

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.

© 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

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, Nikolopoulos, 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.

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:

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

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

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

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):

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

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:

$$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t \quad (5)$$

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

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

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

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

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

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

where $\{A_t\}$ is a sequence of independent, identically distributed binary random variates with $P(A_t = 1) = \phi$ and $P(A_t = 0) = (1 - \phi)$.

At first sight, this is a strange model, but it is easy to see what happens in particular cases. If we wish to model a strongly persistent trend then ϕ will be close to 1, and the sequence $\{A_t\}$ will consist of long runs of 1s interrupted by occasional 0s. This yields long runs of a linear trend model with a similar gradient, one changing smoothly by means of Eq. (3), but which can change suddenly, with a small probability $(1 - \phi)$, to a completely different model. If ϕ is close to 0, there are long runs of 0s with occasional 1s, and so the model displays only a very weak linear trend (if any), with a range of different gradients. With values of ϕ between 0 and 1 we get a mixture, resulting in different models operating over shorter time scales, i.e., low persistence of trends. In passing, we note that the mean length of runs of a similar trend is given by $\phi/(1 - \phi)$, which may also be thought of as a way to measure the persistence of trend. We also note that Eq. (9) is not the only possible form we could use here. For example, if we wish to generate a greater level of variation at the gradient change-point, i.e. when $A_t = 0$, we could replace Eq. (9) by

$$b_t = A_t b_{t-1} + (1 - A_t) d_t + (1 - \beta) \varepsilon_t, \tag{10}$$

where $\{d_t\}$ is another, independent, white noise source, and we would obtain similar results. We will use Eq. (9) here because it is the simplest form.

The new state space model corresponding to the incorporation of the new gradient revision equation (9) is a random coefficient state space model:

$$y_t = \ell_{t-1} + A_t b_{t-1} + \varepsilon_t \tag{11}$$

$$\ell_t = \ell_{t-1} + A_t b_{t-1} + (1 - \alpha^*) \varepsilon_t \tag{12}$$

$$b_t = A_t b_{t-1} + (1 - \beta^*) \varepsilon_t, \tag{13}$$

whose reduced form is a random coefficient ARIMA(1, 1, 2):

$$(1 - A_t B)(1 - B)y_t = \varepsilon_t - (\alpha^* + A_t \beta^*) \varepsilon_{t-1} + A_t \alpha^* \varepsilon_{t-2}. \tag{14}$$

We use (α^*, β^*) here rather than (α, β) in order to emphasise that these coefficients will differ in our discussion of the two models from Eqs. (5)–(7) and (11)–(13), whereas the same value of ϕ will apply to both.

Although this random coefficient state space model may appear complex, it is simply a stochastic mixture of two well known forms. Thus, for example, Eq. (14) may be rewritten as

$$(1 - B)^2 y_t = \varepsilon_t - (\alpha^* + \beta^*) \varepsilon_{t-1} + \alpha^* \varepsilon_{t-2} \tag{15}$$

with probability ϕ , and

$$(1 - B)y_t = \varepsilon_t - \alpha^* \varepsilon_{t-1} \tag{16}$$

with probability $(1 - \phi)$.

In this model, $\{y_t\}$ is generated by the ARIMA(0, 2, 2) given by Eq. (15) or (4), the usual linear trend model, with probability ϕ ; but then, with probability $(1 - \phi)$, the gradient changes completely, with the generation process switching to the ARIMA(0,1,1) given by Eq. (16), the usual underlying model for simple exponential smoothing. The resulting process is a mixture of the two.

Now, in this model (11)–(14), the first differences $\{(1 - B)y_t\}$ are a stationary process, and thus, from the Wold Decomposition, have an infinite moving average representation, and hence an equivalent ARMA form. In this case, the ARMA will be of order (1, 2), with exactly the same autoregressive parameter ϕ , since its autocorrelation function may be shown to be of the form $\rho(k) = \phi^{k-2} \rho(2)$ for $k \geq 2$. It follows that $\{y_t\}$ can be generated by a stochastic difference equation of the form:

$$(1 - \phi B)(1 - B)y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \tag{17}$$

where $\{a_t\}$ is a white noise process which is different from the white noise process $\{\varepsilon_t\}$ in Eq. (14). The variance of $\{a_t\}$ and the parameters θ_1 and θ_2 in Eq. (17) are complicated functions of the parameters ϕ , α^* and β^* , and the variance of the innovation process $\{\varepsilon_t\}$ in Eq. (14). Except for ϕ , the parameters must be different, because those in Eqs. (11)–(13) have random components and those in Eqs. (5)–(7) are constants. Thus, the MMSE forecasts of y_t defined by Eq. (17) are the MMSE forecasts of the random coefficient ARIMA(1, 1, 2) given by Eq. (14), and thus also of our random coefficient state space model (Eqs. (11)–(13)). Moreover, the MMSE forecasts of (17) are clearly damped trend forecasts.

Hence, to summarize these relationships, the standard damped trend forecasts optimal for Eqs. (5)–(7) are also optimal for a random coefficient state space model of the form of Eqs. (11)–(13), with the same parameter value, ϕ , in both, but with different values of

α and β in Eqs. (11)–(13). The values of these corresponding parameters in Eqs. (11)–(13), α^* and β^* say, can be computed from the parameters of the damped trend model in Eqs. (5)–(7) by equating their autocorrelation functions, but our intention here is simply to note that the damped trend forecasts are also optimal for such a broader and more general class of models. We also argue that such a random coefficient state space model is itself often a good approximation to the behaviour of practically occurring non-seasonal time series, and that this is one of the main reasons for the empirical success of the damped trend method.

3. Other models/methods

The same discussion and argument will apply in the cases of other similar models that contain a linear trend component. In particular, we note two important cases here. The first is the additive seasonal model (of period n) which, in random coefficient form, is given by

$$y_t = \ell_{t-1} + A_t b_{t-1} + S_{t-n} + \varepsilon_t \quad (18)$$

$$\ell_t = \ell_{t-1} + A_t b_{t-1} + S_{t-n} + (1 - \alpha)\varepsilon_t \quad (19)$$

$$b_t = A_t b_{t-1} + (1 - \beta)\varepsilon_t \quad (20)$$

$$S_t = S_{t-n} + \gamma\varepsilon_t. \quad (21)$$

If the random coefficient A_t is replaced by the constant value 1 or 0, we obtain models for which the Holt–Winters-type linear trend with additive seasonality, or trend-free seasonality, forecasting methods, respectively, are optimal. If we replace A_t by ϕ , the damped trend version (e.g. Gardner & McKenzie, 1989) is optimal.

The second model we wish to extend is the linear trend version of the very important multiplicative error models of HKOS. It is given by

$$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t) \quad (22)$$

$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + (1 - \alpha)\varepsilon_t) \quad (23)$$

$$b_t = b_{t-1} + (1 - \beta)(\ell_{t-1} + b_{t-1})\varepsilon_t. \quad (24)$$

The importance of models of the form of Eqs. (22)–(24) lies in the fact that although the driving innovation terms have variances that are now functions of the level, exponential smoothing methods can be optimal nevertheless. The random coefficient version of this is given by

$$y_t = (\ell_{t-1} + A_t b_{t-1})(1 + \varepsilon_t) \quad (25)$$

$$\ell_t = (\ell_{t-1} + A_t b_{t-1})(1 + (1 - \alpha)\varepsilon_t) \quad (26)$$

$$b_t = A_t b_{t-1} + (1 - \beta)(\ell_{t-1} + A_t b_{t-1})\varepsilon_t, \quad (27)$$

and, for completeness, we note that the reduced random coefficient ARIMA may be written in the mixture form we have used before: with probability ϕ ,

$$(1 - B)^2 y_t = \omega_t - (\alpha + \beta)\omega_{t-1} + \alpha\omega_{t-2},$$

$$\text{where } \omega_t = (\ell_{t-1} + b_{t-1})\varepsilon_t, \quad (28)$$

and, with probability $(1 - \phi)$,

$$(1 - B)y_t = \omega'_t - \alpha\omega'_{t-1}, \quad \text{where } \omega'_t = \ell_{t-1}\varepsilon_t. \quad (29)$$

This form is essentially the same as Eqs. (15) and (16), except that the innovation process is now dependent on level.

4. Conclusions

We have developed a model, given by Eqs. (11)–(13), or (14), or (15) and (16), 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. Developing these models has led us to reverse our earlier view that a damped trend is a good approximation of a linear trend at short lead times, and is better for longer lead times because the linearity must eventually break down. Now, our argument is that the underlying random coefficient linear trend model is more realistic, i.e. is more often closer to the true process that underlies our time series, and that the linear trend model is simply a good approximation to it for short lead times. Technically, we are arguing that it makes more practical sense to model the uncertainty of the gradient process of our putative linear trend as a random coefficient autoregression (Eq. (13)) than as a random walk (Eq. (3)), thus greatly widening the legitimacy of damped trend forecasting.

We see this model as a natural extension of Brown's (1963) original work. Our aim is to capture the locally constant nature of the linear trend by means of its gradient, which may change either smoothly or suddenly. The random walk form of the gradient revision equation allows smooth change very well, but is less successful with occasional, sudden changes. Our random coefficient model accommodates both kinds of change.

Finally, we note that if we assume that the random coefficient state space model (Eqs. (11)–(13)) does indeed generate our observed time series, then damped trend forecasting may be optimal, but the corresponding prediction intervals will be much wider than if we assume the standard damped trend model of Eqs. (5)–(7). This is because of the extra variation introduced by the presence of the random binary coefficient, and may go some way toward explaining the often conservative performance of prediction intervals in this area. This important topic will be explored elsewhere.

Acknowledgements

We would like to thank Ralph Snyder and Rob Hyndman for their insightful comments on a talk describing this random coefficient model given at the ISF 2008 in Nice, France.

References

- Armstrong, J. S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting*, 22, 583–598.
- Brown, R. G. (1963). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Fildes, R., Nikolopoulos, K., Crone, S., & Syntetos, A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society*, 59, 1–23.
- Gardner, E. S., Jr. (2006). Exponential smoothing: The state of the art — Part II. *International Journal of Forecasting*, 22, 637–666.
- Gardner, E. S., Jr., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, 31, 1237–1246.
- Gardner, E. S., Jr., & McKenzie, E. (1988). Model identification in exponential smoothing. *Journal of the Operational Research Society*, 39, 863–867.
- Gardner, E. S., Jr., & McKenzie, E. (1989). Seasonal exponential smoothing with damped trends. *Management Science*, 35, 372–376.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 5–10.
- Hyndman, R., Koehler, A., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer-Verlag.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1, 111–153.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342.