MONASH UNIVERSITY

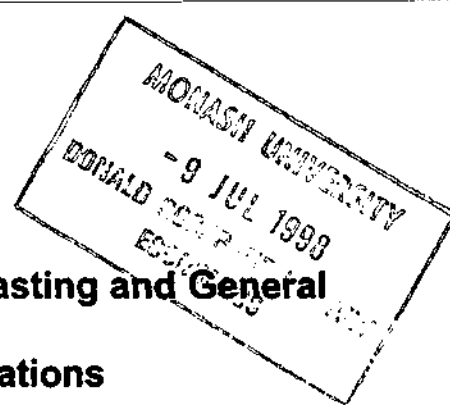**AUSTRALIA**

Exponential Smoothing Methods of Forecasting

and General ARMA Time Series Representations

Roland G. Shami and Ralph D. Snyder

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Exponential Smoothing Methods of Forecasting and General

ARMA Time Series Representations

Roland G. Shami[*] and Ralph D. Snyder[*]

1. ABSTRACT

The focus of this paper is on the relationship between the exponential smoothing methods of forecasting and the integrated autoregressive-moving average models underlying them. In this paper we derive, for the first time, the general linear relationship between their parameters. A method, suitable for implementation on computer, is proposed to determine the pertinent quantities in this relationship. It is illustrated on common forms of exponential smoothing. It is also applied to a new seasonal form of exponential smoothing with seasonal indexes which always sum to zero.

[**] Department of Econometrics & Business Statistics
Faculty of Business and Economics - Monash University
Clayton Campus - Clayton - Vic 3168
[*] e-mail Roland.Shami@buseco.monash.edu.au
[*] e-mail Ralph.Snyder@buseco.monash.edu.au

1. INTRODUCTION

This paper is concerned with the links between exponential smoothing (Holt, 1957; Brown, 1959; Winters, 1960) and the underlying integrated autoregressive-moving average processes. Particular cases of the relationship have been considered by Box and Jenkins (1976) and Roberts (1982). In this paper, however, we seek the relationship in general form.

Our strategy is to cast both the general ARMA model and the data generating process (DGP) underlying exponential smoothing in terms of first-order multi-state recurrence relationships. Then we identify the so-called equivalent transformation linking both formulations. This is used to establish the required general linear relationship between the parameters of the exponential smoothing DGP and the parameters of the general ARMA model. A computational method, suitable for implementation on computer, is outlined. It is applied to some common examples of exponential smoothing. It is also applied to a new version of seasonal exponential smoothing.

2. EXPONENTIAL SMOOTHING

The most general linear form of exponential smoothing emerged as an adjunct of a comparison made by Box and Jenkins (1976, Appendix A5.3) of their approach to time series analysis with that of Brown's general discounted least squares approach (Brown, 1962). It is centred on a first-order error correction relationship. In typical period t, a prediction $\hat{y}_t$ is compared with the actual series value y_t to give the one-step ahead prediction error

$$e_t = y_t - \hat{y}_t. \tag{2.1}$$

Quantities such as level, growth rate and seasonal effects, which define the major components of the series under consideration, are collected together into a k-vector b_t. They are computed recursively with the first-order relationship

2

$$b_t = Tb_{t-1} + \alpha e_t, \tag{2.2}$$

T being a fixed $k \times k$ transition matrix and α a k-vector of smoothing parameters. A linear combination of the components of b_t is used to generate the next prediction using

$$\hat{y}_{t+1} = x'b_t. \tag{2.3}$$

Here x is a fixed k-vector.

The recurrence relationship (2.2) is seeded with a fixed k-vector β. More specifically

$$b_0 = \beta. \tag{2.4}$$

The method is then applied sequentially to the sample $y_1, y_2, \ldots, y_n$ of the time series. Its use is predicated on the assumption that the quantities T, α, x, β are known or have been assigned trial values.

Example 1

The additive form of Winters' method is based on a local trend and local seasonal effects. Letting L_t denote the local level, T_t the local rate of change, and F_t the local seasonal effect, this method for quarterly data, in error correction form (Gardner, 1985), relies on the relationships $L_t = L_{t-1} + T_{t-1} + \alpha_1 e_t$, $T_t = T_{t-1} + \alpha_2 e_t$ and $F_t = F_{t-4} + \alpha_3 e_t$. By defining the state vector in time t as $b_t = (L_t, T_t, F_t, F_{t-1}, F_{t-2}, F_{t-3})'$, these relationships can be converted to first-order form to give $x = (1,1,0,0,0,1)'$ and

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

Future values of the time series are unknown. They may be represented by random variables $\tilde{y}_1$, $\tilde{y}_2$,.... where the notation ~ is used to indicate uncertain quantities. Given the nature of the relationships in exponential smoothing, it is assumed that future values of the series are governed by the data generating process

$$\tilde{y}_{t+1} = x'\tilde{b}_t + \tilde{e}_t \qquad (2.5)$$

$$\tilde{b}_t = T\tilde{b}_{t-1} + \alpha\tilde{e}_t \qquad (2.6)$$

$$\tilde{e}_t \sim NID(0,\sigma^2). \qquad (2.7)$$

The relationship (2.6) is seeded with the value of the current state vector $\tilde{b}_n = b_n$.

It also seems reasonable to assume that the same process underlies past time series values. Thus (2.5), (2.6) and (2.7), together with the seed condition

$$\tilde{b}_0 = \beta, \qquad (2.8)$$

are taken to be the data generating process underlying the Box-Jenkins version of exponential smoothing. This data generating process is the innovations form of the linear state space model (ISSM). It is distinguished from a conventional state space model in that is possesses only one source of randomness - the so-called innovations. It is usually considered in conjunction with Kalman filtering (Snyder, 1985), but is more conveniently used with exponential smoothing. A generalisation, to non-linear state space models, is considered in Ord, Koehler and Snyder (1997).

3. DERIVATION OF GENERAL ARMA REPRESENTATIONS

In this section we show, using the traditional lag operator approach, how to convert any ISSM to its equivalent *general* ARMA representation

4

$$y_t = \sum_{j=1}^{k} \varphi_j y_{t-j} - \sum_{j=1}^{k} \theta_j e_{t-j} + e_t. \tag{3.1}$$

The word 'general' is used to emphasise the fact that (3.1) incorporates both stationary and non-stationary time series. An ARIMA(0,2,2) process, for example, can be written as

$$y_t = 2y_{t-1} - y_{t-2} - \theta_2 e_{t-2} - \theta_1 e_{t-1} + e_t. \tag{3.2}$$

Note that there has been a small change in notation. In the previous section, it was important to maintain the distinction between exponential smoothing and the associated data generating process, two things that are often confused in the literature. Now the focus will be on models alone. The notation is simplified by dropping the convention of the circumflex (~) to designate random variables.

The traditional lag operator L can be applied to the ISSM to derive the corresponding general ARMA model. The lag operator, defined by $Ly_t = y_{t-1}$, can be applied to the recurrence relationship (2.6) to give

$$b_t = (I - TL)^{-1} \alpha e_t. \tag{3.3}$$

This may be used to eliminate b_{t-1} from the measurement equation (2.5) to give

$$y_t = x'(I - TL)^{-1} L\alpha e_t + e_t. \tag{3.4}$$

Let $\lambda_1, \lambda_2, \ldots \lambda_k$ be the roots of the characteristic equation of T. Then

$$(I - TL)^{-1} = A(L) \Big/ \prod_{j=1}^{k} (1 - \lambda_j L). \tag{3.5}$$

Here $A(L)$ is a polynomial matrix of at most degree k-1 in L and can be found by Gaussian elimination. Equation (3.4) may be rewritten as

$$\prod_{j=1}^{k} (1 - \lambda_j L) y_t = x' A(L) L\alpha e_t + \prod_{j=1}^{k} (1 - \lambda_j L) e_t, \tag{3.6}$$

Terms in this expression may be expanded as follows:

$$\prod_{j=1}^{k}\left(1-\lambda_j L\right)=1-\sum_{j=1}^{k}\varphi_j L^j \tag{3.7}$$

$$x'A(L)L\alpha=\sum_{j=1}^{k}\gamma_j L^j. \tag{3.8}$$

Substitution of these expansions into (3.6) gives the general ARMA representation (3.1) where

$$\theta_j=\varphi_j-\gamma_j. \tag{3.9}$$

Example 2

The local linear trend model underlying Holt's trend corrected exponential smoothing (Holt, 1957) is $y_t=L_{t-1}+T_{t-1}+e_t$, $L_t=L_{t-1}+T_{t-1}+\alpha_1 e_t$, $T_t=T_{t-1}+\alpha_2 e_t$ with $e_t\sim NID\left(0,\sigma^2\right)$.

For this model $x=(1,\,1)$, $\alpha=\left(\alpha_1,\alpha_2\right)'$ and $T=\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$. Thus $I-TL=\begin{bmatrix}1-L & -L\\0 & 1-L\end{bmatrix}$,

$\left|I-TL\right|=1-2L+L^2$ and so $\varphi_1=2$, $\varphi_2=-1$, while $A(L)=\begin{bmatrix}1-L & L\\0 & 1-L\end{bmatrix}$. Hence

$x'A(L)L\alpha=(\alpha_1+\alpha_2)L-\alpha_1 L^2$. Using (3.9), it follows that $\theta_1=2-\alpha_1-\alpha_2$ and $\theta_2=-1+\alpha_1$. This confirms the well known result that the ARIMA(0,2,2) model (3.2) underlies Holt's method.

4. EQUIVALENT STATE SPACE MODELS

In this paper we use a first-order representation of general ARMA models from Snyder (1985). State variables are defined by the partial sums $b_{t,i}=\varphi_i y_t-\theta_i e_t+b_{t-1,i+1}$ where $b_{t,k+1}=0$. Hence $y_t=b_{t-1,1}+e_t$ and $b_{ti}=\varphi_i b_{t-1,i}+b_{t-1,j+1}+(\varphi_i-\theta_i)e_t$. These relationships define a particular ISSM, namely

$$y_t=\bar{x}'\bar{b}_{t-1}+e_t \tag{4.1}$$

6

$$\bar{b}_t = \bar{T}\bar{b}_{t-1} + \gamma e_t \qquad (4.2)$$

where $\bar{x}' = (1, 0, ..., 0)$, $\bar{\alpha}' = \gamma' = (\gamma_1, \gamma_2, ..., \gamma_k)$ and

$$\bar{T} = \left[\varphi \left| \frac{I_{k-1}}{0'_{k-1}} \right. \right] . \qquad (4.3)$$

Here φ is the k-vector of autoregressive coefficients, I_{k-1} is a $(k-1) \times (k-1)$ identity matrix and 0_{k-1} is a $(k-1)$-vector of zeroes. $\bar{T}$ is a so-called companion matrix, one that is commonly used in control theory (Skelton, 1988). The model will be referred to as ISSM1.

As already illustrated, any linear exponential smoothing DGP can always be placed in the form of the ISSM

$$y_t = x' b_{t-1} + e_t \qquad (4.4)$$

$$b_t = T b_{t-1} + \alpha e_t . \qquad (4.5)$$

We now explore the possibility of transforming any ISSM into the particular form ISSM1.

First, let Q be a non-zero $k \times k$ matrix. Multiply (4.5) by it to give

$$Q b_t = QT b_{t-1} + Q\alpha e_t . \qquad (4.6)$$

Assume that Q has the property that

$$\bar{T}Q = QT . \qquad (4.7)$$

Then (4.6) becomes

$$Q b_t = \bar{T} Q b_{t-1} + Q\alpha e_t . \qquad (4.8)$$

Furthermore, assume that

$$x' = \bar{x}' Q . \qquad (4.9)$$

Then (4.4) can be written as

$$y_t = \bar{x}' Q b_{t-1} + e_t \qquad (4.10)$$

By letting

$$\bar{b}_t = Q b_t \qquad (4.11)$$

and

$$\gamma = Q\alpha, \qquad (4.12)$$

relationships (4.10) and (4.8) conform to the ISSM1 structure. Note that the condition (4.7) indicates that T and $\bar{T}$ are equivalent matrices and that Q has identical properties to a similar transformation matrix.

Applying (3.9) to (4.12) we get

$$\theta = \varphi - Q\alpha. \qquad (4.13)$$

For the first time, we have a general expression for the linear relationship between the smoothing parameters and moving average parameters. The only problem, therefore, is to find a method for determining φ and Q.

The matrix Q must satisfy the conditions (4.7) and (4.9). The condition (4.9) implies that the first row of Q must be x'. Condition (4.7) is like Liapunov's equation $(AX + XB + Q = 0)$ and similar reasoning to solve this kind of equations is used here. By vectorising the remaining unknown elements of Q and the first column of $\bar{T}$ (the ϕ's) to give a vector X, (see Appendix), the condition (4.7) may be rearranged as system of k^2 linear equations

$$AX = B. \qquad (4.14)$$

If the matrix A is of full rank then it may be solved directly for the unknowns. In cases where it is not of full rank it will be seen that a particular basic solution suffices.

Example 3 (Local Linear Trend)

From example 2 we get

$$\overline{T} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 1 & 1 \\ q_1 & q_2 \end{bmatrix}.$$

The condition (4.7) for this particular case, after vectorisation, is the system

$$\begin{bmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -2 \\ 0 \end{bmatrix}.$$

The rank of A is 4 and the required solution is $X' = (-1, 0, 2, -1)$. Thus

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

The usual ARIMA(0,2,2) representation $y_t = 2y_{t-1} - y_{t-2} - \theta_2 e_{t-2} - \theta_1 e_{t-1} + e_t$ applies where

$\theta_1 = 2 - \alpha_1 - \alpha_2$ and $\theta_2 = -(1 - \alpha_1)$, in agreement with the earlier result obtained with the

backward shift operator.

Example 4

The method was coded in the computer language Gauss to handle more complicated versions of

exponential smoothing. The program was tested on Winters' version of exponential smoothing

(Example 1) for quarterly data. The associated equation system (4.14) consists of 36 equations

with 36 unknown variables. The rank of A, however, is 35. A value of zero is therefore assigned

to one of the unknowns, in this case the last element φ_6 in φ. The effect is to simplify the

general ARMA representation by reducing the maximum lag of the autoregressive component

by 1. The solution, summarised in the form (4.13), was

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

the equivalent general *ARMA* model being $y_t = y_{t-1} + y_{t-4} - y_{t-5} - \sum_{j=1}^{5} \theta_j e_{t-j} + e_t$. This

generalises, for the case where there are p seasons per year, to the result (Roberts, 1982) that a

SARIMA$(0,1,p+1) \times (0,1,0)_p$ underlies Winters method for any seasonal frequency.

Example 5

The seasonal indexes in the model underlying Winters method follow a random walk. It is not

possible to ensure that they sum to zero, a common property of additive seasonal indexes. The

following is an alternative specification that overcomes this problem when generating

predictions. It is an adaptation of an approach developed by Harvey (1984) in the context of

multi-disturbance structural time series models. It takes the form, in the quarterly case:

$$y_t = L_{t-1} + T_{t-1} - F_{t-1} - F_{t-2} - F_{t-3} + e_t.$$

$$L_t = L_{t-1} + T_{t-1} + \alpha_1 e_t$$

$$T_t = T_{t-1} + \alpha_2 e_t$$

$$F_t = -F_{t-1} - F_{t-2} - F_{t-3} + \alpha_3 e_t.$$

The generalisation to any seasonal period is obvious.

This model may be converted to an ISSM where $x' = (1,1,-1,-1,-1)$, $\alpha' = (\alpha_1,\alpha_2,\alpha_3,0,0)$ and

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

Involving one less state variable than the ISSM for Winters' model, the associated equation system (4.14) has full rank. The solution obtained with the aid of the computer program is

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 & 1 & -1 & -1 & -1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & -1 & 0 \\ -1 & 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ 0 \\ 0 \end{bmatrix}.$$

This, like the previous example, generalises to a $SARIMA(0,1,p+1) \times (0,1,0)_p$ model.

5. CONCLUSIONS

In this paper we have found, for the first time, the general relationship between the parameters of the exponential smoothing DGP and the parameters of the corresponding general ARMA model. A computer program, based on this relationship, was developed and then applied to more complicated versions of exponential smoothing. It was verified that the program reproduced results for more common forms of exponential smoothing.

A new seasonal version of exponential smoothing was proposed. It always yields predicted seasonal indexes which sum to zero. It also involves one less state variable if the Kalman filter is to be used in place of exponential smoothing. It is therefore commended for use in practice as an alternative to Winters' method.

REFERENCES

Archibald C. B. (1990) : "Parameter Space of the Holt-Winters' Model", *International Journal of Forecasting*, Vol 6, pp 199-209.

Barnett S. and Cameron G. R. (1985) : *Introduction to Mathematical Control Theory*, Oxford: Clarendon Press.

Box G.E.P. and Jenkins G.M. (1970) : *Time Series Analysis Forecasting and Control*, San Francisco: Holden-Day.

Brown G. R. (1959) : *Statistical Forecasting for Inventory Control*, New York: Mc-Graw Hill, .

Brown G. R. (1962) : *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, Prentice Hall, N.J.

DeRusso P. M. et al (1965) : *State Variables for Engineers*, New York: Wiley.

Gardner S. E. Jr. (1985) : "Exponential Smoothing: The State of the Art", *Journal of Forecasting*, Vol 4, pp 1-28.

Harvey C. A. (1984) : "A Unified View of Statistical Forecasting Procedures", *Journal of Forecasting*, Vol 3, pp 245-275.

Holt C. C. (1957) : "Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages", *Office of Naval Research*, Research Memorandum No 52.

Ord K. J. Koehler B. A. and Snyder D. R. (1997) : "Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models ", *Journal of the American Statistical Association*, Vol 92, pp 1621-1629.

Roberts A. S. (1982) : "A General Class .of Holt-Winters Type Forecasting Model", *Management Science* , Vol 28, pp 808-820..

Skelton E. R. (1988) : *Dynamics Systems Control*, John Wiley & Sons.

Snyder D. R. (1985) : "Recursive Estimation of Dynamic Linear Models ", *The Journal of the Royal Statistical Society - Series B*, Vol 47, pp 272-276.

Winters R. P. (1960) : "Forecasting Sales by Exponentially Weighed Moving Averages", *Management Science* , Vol 6, pp 324-342.

APPENDIX : Vectorisation of Equations (4.7) and (4.9)

Q may be written in terms of row vectors as follows

$$Q = \begin{bmatrix} x^{\cdot} \\ q_2^{\cdot} \\ \vdots \\ q_k^{\cdot} \end{bmatrix} \tag{A.1}$$

Collecting the unknowns from (4.7) to the left hand side, we get

$$\begin{bmatrix} \phi_1 x^{\cdot} \\ \vdots \\ \phi_{k-1} x^{\cdot} \\ \phi_k x^{\cdot} \end{bmatrix} + \begin{bmatrix} q_2^{\cdot} \\ \vdots \\ q_k^{\cdot} \\ 0^{\cdot} \end{bmatrix} - \begin{bmatrix} 0^{\cdot} \\ q_2^{\cdot} T \\ \vdots \\ q_k^{\cdot} T \end{bmatrix} = \begin{bmatrix} x^{\cdot} T \\ 0^{\cdot} \\ \vdots \\ 0^{\cdot} \end{bmatrix} \tag{A.2}$$

Vectorising (A.2) yields

$$AX = b$$

where

$$X^{\cdot} = \begin{pmatrix} q_2^{\cdot} & \cdots & q_k^{\cdot} & \phi^{\cdot} \end{pmatrix}, \tag{A.3}$$

and $\qquad b^{\cdot} = \begin{pmatrix} x^{\cdot} T & 0_1 & \cdots & 0_{k^2-k} \end{pmatrix} \tag{A.4}$

in which 0_r means the rth zero. Furthermore

$$A = \begin{bmatrix} [I_k]_1 & & & & E_1 \\ [-T^{\cdot}]_1 & \ddots & & & E_2 \\ & \ddots & \ddots & & \vdots \\ & & \ddots & [I_k]_{k-1} & E_{k-1} \\ & & & [-T^{\cdot}]_{k-1} & E_k^{\cdot} \end{bmatrix} \tag{A.5}$$

where E_r is a $k{\times}k$ zero matrix with the rth column amended to x. $[I_k]_r$ represents the rth identity matrix of order k with a similar convention for $[-T^{\cdot}]_r$.