MONASH

WP No. 7/92

MONASH
UNIVERSITY
AUSTRALIA

RATIONALIZATION OF EXPONENTIAL SMOOTHING IN TERMS OF A STATISTICAL FRAMEWORK WITH MULTIPLICATIVE DISTURBANCES

Anne Koehler, Keith Ord and Ralph D. Snyder

Working Paper No. 7/92

July 1992

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0357 9

RATIONALIZATION OF EXPONENTIAL SMOOTHING IN TERMS OF A STATISTICAL FRAMEWORK WITH MULTIPLICATIVE DISTURBANCES

Anne Koehler, Keith Ord and Ralph D. Snyder

Working Paper No. 7/92

July 1992

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS COMMERCE & MANAGEMENT

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

RATIONALIZATION OF EXPONENTIAL SMOOTHING IN TERMS OF A STATISTICAL FRAMEWORK WITH MULTIPLICATIVE DISTURBANCES

by

Anne Koehler

Department of Decision Science
Miami University, Oxford, Ohio

Keith Ord

Department of Management Science
PennState University, Pennsylvania

Ralph D. Snyder

Department of Econometrics
Monash University, Melbourne, Australia

Third Draft: June, 1992

ABSTRACT

It is established in this paper that exponential smoothing, in its most general linear form, is an optimal method of forecasting in large samples for time series with an irregular component, the size of which depends on a local mean. As such it is demonstrated that exponential smoothing has a statistical basis that extends beyond the framework of Box and Jenkins.

KEYWORDS:

Exponential smoothing, forecasting, time series analysis.

1. INTRODUCTION

Although the exponential smoothing methods are simple, robust approaches to forecasting, particularly suited to business applications (Gardner, 1985), their statistical foundations have proven to be rather elusive. Brown (1963), the originator of the earliest versions of these techniques, attempted a rationalization in terms of a linear statistical model with constant coefficients: an approach that is incompatible with a major feature of the technique, namely its capacity to accommodate structural change. Explanations for Holt's (1957) methods based on the work of Muth (1960), Theil and Wage (1964), Nerlove and Wage (1964) and culminating in the structural approach of Harvey and Todd (1983), have proven to be more satisfactory in that the coefficients in their models are treated as time dependent random variables which may adapt to structural change. However, independence assumptions imposed on the disturbances of their models to restrict the number of parameters for computational purposes result in unusual autocorrelation requirements which can reduce the effectiveness of their approach (Newbold, 1983). Only the framework of Box and Jenkins (1976) provides a reasonably satisfactory basis for exponential smoothing because, as they show, it underlies all linear forms of this method. The links become particularly transparent when the equivalent state space framework (Snyder, 1985, 1988) is used instead of the more usual ARMA and ARIMA representations. The state of the process, represented in period t by a vector b_t, is assumed to evolve over time according to the first-order Markovian recurrence relationship

$$b_t = Tb_{t-1} + \alpha e_t , \qquad (1a)$$

the e_t being NID$(0,\sigma^2)$ disturbances, T a fixed 'transition' matrix and α a fixed vector reminiscent of the smoothing parameters from exponential smoothing. The series values y_t are related to the state of the process by the so-called measurement equation

$$y_t = x'b_{t-1} + e_t , \tag{1b}$$

x being a fixed vector. The unusual feature of this framework in the state space context is that it relies on only one disturbance source, the e_t, and this occurs because it is a transformation of the ARMA and ARIMA representations which themselves involve only one primary source of randomness.

The fluctuations in economic and business time series often increase with the underlying level over time. Such phenomena may be modelled using a variation of the above framework with multiplicative disturbances:

$$y_t = x'b_{t-1}(1 + e_t) \tag{2a}$$

$$b_t = Tb_{t-1} + \alpha e_t(x'b_{t-1}) , \tag{2b}$$

the e_t[1] being independent, zero mean disturbances with a common variance σ^2. The purpose of the paper is to investigate this framework as the basis for the estimation and prediction of time series.

---

1 Note that these disturbances are not necessarily normally distributed.

2. STRUCTURAL MODELS WITH MULTIPLICATIVE DISTURBANCES

2.1 Local Level Model

Although it is convenient to develop a theory of estimation and prediction for the state space framework (2), most economic and business applications would normally rely on only a few special cases of it. The simplest is the local level model, where the state of the process in typical period t is represented by only a single value: the local level ℓ_t. The model takes the form

$$y_t = \ell_{t-1}(1 + e_t) \tag{3.1}$$

$$\ell_t = \ell_{t-1}(1 + \alpha e_t) \ , \tag{3.2}$$

the scalar α determining the amount of stochastic change in the level. The case, $\alpha = 0$, corresponding to a constant level, accommodates stable situations. But positive values of α ensure that the model accommodates structural change. In the special case of $\alpha = 1$, the model reduces to $y_t = y_{t-1}(1 + e_t)$, a multiplicative form of a random walk process.

2.2 Local Trend Model

In many applications it is advisable to augment the local level model with a growth rate g_t to give

$$y_t = (\ell_{t-1} + g_{t-1})(1 + e_t) \tag{4.1}$$

$$\ell_t = (\ell_{t-1} + g_{t-1})(1 + \alpha_1 e_t) \tag{4.2}$$

$$g_t = g_{t-1} + (\ell_{t-1} + g_{t-1})\alpha_2 e_t \tag{4.3}$$

α_1 and α_2 being the two elements of a vector α. It represents a local trend line, the position and slope of which changes over time in response to structural changes implicit in phenomena such as business cycles.

2.3 Seasonal Components

A third possibility involves seasonal indexes c_t which, in the case of p seasons per year, takes the form

$$y_t = (\ell_{t-1} + g_{t-1} + c_{t-p})(1 + e_t) \tag{5.1}$$

$$\ell_t = \ell_{t-1} + g_{t-1} + (\ell_{t-1} + g_{t-1} + c_{t-p})\alpha_1 e_t \tag{5.2}$$

$$g_t = g_{t-1} + (\ell_{t-1} + g_{t-1} + c_{t-p})\alpha_2 e_t \tag{5.3}$$

$$c_t = -\sum_{j=1}^{p-1} c_{t-j} + (\ell_{t-1} + g_{t-1} + c_{t-p})\alpha_3 e_t , \tag{5.4}$$

the last equation being the stochastic analogue of a common condition for seasonal indexes, namely that they should sum to zero over the period of a year, α_3 being a third parameter. Although this model contains lags in excess of 1, it is always possible, where necessary with the use of additional relationships, to transform it into one involving only first-order Markovian relationships, e.g. see Harvey (1990, p.172).

Between them, these models define the basis of a structural approach to time series analysis along the lines of Harvey and Todd (1983). It may be distinguished from their approach in that (a) it only relies on a single primary disturbance source and (b) the disturbances are multiplicative in form.

3. ESTIMATION OF STATE VECTORS

The dynamic system (2) has the closed form solution

$$b_t = D^t b_0 + \sum_{j=0}^{t-1} D^j \alpha y_{t-j} \tag{6}$$

where $D = T - \alpha x'$. (7)

The role of D is reminiscent of the discount factor in an exponentially weighted average and can be called a discounting matrix in cases where $D^t \to 0$ and $t \to \infty$. Provided that the smoothing parameters α satisfy the conditions shown in Table 1, it may be established that D behaves in this way for the structural models in the previous section. These conditions are reminiscent of those which ensure the invertibility of certain integrated moving average processes (Box and Jenkins, 1976).

Model	Local Level	Local Trend	Seasonal
Conditions	$0 < \alpha < 2$	$2\alpha_1 + \alpha_2 < 4$ $\alpha_1, \alpha_2 > 0$	$2\alpha_1 + \alpha_2 < 4$ $\alpha_3 < 1$ $\alpha_1, \alpha_2, \alpha_3 > 0$

Table 1: Conditions for D to be a discounting Matrix

Let $\hat{b}_0$ be the subjective estimate of the seed vector b_0 used prior to observing the sample. Assume that subsequent estimates $\hat{b}_t$ of the state vectors b_t are generated recursively with each new observation using the most general form of exponential smoothing (Box and Jenkins, 1976)

$$\hat{b}_t = T\hat{b}_{t-1} + \alpha(y_t - x'\hat{b}_{t-1}) \tag{8}$$

The closed form solution of (8) has the same basic structure as (6) and as a consequence

$$b_t - \hat{b}_t = D^t(b_0 - \hat{b}_0) . \tag{9}$$

The estimation error at time t therefore depends on the initial estimation error but is independent of the sample. Whatever the mean and variance of b_0, if D is a discount matrix, then the mean squared error of $\hat{b}_t$ coverges in large samples to zero. Exponential smoothing in its most general linear form, therefore not only provides consistent estimators for the ARIMA class of models, but also those conforming to the non-linear framework (2).

The exponential smoothing algorithm is conditional on specific values of α and so the latter must themselves be estimated. One possibility is to minimize the sum of squared relative errors criterion

$$S = \sum_{t=2}^{n} \hat{e}_t^2 / \hat{y}_t^2 \tag{10}$$

where $\hat{y}_t = x'\hat{b}_{t-1}$ and $\hat{e}_t = y_t - \hat{y}_t$.

Given that the structural models involve at most three parameters apart from σ, namely α_1, α_2, α_3, such an approach with numerical optimization procedures is likely to be viable in many applications. Note that σ can be estimated, for given values of the α's, with

$$\hat{\sigma} = \sqrt{S/(n-1)}$$

4. PREDICTION

After obtaining the estimate of the state vector at the end of period n with a sample of size n, estimates $\hat{b}_{t|n}$ of future state vectors b_t may be generated recursively with

$$\hat{b}_{t|n} = T\hat{b}_{t-1|n} \qquad t = n+1,\ n+2,\ \ldots \tag{11a}$$

and predictions of the series obtained with

$$\hat{y}_{t|n} = x'\hat{b}_{t-1|n}\,. \qquad t = n+2,\ n+2,\ \ldots \tag{11b}$$

The associated prediction error

$$\hat{e}_{t|n} = y_t - \hat{y}_{t|n}$$

reduces to

$$\hat{e}_{t|n} = x'(b_{t-1} - \hat{b}_{t-1|n})(1 + e_t) + x'\hat{b}_{t-1|n}\, e_t$$

so that the mean squared prediction error (scaled) is

$$v_{t|n} = x'B_{t-1|n}x(1 + \sigma^2) + \hat{y}^2_{t|n} \tag{12a}$$

where

$$E(y_t - \hat{y}_{t|n}) = \sigma^2 v_{t|n} \text{ and } E(b_{t-1} - \hat{b}_{t-1|n})(b_{t-1} - \hat{b}_{t-1|n})' = \sigma^2 B_{t-1|n}.$$

Subtracting (11a) from (2b) to give

$$b_t - \hat{b}_{t|n} = T(b_{t-1} - \hat{b}_{t-1|n}) + \alpha e_t x' b_{t-1}\ ,$$

it is readily shown that

$$B_{t|n} = TB_{t-1|n}T' + \alpha\alpha'\left(x'B_{t-1|n}x\sigma^2 + \hat{y}^2_{t|n}\right), \tag{12b}$$

a recurrence relationship which would be seeded, in large samples, with $B_{n|n} = 0$.

Both (12a) and (12b) form the basis of a method for computing mean squared prediction errors. The latter could be used for establishing confidence limits but the prediction errors $\hat{e}_{t|n}$ are unlikely to be normally distributed. Simulation studies would be required to ascertain the form of the distribution for this purpose.

5. CONCLUSIONS

There is a common belief that exponential smoothing is based on only special cases of the Box-Jenkins framework and is therefore a more restrictive approach to forecasting. However, Box and Jenkins themselves have shown that any ARIMA model can be estimated in large samples by linear first-order, error correction type recurrence relationships conforming to the structure in (8). In this paper, matters have been taken further by establishing that exponential smoothing in its most general linear form, can be used to estimate models from another statistical framework, namely the nonlinear one outlined in section 1. This conclusion may help explain why the exponential smoothing methods have proved to be so robust in practice.

One point of particular interest concerns the multiplicative random walk model outlined in Section 2.1. The results of this paper indicate that, like the conventional random walk model, it is best estimated with the so-called naive forecast $\hat{y}_{t|n} = y_n$. Given that price fluctuations are often proportional to the underlying level, this model should provide a more satisfactory basis for analysing the behaviour of share and commodity market prices.

References

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

Brown, R.G., 1959, *Statistical Forecasting for Inventory Control*, McGraw Hill, New York.

Brown, R.G., 1963, *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice Hall, Englewood Cliffs, N.J..

Gardner, E.S., 1985, "Exponential smoothing: the state of the art", *Journal of Forecasting*, 4, 1-28.

Harvey, A.C., 1990, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A.C. and P.H.J. Todd, 1983, "Forecasting economic time series with structural and Box-Jenkins models (with discussion)", *Journal of Business and Economic statistics*, 1, 299-315.

Holt, C.E., 1957, "Forecasting trends and seasonals by exponentially weighted averages", *ONR* memorandum No. 52, Carnegie Institute of Technology.

Muth, J.F., 1960, "Optimal properties of exponentially weighted forecasts", *Journal of the American Statistical Association*, 55, 299-304.

Nerlove, M.I. and S. Wage, 1964, "On the optimality of adaptive forecasting", *Management Science*, 10, 207-229.

Newbold, P., 1983, "Comment", *Journal of Business and Economic Statistics*, 1, 311-312.

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

Snyder, R.D., 1988, "Statistical foundations of exponential smoothing", Department of Econometrics Working Paper No. 5/88, Monash University.

Theil, H. and S. Wage, 1964, "On some observations on adaptive forecasting", *Management Science*, 10, 198-206.