ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

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

Incorporating a Tracking Signal into State Space Models for Exponential Smoothing

Ralph D. Snyder and Anne B. Koehler

July 2006

Working Paper 16/06

Incorporating a Tracking Signal into State Space Models for Exponential Smoothing

Ralph D. Snyder
Monash University

Anne B. Koehler
Miami University

Abstract

It is a common practice to complement a forecasting method such as simple exponential smoothing with a monitoring scheme to detect those situations where forecasts have failed to adapt to structural change. It will be suggested in this paper that the equations for simple exponential smoothing can be augmented by a common monitoring statistic to provide a method that automatically adapts to structural change without human intervention. It is shown that the resulting equations conform to those of damped trend corrected exponential smoothing. In a similar manner, exponential smoothing with drift, when augmented by the same monitoring statistic, produces equations that split the trend into long term and short term components.
Keywords: Forecasting, exponential smoothing, tracking signals.

JEL CLASSIFICATION: C32

Introduction

A tracking signal is used for monitoring forecast errors to detect structural changes in time series. The goal of this oversight is to maintain control of a forecasting system by responding to out-of- control signals that are based on the forecast errors. One of the widely used tracking signals was introduced by Trigg (1964) and is based on a smoothed forecasting error. We will show that the effect of incorporating this smoothed forecasting error directly into the level equation of simple exponential smoothing is equivalent to using the damped trend exponential smoothing method of Gardner and McKenzie (1985). Thus we provide evidence that damped trend exponential smoothing itself adapts to structural change without intervention. This procedure of augmenting a state equation is then extended to the case of exponential smoothing with a drift. In this latter situation, the resulting method includes equations that split the trend into long-term and short-term trend components.

The paper is organized in the following manner. In the second section we introduce the smoothed-error statistic and the innovations state space models for simple exponential smoothing with and without a drift. The equivalence between damped trend exponential smoothing and simple exponential smoothing that is augmented with a tracking signal is shown in the third section. The equivalence of a new exponential smoothing method and an augmented simple smoothing with a drift is explained in the fourth section. In the last section, we summarize the contributions of these theoretical derivations.

State space models and smoothed-error statistic

We begin by introducing state space models for time series that underlie the methods of simple exponential smoothing and simple exponential smoothing with a drift. We also define the smoothed-error statistic that is part of historical tracking signals and adaptive smoothing techniques.

The innovations state space model for simple exponential smoothing is

$$y_t = \ell_{t-1} + \varepsilon_t \quad (1a)$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t \quad (1b)$$

where y_t represents the time series at time *t*, ℓ_t is the level of the time series at time *t*, α is the smoothing parameter, and $\{\varepsilon_t\}$ is a sequence of independent identically and normally distributed random variables with mean 0 and standard deviation σ.

The innovations state space model for the exponential smoothing method with an upward or downward drift *b* is

$$y_t = \ell_{t-1} + b + \varepsilon_t \quad (2a)$$

$$\ell_t = \ell_{t-1} + b + \alpha\varepsilon_t\,. \quad (2b)$$

In both models ε_t is the forecast error. For later use, observe that in these models both Equation (1a) and (2a) can be expressed in terms of the current level as follows:

$$y_t = \ell_t + (1-\alpha)\varepsilon_t \quad (3)$$

The smoothed-error statistic often used in tracking signals is the weighted average

$$\bar{\varepsilon}_t = \phi\bar{\varepsilon}_{t-1} + (1-\phi)\varepsilon_t \quad (4)$$

where the parameter ϕ controls the size of the weights and must lie between 0 and 1. Under the null hypotheses that the forecasts are under control, $\bar{\varepsilon}_t$ has a normal distribution with mean 0 and standard deviation $\sigma(1-\phi)\big/\sqrt{1-\phi^2})$. The smoothed-error statistic $\bar{\varepsilon}_t$ can be employed as a tracking signal, and this standard deviation can be used to establish an out-of-control region for a specified level of significance. This method of monitoring is a simple alternative to the the more traditional approach based on the tracking signal (Trigg, 1964) formed from the ratio of $|\bar{\varepsilon}_t|$ to the smoothed mean absolute deviation. Gardner (1983), however, argued that the distribution for the Trigg tracking signal cannot be derived analytically, a serious impediment to its use in practice. Either way, the parameter(s) α (and b) would have to be re-estimated and ℓ_t re-initialized when an out-of-control signal is produced. Yet another possibility (Trigg and Leach, 1967) is an adaptive method of forecasting where the Trigg ratio itself is used in place of the smoothing parameter α in simple exponential smoothing. By doing this, the smoothing parameter adapts over time in response to variations in the amount of structural change. However, Gardner (2006) states, "In Gardner (1985), I concluded that there was no credible evidence in favor of any of the numerous forms of adaptive smoothing."

A model combining simple exponential smoothing and a tracking signal

In this section, we show that if the level ℓ_t is made self-adjusting in the innovations model for simple exponential smoothing, the result is a model for damped trend exponential smoothing. We start by augmenting Equation (1b) with the smoothed-error statistic in Equation (4) as follows:

$$\ell_t^* = \ell_{t-1}^* + \alpha\varepsilon_t + \delta\overline{\varepsilon}_t \tag{5}$$

where ℓ_t^* is the augmented level and δ is a parameter that controls the amount of adjustment.

Then, define $b_t = \delta\overline{\varepsilon}_t$ and substitute it into Equation (5) to obtain

$$\ell_t^* = \ell_{t-1}^* + b_t + \alpha\varepsilon_t\,. \tag{6}$$

Multiply Equation (4) by δ and use the definition of b_t to give

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

If we let $\beta = (1-\phi)\delta$, then

$$b_t = \phi b_{t-1} + \beta \varepsilon_t . \quad (8)$$

Next use Equation (8) to substitute for b_t in Equation (6), and find that

$$\ell_t^* = \ell_{t-1}^* + \phi b_{t-1} + \beta \varepsilon_t + \alpha \varepsilon_t . \quad (9)$$

If we let $\alpha^* = \alpha + \beta$, then

$$\ell_t^* = \ell_{t-1}^* + \phi b_{t-1} + \alpha^* \varepsilon_t . \quad (10)$$

The analogue to Equation (3) with our new level ℓ_t^* and smoothing parameter α^* is

$$y_t = \ell_t^* + (1 - \alpha^*) \varepsilon_t \quad (11)$$

Using Equation (10), substitute for ℓ_t^* in Equation (11) to obtain

$$y_t = \ell_{t-1}^* + \phi b_{t-1} + \alpha^* \varepsilon_t + (1 - \alpha^*) \varepsilon_t$$

Thus,

$$y_t = \ell_{t-1}^* + \phi b_{t-1} + \varepsilon_t \quad 12)$$

Equation (12), (10), and (8) form an innovations state space model for the damped trend exponential smoothing method of Gardner and McKenzie (1985). Thus, it has been shown that the self-correcting scheme that combines Equations (4) and (5) is equivalent to the damped trend exponential smoothing method of Gardner and McKenzie. We have also provided a statistical model for this forecasting method. Moreover, we offer this equivalence as one reason why damped trend exponential smoothing works so well in practice.

It is easy to see that with the damped trend model, the forecast for *h* periods ahead from time period *n* is

$$\hat{y}_n(h) = \ell_n^* + \phi b_n + \phi^2 b_n + \cdots + \phi^h b_n$$

The forecasted growth rate at time period $t+h$ is $\phi^h b_n$. When $0 < \phi < 1$, it converges to 0 as the forecast horizon *h* increases so that the forecasts converge to the fixed value $\ell_n^* + b_n\phi/(1-\phi)$ for the level. When there is a persistent long term growth in the data, one may prefer to use a model in which the forecasted growth rate does not disappear, such as model (2), Holt's linear exponential smoothing, or the model that arises in the next section.

Extension to simple exponential smoothing with a drift

The approach of augmenting the level equation with the smoothed-error statistic can be applied to the model for simple exponential smoothing with drift in Equation (2). This time the effect is to split the trend into two parts: a long-term trend and a damped trend.

As in the earlier case, the smoothed-error statistic $\bar{\varepsilon}_t$ is incorporated into the model by augmenting the level in Equation (2b) as follows:

$$\ell_t^* = \ell_{t-1}^* + b + \alpha\varepsilon_t + \delta\bar{\varepsilon}_t . \tag{13}$$

This time, in order to simplify the derivation, we use the lag operator *L*, where $La_t = a_{t-1}$ for any time series a_t. Equation (4) can be written in terms of $\bar{\varepsilon}_t$ as

$$\bar{\varepsilon}_t = \frac{(1-\phi)\varepsilon_t}{1-\phi L} . \tag{14}$$

When Equation (14) is substituted into Equation (13), we have

$$\ell_t^* = \ell_{t-1}^* + b + \alpha\varepsilon_t + \frac{\delta(1-\phi)\varepsilon_t}{1-\phi L}.$$

Define b_t by

$$b_t = b + \frac{\delta(1-\phi)\varepsilon_t}{1-\phi L}.$$

It follows that

$$b_t = \phi b_{t-1} + (1-\phi)b + \delta(1-\phi)\varepsilon_t \tag{15}$$

and that

$$\ell_t^* = \ell_{t-1}^* + b_t + \alpha\varepsilon_t = \ell_{t-1}^* + \phi b_{t-1} + (1-\phi)b + \delta(1-\phi)\varepsilon_t + \alpha\varepsilon_t \tag{16}$$

Letting $\alpha^* = \delta(1-\phi) + \alpha$ and $\beta = \delta(1-\phi)$, Equations (15) and (16) become

$$\ell_t^* = \ell_{t-1}^* + \phi b_{t-1} + (1-\phi)b + \alpha^*\varepsilon_t \tag{17}$$

$$b_t = \phi b_{t-1} + (1-\phi)b + \beta\varepsilon_t. \tag{18}$$

The analogue for y_t in Equation (3) is

$$\begin{aligned} y_t &= \ell_t^* + (1-\alpha^*)\varepsilon_t \\ &= \ell_{t-1}^* + \phi b_{t-1} + (1-\phi)b + \varepsilon_t \end{aligned} \tag{19}$$

In this case we have shown that the self-correcting scheme that combines Equations (13) and (4) is equivalent to Equations (17), (18), and (19). The latter form a very interesting new innovations state space model in which b and b_{t-1} may be interpreted as the long-term growth and short-term growth respectively. The forecast for *h* periods into the future from time period *n* with this model (see the Appendix for proof) is

$$\hat{y}_n(h) = \ell^*_n + \sum_{j=1}^{h}\left[\phi^j b_n + (1-\phi^j)b\right] \tag{20}$$

Thus, the forecasted growth rate in time period $n+j$ combines the short-term and long-term growth rates as

$$\hat{b}_n(j) = \phi^j b_n + (1-\phi^j)b\,.$$

When $0 \le \phi < 1$, the forecasted growth $\hat{b}_n(j)$ converges to the long term growth rate b as the prediction horizon *j* increases. The local level eventually increases at a constant rate rather than leveling out to a constant value.

Another interesting feature of this model is that by allowing ϕ to range from 0 to 1, it ranges between a local level model with drift and a local trend model. In other words, the model ranges between using one and two differencing operators. If $\phi = 0$, the reduced form of the model is an ARIMA(0,1,1) with a constant term *b*, that is, $(1-L)y_t = b + (1-\theta L)\varepsilon_t$. If $\phi = 1$, the reduced form of the model is an ARIMA(0,2,2) model, that is, $(1-L)^2 y_t = \left(1-\theta_1 L - \theta_2 L^2\right)\varepsilon_t$. Thus, by using this model when there is an upward trend in the data, one avoids the decision of whether the data contains a single or double unit root.

As an illustration of the application of the model, a time series that is composed of the gross domestic product GDP values in the U.S. from 1970 to 2004 was examined. We used the following restrictions on the parameters: $0 \le \phi \le 1$, $(\phi-1)\alpha^* < \phi\beta$, $(\phi\_1) < \phi\alpha^*$, $(1+\phi)\alpha^* + \phi\beta < 2(1+\phi)$. The latter three conditions are the invertibility conditions for the corresponding reduced model. The long run growth rate *b* was set to the slope of a classical tend line fitted to all 35 observations. To capture the initial local features of the series, the seed states ℓ^*_0 and b_0 were set to the intercept and slope respectively of a classical trend line fitted to only the first 5 observations. The optimal values for the parameters were found by minimizing the sum of the squared forecast errors (SSE) with Solver in Microsoft Excel. The results were $\phi = .939$, $\alpha^* = 1.004$, $\beta = .310$, b (long term growth rate) = .071, and b_{35} (the short term growth rate) = .055. In this case, the value of ϕ is close to 1 showing that short term growth rate is the most important component of the trend in the initial forecasts. Repeating this process for Australian GDP over the same time frame, we found $\phi = .916$, $\alpha^* = 1.042$, $\beta = .166$, b (long term growth rate) = .070, and b_{35} (the short term growth rate) = .060.

Conclusions

The most significant contribution of this paper is that it provides an explanation for why damped trend exponential smoothing has been so successful in practice. We show that it can be thought of as simple exponential smoothing with an embedded tracking signal that allows the forecast to adjust to structural change. Almost as important is a new method for exponential smoothing that resulted from incorporating a statistic for tracking forecast errors directly into a model for simple exponential smoothing with a drift. This method, and its corresponding innovations state space model, allows one to forecast a time series with two types of trend components: one for long-term growth and a second for short-term growth. In both cases, it is enlightening to know that some models essentially contain tracking signals.

References

Gardner, E.S. (1983). Automatic monitoring of forecast errors. *J Forecasting* **2**: 1-21.

Gardner, E.S. (1985). Exponential smoothing: The state of the art. *J Forecasingt* **5**: 1-28.

Gardner, E.S. (2006). Exponential smoothing: The state of the art – Part II, *Int J Forecastingt* (to appear in 2006).

Gardner, E.S. and McKenzie, E. (1985). Forecasting trends in time. *Mgmt Sci* **31:** 1237-1246.

Trigg, D.W. (1964). Monitoring a forecasting system. *Opl Res Q* **15**: 271-274.

Trigg, D.W. and Leach, A.G. (1967). Exponential smoothing with an adaptive response rate. *Opl Res Q* **18**: 53-59.

Appendix

Forecasts for the New Trend Model

In this appendix we will derive the h-period-ahead forecast in Equation (20) for the new combined trend model defined by Equations (17), (18), and (19). We will use $E(\cdot)$ to denote the conditional expected value of a variable at time t, given the initial values of the states, the values of the parameters, and the values of y_t for $t = 1,\ldots,n$. Thus, the forecast of y_{n+h} at time *n*, $\hat{y}_n(h)$, is $E(y_{n+h})$. We now show by mathematical induction for $h \geq 1$ that

$$\hat{y}_n(h) = \ell^*_n + \sum_{j=1}^{h}\left[\phi^j b_n + (1-\phi^j)b\right] \tag{21a}$$

$$E(\ell^*_{n+h}) = \ell^*_n + \sum_{j=1}^{h}\left[\phi^j b_n + (1-\phi^j)b\right] \tag{21b}$$

$$E(b_{n+h}) = \phi^h b_n + (1-\phi^h)b \tag{21c}$$

Using the equations for the new combined trend model when $h = 1$,

$$\begin{aligned}\hat{y}_n(1) = E(y_{n+1}) &= E(\ell^*_n + \phi b_n + (1-\phi)b + \varepsilon_{t+1}) \\ &= \ell^*_n + \phi b_n + (1-\phi)b\end{aligned}$$

$$E(\ell^*_{n+1}) = E(\ell^*_n + \phi b_n + (1-\phi)b + \alpha\varepsilon_t)$$
$$= \ell^*_n + \phi b_n + (1-\phi)b$$

$$E(b_{n+1}) = E(\phi b_n + (1-\phi)b) + \beta\varepsilon_t)$$
$$= \phi b_n + (1-\phi)b$$

Now we assume that (21) is true for *h-1* and show it is then true for *h*.

$$\hat{y}_n(h) = E(y_{n+h}) = E(\ell^*_{n+h-1} + \phi b_{n+h-1} + (1-\phi)b + \varepsilon_{n+h})$$
$$= \ell^*_n + \sum_{j=1}^{h-1}\left[\phi^j b_n + (1-\phi^j)b\right] + \phi\left[\phi^{h-1}b_n + (1-\phi^{h-1})b\right] + (1-\phi)b$$
$$= \ell^*_n + \sum_{j=1}^{h-1}\left[\phi^j b_n + (1-\phi^j)b\right] + \phi^h b_n + \left[\phi(1-\phi^{h-1}) + (1-\phi)\right]b$$
$$= \ell^*_n + \sum_{j=1}^{h-1}\left[\phi^j b_n + (1-\phi^j)b\right] + \phi^h b_n + (1-\phi^h)b$$
$$= \ell^*_n + \sum_{j=1}^{h}\left[\phi^j b_n + (1-\phi^j)b\right]$$

Similarly,

$$E(\ell^*_{n+h}) = E(\ell^*_{n+h-1} + \phi b_{n+h-1} + (1-\phi)b + \alpha\varepsilon_{n+h})$$
$$= \ell^*_n + \sum_{j=1}^{h}\left[\phi^j b_n + (1-\phi^j)b\right]$$

and

$$E(b_{n+h}) = E(\phi b_{n+h-1} + (1-\phi)b + \beta\varepsilon_{n+h})$$
$$= \phi^h b_n + (1-\phi^h)b$$