Available online at www.sciencedirect.com

ScienceDirect

International Journal of Forecasting 26 (2010) 661–665

international journal of forecasting

www.elsevier.com/locate/ijforecast

Damped trend exponential smoothing: A modelling viewpoint

Eddie McKenzie[a,1], Everette S. Gardner Jr.[b,*]

[a] University of Strathclyde, Department of Statistics and Modelling Science, Glasgow G11XH, Scotland, United Kingdom
[b] University of Houston, Bauer College of Business, Houston, TX 77204-6021, United States

Abstract

Over the past twenty years, damped trend exponential smoothing has performed well in numerous empirical studies, and it is now well established as an accurate forecasting method. The original motivation for this method was intuitively appealing, but said very little about why or when it provided an optimal approach. The aim of this paper is to provide a theoretical rationale for the damped trend method based on Brown's original thinking about the form of underlying models for exponential smoothing. We develop a random coefficient state space model for which damped trend smoothing provides an optimal approach, and within which the damping parameter can be interpreted directly as a measure of the persistence of the linear trend.

Keywords: Time series; Exponential smoothing; ARIMA models; State space models

1. Introduction

In a series of three papers (Gardner & McKenzie, 1985, 1988, 1989), we developed new versions of the Holt–Winters (Holt, 2004; Winters, 1960) methods of exponential smoothing that damp the trend as the forecast horizon increases. Since those papers appeared, damped trend exponential smoothing has performed well in numerous empirical studies, as discussed by Gardner (2006). In a review of evidence-based forecasting, Armstrong (2006) recommended the damped trend as a well established forecasting method that should improve accuracy in practical applications. In a review of forecasting in operational research, Fildes, Nikolopous, Crone, and Syntetos (2008) concluded that the damped trend can "reasonably claim to be a benchmark forecasting method for all others to beat". Additional empirical evidence using the M3 competition data (Makridakis & Hibon, 2000) is given by, Hyndman, Koehler, Ord, and Snyder (HKOS) (2008), who found that the use of the damped trend method alone compared favourably to model selection via information criteria.

Despite this record of empirical success, we still have no compelling rationale for the damped trend. Our original approach was pragmatic, based on the findings of the M-competition (Makridakis et al., 1982), which showed that the practice of projecting

* Corresponding author. Tel.: +1 713 743 4744.
E-mail addresses: ed@stams.strath.ac.uk (E. McKenzie), egardner@uh.edu (E.S. Gardner Jr.).
[1] Tel.: +44 0 141 548 3807.

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.07.001

a straight line trend into the future indefinitely was often too optimistic (or pessimistic). Thus, we added an autoregressive-damping parameter (ϕ) to modify the trend component in Holt's linear trend method. The result is a method which is stationary in first differences, rather than in second differences as is the case for Holt's method. If there is a strong, consistent trend in the data, we hypothesized that ϕ would be fitted at a value near 1, and the forecasts would be very nearly the same as Holt's; if the data are extremely noisy or if the trend is erratic, ϕ would be fitted at a value less than 1 to create a damped forecast function. This explanation may be intuitively appealing, but it says nothing about when trend damping is the optimal forecasting approach.

The aim of this paper is to provide a theoretical rationale for the damped trend based on Brown's (1963) original thinking about the form of underlying models for exponential smoothing. His preference was for processes that are thought to be *locally constant*. Brown argued that although the parameters of the model may be constant within any local segment of time, they may change from one segment to the next, and the changes may be either sudden or smooth. We present a new model for the damped trend method that accommodates both types of change. Interestingly, our interpretation of this model essentially reverses our original thinking on the use of damped trend forecasting in practice.

2. A modelling viewpoint

Our development is based on the class of single source of error (SSOE) state space models (HKOS). We begin with the model for a linear trend with additive errors:

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t \tag{1}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + (1-\alpha)\varepsilon_t \tag{2}$$

$$b_t = b_{t-1} + (1-\beta)\varepsilon_t, \tag{3}$$

where $\{y_t\}$ is the observed series, $\{\ell_t\}$ is its level and $\{b_t\}$ is the gradient of its linear trend. This model has a single source of error $\{\varepsilon_t\}$, hence the name. We note that what we have to say here still applies even if we consider models with multiple sources of error. Compared to the presentation in HKOS, we have written the coefficients of the innovations in the level (2) and gradient (3) revision equations in a slightly unusual way to simplify some of the results which follow. The model (Eqs. (1)–(3)) has a reduced form as the ARIMA(0, 2, 2):

$$(1-B)^2 y_t = \varepsilon_t - (\alpha+\beta)\varepsilon_{t-1} + \alpha\varepsilon_{t-2}. \tag{4}$$

The two models are equivalent, but the state space expression is easier to interpret, especially when the parameters take on extreme values. The usual minimum mean square error (MMSE) forecasts of this model can be generated using the recursive formulae of Holt.

To damp the trend component in Eqs. (1)–(3), we incorporate an autoregressive-damping parameter ϕ to create another SSOE model:

$$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t \tag{5}$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + (1-\alpha)\varepsilon_t \tag{6}$$

$$b_t = \phi b_{t-1} + (1-\beta)\varepsilon_t. \tag{7}$$

This model (Eqs. (5)–(7)) has a reduced form as the ARIMA(1, 1, 2):

$$(1-\phi B)(1-B)y_t = \varepsilon_t - (\alpha+\phi\beta)\varepsilon_{t-1} + \phi\alpha\varepsilon_{t-2}. \tag{8}$$

Note that the gradient revision equation (7) is an AR(1) rather than the random walk form used in Eq. (3). Thus, revision equation (7) allows the gradient to change, but in a stationary way, whereas in Eq. (3) such changes are non-stationary and the longer-term behaviour is quite different.

In Eqs. (5)–(7), we can interpret ϕ as a direct measure of the persistence of the linear trend. With ϕ close to 1, the linear trend is highly persistent, but values of ϕ moving away from 1 toward zero indicates weaker persistence. And, of course, $\phi = 0$ would indicate the complete absence of any linear trend.

Now we recall Brown's idea of a locally constant model and apply it to the *gradient* of the linear trend. For the model in Eqs. (1)–(3), this means that the usual random walk form of the gradient revision equation (3) holds for a while, but then the gradient changes to a new value, which holds for a while, and then changes again, and so on. Thus, we have runs of the linear trend model given by Eqs. (1)–(3), but each run ends when the gradient revision equation (3) restarts with a new gradient. Such behaviour may be modelled by rewriting the gradient revision equation in the form

$$b_t = A_t b_{t-1} + (1-\beta)\varepsilon_t, \tag{9}$$