ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Exponential Smoothing: A Prediction Error Decomposition Principle

Ralph D Snyder

August 2004

Working Paper 15/04

Exponential Smoothing: A Prediction Error Decomposition Principle

Ralph D. Snyder
Department of Econometrics and Business Statistics,
Monash University
Clayton, Victoria 3800, Australia

August 25, 2004

Abstract

In the exponential smoothing approach to forecasting, restrictions are often imposed on the smoothing parameters which ensure that certain components are exponentially weighted averages. In this paper, a new general restriction is derived on the basis that the one-step ahead prediction error can be decomposed into permanent and transient components. It is found that this general restriction reduces to the common restrictions used for simple, trend and seasonal exponential smoothing. As such, the prediction error argument provides the rationale for these restrictions.

Keywords: time series analysis, prediction, exponential smoothing, ARIMA models, state space models

JEL Classification: C22, C53

1 Introduction

The exponential smoothing approach to forecasting relies on smoothing parameters, the sizes of which reflect the effect of structural change. When the smoothing parameters are chosen so that the invertibility conditions of equivalent ARIMA models are satisfied, it has been shown that exponential smoothing discounts the importance of older quantities in associated calculations (Brenner, 1968; McClain and Thomas, 1973, Sweet, 1983). However, these conditions differ markedly from much tighter restrictions (Gardner, 1985) commonly used in practice. It has been found (Hyndman, Koehler, Snyder and Grose, 2002) that tighter restrictions can translate into better forecasts, a point that supports current practice. The practical restrictions, however, are somewhat arbitrary. They have never been properly justified with respect to an underlying principle. The purpose of this paper is to identify and apply the missing principle.

2 Multiple Source of Error State Space Model (MSOE)

State space models and exponential smoothing are known to be closely linked (Harrison, and Stevens, 1971; Harvey, 1990). The state space framework that serves the purpose of this paper best is:

$$y_t = \overline{h}' x_t + u_t \tag{1a}$$

$$x_t = F x_{t-1} + G v_t. \tag{1b}$$

Equation (1a) is the *measurement equation*. It shows how the observable series value y_t is related to a random $k-$vector x_t called the *state vector*, and a random variable u_t called the *measurement disturbance*. The u_t are normally and independently distributed with mean 0 and a common variance σ^2. Each u_t measures *temporary unanticipated* change, that is stochastic change that impacts on only the period in which it occurs. The k-vector $\overline{h}$ is fixed.

The state vector x_t summarises the history of the process. Its evolution through time is governed by the first-order recurrence relationship (1b) where F is a fixed $k \times k$ matrix called the *transition matrix*, G is also a fixed $k \times k$ matrix, and v_t is a random k-vector of what are called the *system disturbances*. The v_t are normally and independently distributed with mean 0 and variance matrix $\sigma^2 \Omega$ where Ω is a symmetric, positive semi-definite matrix. The purpose of v_t is to model the effect of *structural change*, that is unanticipated change that persists through time.

The covariance between v_t and u_t is given by $\sigma^2\omega$ where ω is a fixed k-vector. u_t and v_s are independent for all distinct periods s and t. The model (1) is *invariant* because the vectors $\overline{h}$, ω and matrices F, G and Ω are independent of time.

The elements of ω and Ω are usually unknown. In the quest for parsimony the following additional assumptions are often made:

1. The elements of v_t are mutually independent; hence the off-diagonal elements of Ω are zero.

2. u_t and v_t are independent; hence $\omega = 0$.

The effect of these assumptions is to reduce the number of unknown parameters in Ω and ω from $k^2 + k$ to k.

3 Single Source of Error State Space Models (SSOE)

If Equation (1b) is substituted into Equation (1a) the equation $y_t = h'b_{t-1} + g'v_t + u_t$ is obtained where $h' = \overline{h}'F$ and $g' = \overline{h}'G$. The term $h'b_{t-1}$ is the one-step ahead prediction of y_t. The remainder $e_t = g'v_t + u_t$ is the one-step ahead prediction error. Its composition reflects that fact that prediction errors can possess two sources of error: the error $g'v_t$ induced by structural change and the transient error u_t.

An alternative to the independence assumptions, to achieve a parsimonious representation, is to assume that u_t and v_t are perfectly correlated. Then u_t and v_t are perfectly correlated with e_t so that $v_t = \overline{\alpha}e_t$ and $u_t = \delta e_t$ where $\overline{\alpha}$ is a non-negative fixed k-vector and δ is a non-negative scalar. The state space model can then be rewritten as

$$y_t = h'x_{t-1} + e_t \tag{2a}$$
$$x_t = Fx_{t-1} + G\overline{\alpha}e_t. \tag{2b}$$

In effect, the number of parameters is again reduced from $k^2 + k$ to k. The scalar δ is ignored because it does not directly appear in this single source of error specification. At first sight it might be thought that this perfect correlation assumption is likely to be very restrictive. However, the examples considered in Sections 4-6 indicate that this need not be the case.

An important byproduct of this specification is that $e_t = g'\overline{\alpha}e_t + \delta e_t$, something that must be true for all non-zero values of e_t. It follows that the parameter vector $\overline{\alpha}$, as well as being non-negative, must satisfy the linear restriction

$$g'\overline{\alpha} \leq 1. \tag{3}$$

It suggests that the elements of $\overline{\alpha}$ effectively allocate the prediction error amongst the unobserved components of the model. In effect we have used what might be termed a *one-step ahead prediction error decomposition principle* to derive the restriction (3). The restriction will be referred to as the *prediction condition.*

An equivalent variation of the specification of the the single source of error state space model is

$$\begin{aligned} y_t &= h'x_{t-1} + e_t && \text{(4a)} \\ x_t &= Fx_{t-1} + \alpha e_t. && \text{(4b)} \end{aligned}$$

where $\alpha = G\overline{\alpha}$. It is the more traditional form of the single source of error state space model (Ord, Koehler and Snyder, 1997). Equivalent restrictions on the k-vector α can be derived from the prediction condition on $\overline{\alpha}$. If G is non-singular, the restrictions take the form

$$\begin{aligned} G^{-1}\alpha &\geq 0 && \text{(5a)} \\ h'\alpha &\leq 1. && \text{(5b)} \end{aligned}$$

In some applications G may be singular, in which case it is simplest to elucidate the restrictions on a case by case basis.

The recurrence relationship

$$x_t = Dx_{t-1} + \alpha y_t, \tag{6}$$

where $D = F - \alpha h'$, may be derived by eliminating the error from (4). The solution to this relationship is

$$x_t = D^t x_0 + \sum_{j=0}^{t-1} D^j \alpha y_{t-j} \tag{7}$$

It shows that the state vector depends on past values of a series. In the presence of structural change, it would be expected that the state vector is influenced less by older series values than more recent ones. Structural change implies that α should take values that ensure that $\alpha D^j \rightarrow 0$ as $j \rightarrow \infty$. Unless $\alpha = 0$, the case of no structural change, this condition holds when the eigenvalues of D lie within the unit circle. This leads to additional restrictions on the vector α, herein referred to as the *structural* change conditions. It was shown in Snyder, Ord and Koehler (2001) that these conditions are equivalent to the invertibility conditions for equivalent ARIMA models.

4 Local Level Model

One of the simplest state space models involves a local level A_t that follows a random walk over time. The series values are randomly scattered about the local levels. More specifically

$$y_t = a_t + u_t \tag{8a}$$
$$a_t = a_{t-1} + v_t. \tag{8b}$$

The correlation between u_t and v_t is designated by ρ.

Equation (8b) may be substituted into Equation (8a) to give

$$y_t = a_{t-1} + v_t + u_t$$

The term a_{t-1} is the one-step ahead prediction, while $e_t = v_t + u_t$ is the one-step ahead prediction error. The prediction error has two components: one *permanent* (v_t) and the other *temporary* (u_t). In demand forecasting applications, the permanent component might reflect the effect of new customers or the impact of new suppliers (competitors) in a market.

Applying the restriction $\rho = 1$ instead of the independence restriction, the permanent and temporary disturbances must also correlate perfectly with the one-step ahead prediction error e_t; in other words $v_{1t} = \alpha e_t$ and $u_t = \delta e_t$ where α and δ are non-negative parameters. The local level model (8) can then be rewritten as

$$y_t = a_{t-1} + e_t \tag{9a}$$
$$a_t = a_{t-1} + \alpha e_t \tag{9b}$$

It is the single source of error version of the local level model (Ord et. al., 1997). It is the statistical model underlying simple exponential smoothing (Brown, 1959). The associated prediction condition is

$$0 \leq \alpha \leq 1. \tag{10}$$

The size of the parameter α is a measure of the impact of structural change in a time series. When $\alpha = 0$, successive levels are equal: the case of no structural change. When $\alpha = 1$, the model reduces to a random walk, a case at the other extreme where a time series has no parametric structure (except the variance parameter).

A recurrence relationship corresponding to the general relationship (6) is

$$a_t = \delta a_{t-1} + \alpha y_t. \tag{11}$$

It describes the evolution of the level over time. Note that $\delta = 1 - \alpha$. Under the condition (10), the level can be viewed as a weighted average. In traditional expositions of exponential smoothing (Brown, 1959), the condition (10) is imposed to permit this interpretation. As has been seen here, there is a more fundamental reason for it. It was derived from the prediction error decomposition principle.

The structural change condition requires that $\alpha\delta^j \to 0$ as $j \to \infty$. This occurs if $-1 < \delta \leq 1$. The equivalent condition, in terms of α, is

$$0 \leq \alpha < 2. \tag{12}$$

Advocates of the broader condition (12) argue that it provides greater flexibility. Indeed maximum likelihood estimates of α obtained under this restriction often exceed one on typical economic time series. Proponents of the narrower condition (10), however, argue that the added flexibility is counterproductive. An α in excess of one is seen as evidence of the existence of patterns in a time series such as a trend that are not covered by a local level model. It is interpreted as a signal that the local level model is not appropriate for the data and is likely to yield inferior forecasts.

5 Local Trend Model

A local level may be supplemented by a time dependent growth rate b_t which follows a random walk $b_t = b_{t-1} + v_{2t}$ where v_{2t} is another disturbance. The resulting local trend model is

$$\begin{aligned} y_t &= a_t + u_t && \text{(13a)} \\ a_t &= (a_{t-1} + v_{1t}) + (b_{t-1} + v_{2t}) && \text{(13b)} \\ b_t &= b_{t-1} + v_{2t} && \text{(13c)} \end{aligned}$$

Unlike the usual local trend model (Harvey, 1991), the equation for the current level contains two disturbances. This local trend model is a special case of the general framework (1).

The equation $y_t = a_{t-1} + b_{t-1} + v_{1t} + v_{2t} + u_t$ is obtained when a_t is eliminated from Equation (13a). Given that $a_{t-1} + b_{t-1}$ is now the one-step ahead prediction, the prediction error is now given by $e_t = v_{1t} + v_{2t} + u_t$. The prediction error has three components, two of them permanent. As before, one of the permanent disturbances is associated with the change in the underlying level. The other is the permanent change in the rate of growth. It is assumed that the three disturbances are potentially correlated.

When it is assumed that the three disturbances are perfectly correlated, they are also perfectly correlated with the one-step ahead prediction error, so that $v_{1t} = \overline{\alpha}_1 e_t$, $v_{2t} = \overline{\alpha}_2 e_t$ and $u_t = \delta E e_t$ where $\overline{\alpha}_2$ is a parameter. The resulting single source of error model is

$$y_t = a_{t-1} + b_{t-1} + e_t \tag{14a}$$
$$a_t = a_{t-1} + b_{t-1} + (\overline{\alpha}_1 + \overline{\alpha}_2)\, e_t \tag{14b}$$
$$b_t = b_{t-1} + \overline{\alpha}_2 e_t \tag{14c}$$

It can be rewritten as

$$y_t = a_{t-1} + b_{t-1} + e_t \tag{15a}$$
$$a_t = a_{t-1} + b_{t-1} + \alpha_1 e_t \tag{15b}$$
$$b_t = b_{t-1} + \alpha_2 e_t \tag{15c}$$

where $\alpha_1 = \overline{\alpha}_1 + \overline{\alpha}$ and $\alpha_2 = \overline{\alpha}_2$ This is the more traditional form of the local linear trend model found in Hyndman et. al. (2002). It may be established that the region for the parameters then becomes $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, $\alpha_1 < 1$ and $\alpha_2 \leq \alpha_1$.

Yet another way of writing the model is

$$y_t = a_{t-1} + b_{t-1} + e_t \tag{16}$$
$$a_t = (1-\alpha)\,(a_{t-1} + a_{t-1}) + \alpha y_t \tag{17}$$
$$b_t = (1-\beta)\, b_{t-1} + \beta\,(b_t - b_{t-1}) \tag{18}$$

where $\alpha = \alpha_1$ and $\beta = \alpha_2/\alpha_1$. It is obtained by solving (15b) for e_t and substituting the result into Equation (15c). It is the model underlying the original form of trend corrected exponential smoothing (Holt, 2002). The above feasible region for the parameters can be re-expressed as $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$, conditions that have been traditionally advocated (Makridakis, Wheelwright and Hyndman, 1998) for trend corrected exponential smoothing. A contribution of this paper has been to show that these conditions can be derived from the prediction error decomposition principle, instead of being imposed by assumption as has been the tradition.

The invertibility conditions for the equivalent ARIMA(0,2,2) process are $\alpha \geq 0$, $\alpha_2 \geq 0$ and $2\alpha_1 + \alpha_2 \leq 4$. This region is larger than the one derived from structural considerations.

6 Local Seasonal Model

An extension involving a seasonal factor c_t is

$$\begin{aligned} y_t &= a_t + c_t + u_t && \text{(19a)} \\ a_t &= a_{t-1} + b_{t-1} + v_{1t} + v_{2t} && \text{(19b)} \\ b_t &= b_{t-1} + v_{2t} && \text{(19c)} \\ c_t &= c_{t-m} + v_{3t}. && \text{(19d)} \end{aligned}$$

where m is the number of seasons per year. Substituting Equations (19b) and (19d) into Equation (19a) yields $y_t = a_{t-1} + b_{t-1} + c_{t-m} + e_t$ where $e_t = v_{1t} + v_{2t} + v_{3t} + u_t$. Adapting the perfect correlation argument above to include $v_{3t} = \overline{\alpha}_3 e_t$, the equivalent single source of error model is

$$\begin{aligned} y_t &= a_{t-1} + b_{t-1} + c_{t-m} + e_t && \text{(20a)} \\ a_t &= a_{t-1} + b_{t-1} + (\overline{\alpha}_1 + \overline{\alpha}_2)\, e_t && \text{(20b)} \\ b_t &= b_{t-1} + \overline{\alpha}_2 e_t && \text{(20c)} \\ c_t &= c_{t-m} + \overline{\alpha}_3 e_t && \text{(20d)} \end{aligned}$$

where $\overline{\alpha}_1 \geq 0$, $\overline{\alpha}_2 \geq 0$, $\overline{\alpha}_3 \geq 0$ and $\overline{\alpha}_1 + \overline{\alpha}_2 + \overline{\alpha}_3 \leq 1$. An equivalent representation is

$$\begin{aligned} y_t &= a_{t-1} + b_{t-1} + c_{t-m} + e_t && \text{(21a)} \\ a_t &= a_{t-1} + b_{t-1} + \alpha_1 e_t && \text{(21b)} \\ b_t &= b_{t-1} + \alpha_2 e_t && \text{(21c)} \\ c_t &= c_{t-m} + \alpha_3 e_t && \text{(21d)} \end{aligned}$$

where $0 \leq \alpha_1 \leq 1$, $0 \leq \alpha_2 \leq \alpha_1$, $\alpha_3 \geq 0$ and $\alpha_1 + \alpha_3 \leq 1$. Another equivalent form that corresponds to the traditional expression of the Winters (1960) additive method is

$$\begin{aligned} y_t &= a_{t-1} + b_{t-1} + c_{t-m} + e_t && \text{(22a)} \\ a_t &= (1-\alpha)(a_{t-1} + b_{t-1}) + \alpha y_t && \text{(22b)} \\ b_t &= (1-\beta)\, b_{t-1} + \beta (a_t - a_{t-1}) && \text{(22c)} \\ c_t &= (1-\gamma)\, c_{t-m} + \gamma (y_t - a_t) && \text{(22d)} \end{aligned}$$

where $\alpha = \alpha_1$, $\beta = \alpha_2/\alpha_1$ and $\gamma = \alpha_3/(1-\alpha_1)$. The parameters α, β and γ must all lie in the unit interval $[0, 1]$, conditions advocated by Winters to ensure in part that the seasonal effects can be interpreted as weighted averages. Again, this paper provides a more fundamental reason for the same restrictions. Extensive calculations indicate that the region defined by these restrictions is smaller than the region associated with the invertibility conditions [1](Hyndman, Akram and Archibald, 2003).

7 Conclusions

The one-step ahead prediction error decomposition principle was introduced in the paper and used to obtain restrictions on the smoothing parameters. When applied to the models underpinning the simple, trend and seasonal exponential smoothing methods, the restrictions imply that the smoothing parameters are effectively allocation parameters in the sense that they allocate the one-step ahead prediction error between the model components. It was established that the restrictions on the smoothing parameters are equivalent to those commonly used in practice. It was also shown that they are tighter than restrictions obtained with the traditional invertibility principle.

8 References

Brenner, J. L., D'Esopo, D. A. and Fowler, A. G. (1968), "Difference equations in in forecasting formulas," Management Science 15, 141-159.

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

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

Harrison, P. J., and Stevens, C. F. (1971), "A Bayesian Approach to Short-Term Forecasting," Operational Research Quarterly, 22, 341-362.

Harvey, A. C. (1991), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge: Cambridge University Press.

Harvey, A. C., and Koopman, S. (2000), "Signal Extraction and the Formulation of Unobserved Components Models," Econometrics Journal, 3, 84-107.

Holt, C. C. (1957). Forecasting Trends and Seasonal by Exponentially Weighted Averages, ONR Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh, USA (republished in International Journal of Forecasting 2004, 20, 5-13).

Hyndman, R., Koehler, A. B., Snyder, R. D., and Grose, S. (2002), "A State Space Framework for Automatic Forecasting Using Exponential

[1] My thanks to Muhammad Akram for producing plots that confirm this relationship.

Smoothing Methods," International Journal of Forecasting, 18, 439-454.

Hyndman, R. J., Akram, M., and Archibald, B. (2003), "Invertibility Conditions for Exponential Smoothing Models," Department of Econometrics and Business Statistics, Monash University, Working Paper Series.

McClain, J. O. and Thomas, L. J. "Response-variance tradeoffs in adaptive forecasting", Operations Research, 21, 554-568.

Makridakis, S., Wheelwright, S. C. and Hyndman, R. J.(1998), Forecasting: Methods and Applications, New York: John Wiley & Sons.

Ord, J. K., Koehler, A. B., and Snyder, R. D. (1987), "Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models," Journal of the American Statistical Association, 92, 1621-1629.

Snyder, R.D., Ord, J.K. and Koehler, A.B. (2001), Prediction Intervals for ARIMA Models, Journal of Business and Economic Statistics, 19, 217-225

Sweet, A. L. "Adaptive smoothing for forecasting seasonal series', AIIE Transactions, 13, 243-248.

Winters, P. R. (1960), "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, 1960, 324-342.