpose. The generalization of their work in Snyder (1988a) together with its square root equivalent is applicable in this context.

7. ESTIMATION OF PARAMETERS

Both exponential smoothing and the Kalman filter assume that all the elements of T and α are known. But as indicated by the model undepinning damped trend corrected exponential smoothing some elements such as ϕ, α_1, α_2 must be assigned values before these methods can be used. Given that the forecasting performance of a model depends on our choice, it is appropriate to investigate methods designed to select good values.

7.1 Exponential Smoothing

In the context of exponential smoothing it is a common practice to evaluate trial values for the parameters using the sum of the squared one step ahead forecast errors

$$S_1 = \sum_{t=1}^{n} e(t)^2 \tag{7.1}$$

where n is the sample size. It is readily established from (3.8) that the typical one step ahead forecast error can be written as

$$e(t) = x' \; D^t \; (\; \beta(0) - b(0) \;) + \varepsilon(t) \tag{7.2}$$

The one step ahead forecast errors, under normal conditions, are biased estimates of the corresponding disturbances. However, when the discount matrix D is stable, the bias disappears in large samples and the resulting estimates approximate those obtained with the more usual residual sum of squares criterion. One interesting point is that the existance of this bias is not entirely disadvantageous. When values for the unknown parameters are chosen which lead to an unstable discount matrix D, according to (7.2) the forecast errors explode in size, leading to a large value for the criterion function (7.1). Such values are automatically penalized which means that any numerical optimization procedure conveniently rejects them when they are found. There is then no need to implement special procedures to constrain the parameters to stable values.

7.2 Kalman Filter

When reliable small sample results are required then the Kalman filter must be used in place of exponential smoothing. It is well known that the one step ahead forecast errors $\hat{e}(t)$ from a Kalman filter are *independent* $N(0, v_t \sigma^2)$ random variables where the heteroskedastic factor v_t can be obtained as a by-product of the associated computations. The essential idea is that as the sample size increases, the estimates and hence the forecasts become more reliable. This is reflected in the values taken by v_t which decline towards 1 as t increases. An appropriate criterion, in these circumstances, for evaluating trial values of the unknown parameters is

$$S_2 = \sum_{t=1}^{n} \hat{e}(t)^2 / v_t. \qquad (7.3)$$

The results obtained from numerically minimizing (7.3) can be shown to be minimum variance estimators.

7.3 Mixed Method

Despite being more accurate in small samples, the Kalman filter involves substantially higher computational loads than exponential smoothing. Furthermore, because its forecast errors are unbiased, there is no built-in penalty when a numerical optimisation routine strays into an unstable region of the parameter values. To circumvent both these difficulties, it seems sensible to divide the search into two stages. The first would invoke exponential smoothing to minimize (7.1). The resulting parameter values would then be used to seed the search with the Kalman filter in order to obtain more refined results in the second stage.

7.4 Single Pass Method

Because the method in section 7.3 employs a numerical optimization procedure, it involves quite heavy computational loads which, with the state of current computing technology, mitigates against its use in the important area of sales forecasting in large scale inventory systems with typically tens of thousands of line items. One possibility, in these circumstances, is to recognize that the state space framework (2.2) can be written in the equivalent form

$$y(t) = x' \beta(t-1) + \varepsilon(t)$$

$$\begin{bmatrix} \beta(t) \\ \alpha(t) \end{bmatrix} = \begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \beta(t-1) \\ \alpha(t-1) \end{bmatrix} + \begin{bmatrix} \alpha(t-1) \\ 0 \end{bmatrix} \varepsilon(t) \qquad (7.4)$$

This, itself conforms to a time variant counterpart of the framework (2.2). Initializing the associated Kalman filter with a diffuse prior

N(0,kI) on $\alpha(0)$, each stage of this algorithm yields a new estimate of α which can, in turn, be fed to the following stage. Thus in only one pass of the filter, without the use of any cumbersome numerical optimization procedure, this method yields estimates of the *smoothing vector* α.

The statistical properties of the resulting estimates are difficult to establish analytically because of the nonlinear nature of the problem. Some preliminary results with an equivalent procedure but confined to the model described in section 5.1 for the exponentially weighted average are presented in Snyder (1988b). Here it was established analytically that the estimates of the scalar α are consistent. Furthermore, a simulation study concerned with their statistical efficiency indicated that their rate of convergence to their true values was only marginally lower than estimates from a progressive implementation of the optimal method in section 7.2. Curiously, despite this marginal loss in efficiency, the single pass method, more often than not, gave marginally better forecasts.

One conceivable application of the single pass method is to employ its results as seed values for the Kalman filter approach in section 7.2. Casual experience with a local pilot computer package for estimating ARMA models based on this idea suggests that the seed values can be quite close to the optimal values provided that the models concerned are satisfactory representations of the data being analysed. However, more systematic studies are required before drawing definitive conclusions about the statistical properties of the estimators for more complex cases.

8. CONCLUSIONS

In this paper we explored the relationship between a particular version of the state space framework with only a single source of randomness and the forecasting technique of exponential smoothing. It was shown that the framework was equivalent to the ARMA class of models in its most general sense, representing an alternative, clearer statistical basis for exponential smoothing.

The proposed approach was contrasted with its more traditional state space counterpart containing many primary sources of randomness. An example indicated that the proposed approach is more general and may yield better forecasts. Other advantages flowed from the fact that the smoothing parameters explictly appear in the framework. Not only does this provide a clearer perception of the link with exponential smoothing but it admits new possibilities for reducing computational overheads with the mixed and the single pass methods described in sections 7.3 and 7.4. Taken together, the results suggest that the framework (2.2) is the natural and most straight forward statistical explanation of the linear exponential smoothing methods of forecasting.

REFERENCES

Akaike, H., 1974, Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, Annals of the Institute of Statistical Mathematics, 26, 363-387.

Ansley, C.F. and R. Kohn, 1985, Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions, The Annals of Statistics, 13, 1286-1316.

Brenner, J. L., D.A. D'Esopo and A.G.Fowler, 1968, Difference equations in forecasting formulas, Management Science, 14, 141-159.

Box, G.E.P. and G.M. Jenkins, 1976, _Time series analysis: Forecasting and control_, revised ed. (Holden Day, San Fransisco, CA).

Brown, R. G., 1959, _Statistical forecasting for inventory control_ (Mc.Graw-Hill, New York).

Brown, R.G. 1963, _Smoothing, Forecasting and prediction of discrete time series_, (Prentice Hall, Englewood Cliffs, N.J.).

Cogger, K.O., 1974, The optimality of general-order exponential smoothing, _Operations Research_, 22, 858-867.

Cogger, K.O., 1985, Introduction to the commentaries on Gardner (1985).

Gardner, E. S., 1985, Exponential smoothing: the state of the art, _Journal of Forecasting_, 4, 1-28.

Gardner, G., A.C. Harvey and G.D.A. Phillips, 1980, An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering, _Applied Statistics_, 29, 311-322.

Godolphin, E.J. and P.J. Harrison, 1975, Equivalence theorems for polynomial-projecting predictors, _Journal of the Royal Statistical Society_, B, 37, 205-215.

Goodman, M.L., 1974, A new look at higher-order exponential smoothing for forecasting, _Operations Research_, 22, 880 - 888.

Harrison, P.J., 1967, Exponential smoothing and short-term sales forecasting, _Management Science_, 13, 821-842.

Harrison, P.J. and C.F.Stevens, 1976, Bayesian Forecasting, Journal of the Royal Statistical Society, B, 38, 205-228.

Harvey, A. C., 1984, A unified view of statistical forecasting procedures, Journal of Forecasting, 3, 245-275.

Harvey, A.C. and R.G. Peirse, 1984, Estimating missing observations in economic time series, Journal of the American Statistical Association, 79, 125-131.

Holt, C. E., 1957, Forecasting trends and seasonals by exponentially weighted averages, O.N.R. memorandum No. 52, Carnegie Institue of Technology.

Koehler, A.B. and E.S. Murphree, 1988, A comparison of results from state space forecasting with forecasts from the Makridakis competition, International Journal of Forecasting, 4, 45-55.

Kohn, R. and C.F. Ansley, 1986, Estimation, prediction and interpolation for ARIMA models with missing data, Journal of the American Statistical Association, 81, 751-761.

Ledolter, J. and G.E.P. Box, 1978, Conditions for the optimality of exponential smoothing forecast procedures, Metrica, 25, 77-93.

McKenzie, E, 1974, A comparison of standard forecasting systems with the Box-Jenkins approach, The Statistician, 23, 107-116.

McKenzie, E., 1976, A comparison of some standard seasonal forecasting systems, The Statistician, 25, 3-14.

Makridakis, S., A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewkandowski J. Newton, R. Parzen, and R. Winkler, 1982, The accuracy of extrapolation (time series) methods: results of a forecasting competition', Journal of Forecasting, 1, 111-153.

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

Nerlove, M. and S. Wage, 1964, Some observations on adaptive forecasting, Management Science, 10, 207-224.

Roberts, S.A., 1982, A general class of Holt-Winters type forecasting models', Management Science, 28, 808-820.

Snyder, R.D., 1985, Recursive estimation of dynamic, linear statistical models, Journal of the Royal Statistical Society, B, 47, 272-276.

Snyder, R.D., 1988a, Computational aspects of Kalman filtering with a diffuse prior distribution, Journal of Statistical Computation and Simulation, (forthcoming).

Snyder, R.D., 1988b, Progressive tuning of simple exponential smoothing forecasts, Journal of the Operational Research Society (forthcoming).

Theil, H. and S.Wage, 1964, 'Some observations on adaptive forecasting, Management Science, 10, 198-206.

Winters, P.R., 1960, Forecasting sales by exponentially weighted averages, Management Science, 6, 324 - 342.

APPENDIX

Here we adapt the result in Akaike (1974) to the particular version of the state space framework of this paper. Essentially we demonstrate how the equations of any model conforming to the framework (2.2) can be manipulated to eliminate the state variables and leave us with an ARMA model. More specifically, we take the expression for $\beta(t-j)$ corresponding to (2.2b) and backsolve it to period t-r. We then multiply the result by the corresponding coefficient a_j from the characteristic equation (2.4c) of T and sum with respect to j to give

$$\sum_{j=0}^{r} a_j \ \beta(t-j) = \left(\sum_{j=0}^{r} a_j \ T^{r-j} \right) \beta(t-r) + \sum_{j=0}^{r-1} a_j \sum^{r} T^{i-j-1} \ \alpha \ \varepsilon(t-i)$$

By the Hamilton-Cayley theorem any matrix satisfies its own characteristic equation so that the term in brackets representing the coefficent of $\beta(t-r)$ in the above expression equals zero. Furthermore, the double summation can be rearranged so that we get

$$\sum_{j=0}^{r} a_j \ \beta(t-j) = \sum_{j=1}^{r} \left(\sum_{i=0}^{j-1} a_i \ T^{j-i-1} \right) \alpha \ \varepsilon(t-j) \qquad \text{(A1)}$$

Then lag (2.2a) by i, multiply the result a_i, sum with respect to i and use (A1) to eliminate the offending $\beta(t-i)$, to give

$$\sum_{j=0}^{r} a_j \ y(t-j) = x' \sum_{j=1}^{r} \left(\sum_{i=0}^{j-1} a_i \ T^{j-i-1} \right) \alpha \ \varepsilon(t-j) + \sum_{j=0}^{r} a_j \ \varepsilon(t-j)$$

Given that $a_0 = 1$, this conforms to the ARMA model (2.3) where the autoregressive and moving average parameters are given by (2.4a) and (2.4b). This completes the proof but it should be noted that although Akaike's (1974) original proof was undertaken in the context of stationarity assumption.