Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Forecasting 22 (2006) 239–247

international journal of forecasting

www.elsevier.com/locate/ijforecast

# Exponential smoothing model selection for forecasting

Baki Billah [a], Maxwell L. King [b], Ralph D. Snyder [b,*], Anne B. Koehler [c,1]

[a] Department of Epidemiology and Preventive Medicine, Monash University, VIC 3800, Australia
[b] Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia
[c] Department of Decision Sciences and Management Information Systems, Miami University, Oxford, OH 45056, USA

**Abstract**

Applications of exponential smoothing to forecasting time series usually rely on three basic methods: simple exponential smoothing, trend corrected exponential smoothing and a seasonal variation thereof. A common approach to selecting the method appropriate to a particular time series is based on prediction validation on a withheld part of the sample using criteria such as the mean absolute percentage error. A second approach is to rely on the most appropriate general case of the three methods. For annual series this is trend corrected exponential smoothing: for sub-annual series it is the seasonal adaptation of trend corrected exponential smoothing. The rationale for this approach is that a general method automatically collapses to its nested counterparts when the pertinent conditions pertain in the data. A third approach may be based on an information criterion when maximum likelihood methods are used in conjunction with exponential smoothing to estimate the smoothing parameters. In this paper, such approaches for selecting the appropriate forecasting method are compared in a simulation study. They are also compared on real time series from the M3 forecasting competition. The results indicate that the information criterion approaches provide the best basis for automated method selection, the Akaike information criteria having a slight edge over its information criteria counterparts.


*Keywords:* Model selection; Exponential smoothing; Information criteria; Prediction; Forecast validation

## 1. Introduction

The exponential smoothing methods are relatively simple but robust approaches to forecasting. They are widely used in business for forecasting demand for inventories (Gardner, 1985). They have also performed surprisingly well in forecasting competitions against more sophisticated approaches (Makridakis et al., 1982; Makridakis & Hibon, 2000).

Three basic variations of exponential smoothing are commonly used: simple exponential smoothing (Brown, 1959); trend-corrected exponential smoothing (Holt, 1957); and Holt–Winters' method (Winters, 1960). A distinctive feature of these approaches is that a) time series are assumed to be built from unobserved components such as the level, growth and seasonal effects; and b) these components need to be adapted over time when demand series display the effects of

* Corresponding author. Fax: +61 613 9905 5474.
*E-mail addresses:* baki.billah@med.monash.edu.au (B. Billah), max.king@buseco.monash.edu.au (M.L. King), ralph.snyder@buseco.monash.edu.au (R.D. Snyder), koehleab@muohio.edu (A.B. Koehler).
[1] Fax: +1 513 529 9689.

0169-2070/$ - see front matter © 2005 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2005.08.002

structural changes in product markets. As these components may be combined by addition or multiplication operators, 24 variations of the exponential smoothing methods may be identified (Hyndman, Koehler, Snyder, & Grose, 2002). Given this proliferation of options, an automated approach to method selection becomes most desirable (Gardner, 1985; McKenzie, 1985).

Hyndman et al. (2002) provided a statistical framework for exponential smoothing based on the earlier work of Ord, Koehler, and Snyder (1997). The framework incorporated stochastic models underlying the various forms of exponential smoothing and enabled the calculation of maximum likelihood estimates of smoothing parameters. It also enabled the use of Akaike's information criterion (Akaike, 1973) for method selection. One issue not addressed was the preference for Akaike's information criterion over possible alternatives such as Schwarz (1978), Hannan and Quinn (1979), Mallows (1964), Golub, Heath, and Wahba (1979), and Akaike (1970). One aim, therefore, is to determine whether Akaike's information criterion (AIC) has a superior performance compared to its alternatives. Given that it was developed to minimise the forecast mean squared error, it might be hypothesised that the AIC has a natural advantage over the alternatives in forecasting applications, except possibly for Akaike's FPE which is asymptotically equivalent to the AIC.

The exponential smoothing methods were traditionally implemented without reference to a statistical framework so that other approaches were devised to resolve the method selection problem. Prediction validation (Makridakis, Wheelwright, & Hyndman, 1998) is one such approach. The sample is divided into two parts: the fitting sample and the validation sample. The fitting sample is used to find sensible values for the smoothing parameters, often with a sum of squared one-step ahead prediction error criterion. The validation sample is used to evaluate the forecasting capacity of a method with a criterion such as the mean absolute percentage error (MAPE). Another approach applies a general version of exponential smoothing on the assumption that it effectively reduces to an appropriate nested method when this is warranted by the data. Trend corrected exponential smoothing is applied to annual time series; Winter's method is applied to sub-annual time series. A second aim is to gauge the effectiveness of these traditional approaches relative to the information criterion approach to method selection.

The plan of this paper is as follows. State space models for exponential smoothing and an approach to their estimation are introduced in Section 2. Criteria to be used in model selection and a measure for comparing resulting forecast errors are explained in Section 3. A simulation study is discussed in Section 4. An application of the model selection criteria to the M3 competition data (Makridakis & Hibon, 2000) is given in Section 5. The paper ends with some concluding remarks in Section 6.

## 2. State space models

The state space framework in Snyder (1985), and its extension in Ord et al. (1997), provides the basis of an efficient method of likelihood evaluation, a sound mechanism for generating prediction distributions, and the possibility of model selection with information criteria. Important special cases, known as structural models, that capture common features of time series such as trend and seasonal effects, provide the foundations for simple exponential smoothing, trend corrected exponential smoothing and Holt–Winters' seasonal exponential smoothing. Of the 24 versions of exponential smoothing found in Hyndman et al. (2002), the scope of this study is limited to three linear cases.

The focus is on a time series that is governed by the innovations model (Snyder, 1985):

$$y_t = \boldsymbol{h}'\boldsymbol{x}_{t-1} + \varepsilon_t \tag{2.1}$$

$$\boldsymbol{x}_t = F\boldsymbol{x}_{t-1} + \boldsymbol{\alpha}\varepsilon_t. \tag{2.2}$$

Eq. (2.1), called the measurement equation, relates an observable time series value $y_t$ in typical period $t$ to a random $k$-vector $\boldsymbol{x}_{t-1}$ of unobservable components from the previous period. $\boldsymbol{h}$ is a fixed $k$-vector, while the $\varepsilon_t$, the so-called innovations, are independent and normally distributed random variables with mean zero and a common variance $\sigma^2$. The inter-temporal dependencies in the time series are defined in terms of the unobservable components with the so-called transition equation (2.2). $F$ is a fixed $k \times k$ 'transition' matrix and $\boldsymbol{\alpha}$ is a $k$-vector of smoothing parameters.