Why the damped trend works

Everette S. Gardner, Jr.
University of Houston, Bauer College of Business
334 Melcher Hall, Houston, Texas 77204-6021
Tel. 713-743-4744; fax 713-743-4940
egardner@uh.edu

Eddie McKenzie
University of Strathclyde, Department of Statistics and Modelling Science,
Glasgow G1 1XH, Scotland
Tel. 0141 548 3807; fax 0141 552 2079
ed@stams.strath.ac.uk

Revised October 22, 2009

Why the damped trend works

The damped trend method of exponential smoothing is a benchmark that has been difficult to beat in empirical studies of forecast accuracy. One explanation for this success is the flexibility of the method, which contains a variety of special cases that are automatically selected during the fitting process. That is, when the method is fitted, the optimal parameters usually define a special case rather than the method itself. For example, in the M3-competition time series, the parameters defined the damped trend method only about 43% of the time using local initial values for the method components. In the remaining series, a special case was selected, ranging from a random walk to a deterministic trend. The most common special case was a new method, simple exponential smoothing with a damped drift term.

Key words: forecasting, time series, exponential smoothing

Why the damped trend works

Introduction

In forecasting with exponential smoothing, it is common to apply the damped trend method to every time series, although many attempts have been made to improve on this practice by selecting individual methods for each series. Examples include selection based on information criteria (Hyndman *et al*, 2008), expert systems (Flores and Pearce, 2000), and time series characteristics (Gardner and McKenzie, 1988). Although method selection procedures can result in simpler methods than the damped trend, they have yet to produce better forecast accuracy. For a review of the evidence, see Gardner (2006). See also Fildes (2001), who concluded that it is difficult to beat the damped trend when a single forecasting method is applied to a collection of time series. If individual methods are selected for each series, Fildes argued that it may be possible to beat the damped trend, although this has not been demonstrated and it is not clear how one should proceed. In a later review of forecasting in operational research, Fildes *et al* (2008) concluded that the damped trend can “reasonably claim to be a benchmark forecasting method for all others to beat.”

How do we explain the success of the damped trend method? In McKenzie and Gardner (2009), we presented a theoretical rationale based on an underlying random coefficient state space (RCSS) model, which we view as an extension of Brown’s (1963) original thinking about the form of underlying models for exponential smoothing. We aim to capture *local* time series behavior with a constant model whose parameters may change smoothly or suddenly. The RCSS model adapts to both types of change, and the damping parameter in the model may be interpreted as a measure of the persistence of trends.

This paper presents an alternative rationale for the damped trend aimed at the practical forecaster faced with the problem of method selection. We show that fitting the damped trend method is actually a means of automatic selection from a variety of special cases, ranging from a random walk to a deterministic trend. The next section derives the special cases, including a new method of exponential smoothing. Next, we show how each special case method can be justified by an underlying RCSS model. Finally, we demonstrate the frequency with which special cases occur in the time series from the M3 competition (Makridakis and Hibon, 2000).

The damped trend method and its special cases

Following the notation of Hyndman *et al* (2008), the damped trend method can be written in several different forms. The original recurrence form (Gardner and McKenzie, 1985) is written:

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \qquad (1)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1} \qquad (2)$$

$$\hat{y}_{t+h\,|\,t} = \ell_t + (\phi + \phi^2 + \ldots + \phi^h) b_t \qquad (3)$$

where ℓ_t is the level and b_t is the trend. The smoothing parameters for level and trend are α and β, while ϕ is the damping or autoregressive parameter.

Equations (1) and (2) can be rewritten in the simpler error-correction form:

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha e_t \qquad (4)$$

$$b_t = \phi b_{t-1} + \alpha\beta e_t \qquad (5)$$

where e_t is the one-step-ahead error. It appears that (1) and (2) always produce the same level and trend components as (4) and (5), but this is not true when $\alpha = 0$. The difficulty lies in (5), which contains the product $\alpha\beta$; when $\alpha = 0$, the optimal value of β cannot be determined, and the recurrence and error-correction forms of the method are not equivalent. Some forecasters simply drop the α parameter in (5) and smooth the trend component using β only, but again we lose equivalence to the recurrence form. In the results below, we use the recurrence form of the method to avoid these problems.

When all parameters are selected from the [0, 1] interval, at least eleven different methods can be defined. The damped trend itself is defined by optimal parameters in the ranges $0 \le \alpha \le 1$, $0 < \beta \le 1$, and $0 < \phi < 1$. Another well-known method occurs with the same α and β ranges and $\phi = 1$; there is no damping of the trend component and the method is Holt. An interesting variation on the Holt method occurs when we allow $\alpha = 1$, with $0 < \beta < 1$ and $\phi = 1$, a method sometimes called the smoothed trend method, although for the sake of simplicity we counted it as the Holt method.

Three versions of simple exponential smoothing (SES) can be obtained. When $\phi = \beta = 0$ and $0 < \alpha < 1$, there is no trend and the method is standard SES. When $0 < \alpha < 1$, $\beta = 0$, and $\phi = 1$, the method becomes SES with drift, as discussed in Hyndman and Billah (2003):

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} + b \tag{6}$$

$$\hat{y}_{t+h\,|\,t} = \ell_t + hb \tag{7}$$

With the same α and β parameters and $0 < \phi < 1$, we have a new method, SES with damped drift:

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} + \phi b \quad (8)$$

$$\hat{y}_{t+h\,|\,t} = \ell_t + (\phi + \phi^2 + \ldots + \phi^h)b \quad (9)$$

Three versions of the random walk are possible. When $\alpha = 1$ and $\phi = \beta = 0$, the method is the standard random walk. When $\alpha = 1$, $\beta = 0$, and $\phi = 1$, the method is a random walk with drift. With the same α and β parameters and $0 < \phi < 1$, we have another new method, a random walk with damped drift.

Finally, with $\alpha = 0$ and $\beta = 0$, three deterministic methods are possible depending on the value of ϕ. If $0 < \phi < 1$, the method is a deterministic modified exponential trend. If $\phi = 1$, the method is a deterministic linear trend because parameter optimization does not change the initial values of level and trend. Finally, if $\phi = 0$, the method reduces to a simple average of the data in the fit periods.

Random coefficient models that underlie the special cases

Although the state space models of Hyndman *et al* (2008) provide a theoretical rationale for exponential smoothing methods, we prefer the RCSS models of McKenzie and Gardner (2009) on the grounds that they are more realistic. Hyndman *et al* (2008) show that damped trend exponential smoothing is optimal for a single source of error state space model with constant coefficients:

$$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t \quad (10)$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + h_1 \varepsilon_t \quad (11)$$

$$b_t = \phi b_{t-1} + h_2 \varepsilon_t \quad (12)$$

In this model, $h_1 = \alpha$ and $h_2 = \alpha\beta$. In McKenzie and Gardner (2009), we demonstrate that the approach is also optimal for an RCSS model of the form

$$y_t = \ell_{t-1} + A_t b_{t-1} + v_t \tag{13}$$

$$\ell_t = \ell_{t-1} + A_t b_{t-1} + h_1^* v_t \tag{14}$$

$$b_t = A_t b_{t-1} + h_2^* v_t \tag{15}$$

where $\{A_t\}$ is a sequence of independent, identically distributed binary (0,1) random variates with $P(A_t = 1) = \phi$. The white noise innovations processes of these two models, (10-12) and (13-15), are different, while the RCSS coefficients (h_1^*, h_2^*) are related to, but usually different from those of the constant coefficient model (h_1, h_2). The RCSS model has two advantages as an underlying model for the damped trend. First, the gradient revision equation (15) allows both smooth and sudden changes of gradient. When sudden changes occur, $A_t = 0$, which yields consecutive runs of different linear trends. The second advantage is that the parameter ϕ can be interpreted directly as a measure of the persistence of these different linear trends.

Now consider the three damped models other than the standard damped trend:

SES with damped drift: $0 < \alpha < 1,\ \beta = 0: \Rightarrow h_1 = \alpha,\ h_2 = 0$

Random Walk with damped drift: $\alpha = 1,\ \beta = 0: \Rightarrow h_1 = 1,\ h_2 = 0$

Modified exponential trend: $\alpha = 0,\ \beta = 0: \Rightarrow h_1 = 0,\ h_2 = 0$

In all three models, $h_1^* = h_1$ and $h_2^* = h_2$, so the corresponding forms of (13-15) are easily derived. All three models have gradient revision equations of the form $b_t = A_t b_{t-1}$, so there is an initial linear trend of constant gradient which changes to zero gradient at a random time. Such behavior is reflected in the series plotted in Figure 1, and in many other series in the M3 data base.

Furthermore, for all three models we can specify an alternative gradient revision equation of the form $b_t = A_t b_{t-1} + (1 - A_t) d_t$ where $\{d_t\}$ is a sequence of independent, identically distributed random variates of zero mean. Use of this form implies a gradient which remains the same as long as $A_t = 1$, and changes suddenly to another (non-zero) value when $A_t = 0$. Thus, we would get consecutive runs of linear trend , each with a distinct gradient, and each of random length. Such a class of models forms a natural generalization of the other special cases identified here, for example the single linear trend of constant gradient and the random walk with globally constant drift.

The special cases demonstrated

To demonstrate the special cases, we used the 3,003 series from the M3 competition (Makridakis and Hibon, 2000). The damped trend method was fitted after holding out the last 6, 8, and 18 observations for *ex ante* testing in the annual, quarterly, and monthly series, respectively. There is also a group of "other" series for which no sampling frequency was given and for which the last 6 observations were held out. The series were deseasonalized using multiplicative seasonal indices computed from data in the fit periods. To obtain initial values for level and trend, we used two common procedures. First, *local* initial values were computed by fitting an OLS regression on time to the first five observations in the fit periods. Because many of the special cases include a fixed drift or trend component, we also tested *global* initial values, computed by extending the regression to include all observations in the fit periods. For each set of initial values, the Excel Solver was applied to find the parameter set from the [0, 1] interval that minimized the sum of squared errors in the fit periods.

Tables 1 and 2 summarize the methods identified using local and global initial values, respectively. There are some surprising findings in both tables. Either a drift or a smoothed trend component was identified in about 99% of the series for both local and global initial values. Methods with a drift component were identified in about 38% of the series using local initial values and 53% with global. The drift or trend component was usually damped, which happened in 84% of the series using local initial values and in 70% with global.

(Insert Tables 1 and 2 here)

The damped trend method itself was identified in only 43% of the series with local initial values and 28% with global. Notice that the frequency of identification of the damped trend increased with sampling frequency in both tables. The most common special case of the damped trend was SES with damped drift, which occurred in almost a quarter of the series for both types of initial values. This method describes a fixed early trend that gradually dies out, behavior that may seem strange, but is actually quite common in the M3 series; an example for one of the annual series is given in Figure 1.

(Insert Figure 1 here)

In Gardner and McKenzie (1985), we hypothesized that the damped trend would often reduce to SES, but this method was identified in less than 1% of the series with both types of initial values. We hypothesized that the damped trend would often reduce to the Holt method, but this happened in only 10% of the series with local initial values and 2% with global. We also thought that the standard random walk method would be identified with some frequency, but this happened not at all with local initial values and in only 0.1% of the monthly series using global initial values. However, we did find that the random walk with damped drift was a fairly common special case using both types of initial values.

Our forecast accuracy results are not presented in detail here because they are not significantly different from the results reported for the Makridakis and Hibon implementation of the damped trend, which used the recurrence form in (1) - (3), with backcasting to obtain initial values. Over all series and forecast horizons, Makridakis and Hibon reported a mean symmetric absolute percentage error of 13.6%, compared to 13.5% for our implementation with local initial values, and 13.8% with global initial values.

Conclusions

In an evaluation of progress in forecasting over the last 25 years, Armstrong and Fildes (2006) concluded that the diffusion of useful methods such as the damped trend has been slow. Many textbooks continue to ignore the damped trend despite it having been shown to improve accuracy in multiple hypotheses studies since 1985. Software companies have been slow to adopt methods that should improve accuracy, and few software programs include the damped trend as an extrapolative option. We hope that the findings in this paper will encourage adoption of the damped trend. In particular, we note that damping was necessary in 84% of the M3 series using local initial values and in 70% when global values were used. The interpretation of this damping in terms of the RCSS models serves to emphasize the continuing importance of

Brown's (1963) advice that our forecasts must attempt to capture local behavior and allow both smooth and sudden changes to occur.

One explanation for the empirical success of the damped trend is the flexibility of the method, which adapts to the time series by automatically selecting from a variety of special cases during the fitting procedure. Each special case method can be justified by an underlying RCSS model. If forecasters wish to exclude any of the special cases, this is easily accomplished by constraining the parameters of the method.

Perhaps the special cases of the damped trend are obvious, but what is surprising is the frequency with which they are selected. Most of the time, fitting the damped trend produces a special case rather than the damped trend itself. An unusual special case is a new variant of exponential smoothing, SES with a damped drift term. This may seem an unlikely method, but in our opinion it is no more unlikely than any of the other time series methods that contain a fixed drift. Given that SES with damped drift was identified so often in the M3 series, this method should receive some consideration in both empirical research and practice.

References

Armstrong, JS and Fildes R (2006). Making progress in forecasting, *Int J Forecasting* **22**: 433-441.

Brown, R.G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series,* Englewood Cliffs, NJ: Prentice-Hall.

Fildes R (2001). Beyond forecasting competitions, *Int J Forecasting* **17**: 556-560.

Fildes R, Nikolopous K, Crone S, and Syntetos A (2008). Forecasting and operational research: a review. *J Opl Res Soc* **59**: 1-23.

Flores, BE and Pearce, SL (2000). The use of an expert system in the M3 competition, *Int J Forecasting* **16**: 485-496.

Gardner ES and McKenzie E (1985). Forecasting trends in time series. *Mngt Sci* **31**: 1237–1246.

Gardner ES and McKenzie E (1988). Model identification in exponential smoothing, *J Opl Res Soc* **39**: 863-867.

Gardner ES (2006). Exponential smoothing: The state of the art – Part II. *Intl J Forecasting* **22**: 637-666.

Hyndman RJ, Koehler AB, Ord JK and Snyder RD (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer: Berlin.

Hyndman RJ and Billah B (2003). Unmasking the Theta method. *Intl J Forecasting* **19**: 287-290.

Makridakis S and Hibon M (2000). The M3-Competition: Results, conclusions and implications. *Intl J Forecasting* **16**: 451-476.

McKenzie E and Gardner ES (2009). Damped trend exponential smoothing: A modelling viewpoint. *Intl J Forecasting*: forthcoming.

Figure 1.

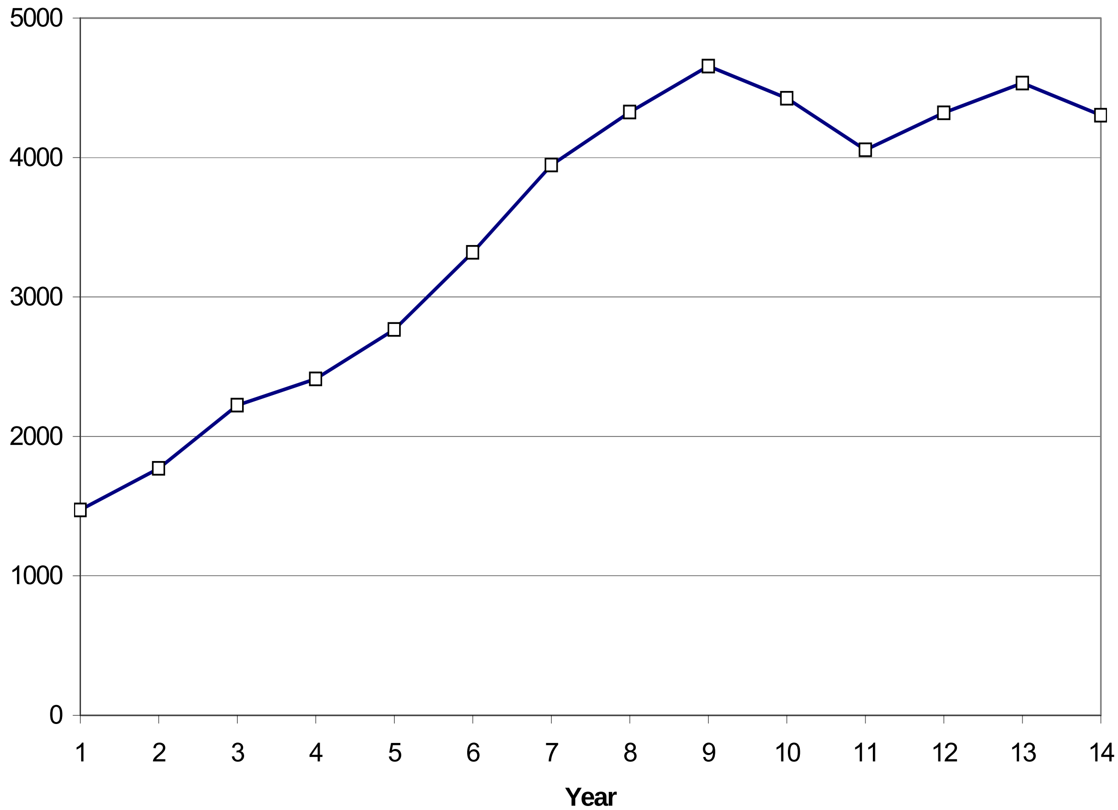

Table 1.

	Parameter values				Percent of series				
Case	Level	Trend	Damping	Method	Ann.	Qtr.	Mon.	Other	All
1	$0 \leq \alpha \leq 1$	$0 < \beta \leq 1$	$0 < \phi < 1$	Damped trend	25.9	47.1	47.5	51.1	43.0
2	$0 \leq \alpha \leq 1$	$0 < \beta \leq 1$	1	Holt	17.4	14.2	3.6	17.2	10.0
3	$0 < \alpha < 1$	0	$0 < \phi < 1$	SES with damped drift	17.7	16.7	33.6	14.4	24.8
4	$0 < \alpha < 1$	0	1	SES with drift	3.6	3.7	1.1	2.3	2.4
5	$0 < \alpha < 1$	0	0	SES	0.2	0.4	1.5	0.0	0.8
6	1	0	$0 < \phi < 1$	Random walk with damped drift	18.3	9.0	1.9	12.6	7.8
7	1	0	1	Random walk with drift	7.8	2.1	0.4	2.3	2.5
8	1	0	0	Random walk	0.0	0.0	0.0	0.0	0.0
9	0	0	$0 < \phi < 1$	Modified exponential trend	9.1	6.0	10.1	0.0	8.3
10	0	0	1	Linear trend	0.2	0.1	0.1	0.0	0.1
11	0	0	0	Simple average	0.0	0.8	0.2	0.0	0.3
				Total	100.0	100.0	100.0	100.0	100.0

Table 2.

	Parameter values				Percent of series				
Case	Level	Trend	Damping	Method	Ann.	Qtr.	Mon.	Other	All
1	$0 \le \alpha \le 1$	$0 < \beta \le 1$	$0 < \phi < 1$	Damped trend	11.8	32.0	32.6	29.3	27.8
2	$0 \le \alpha \le 1$	$0 < \beta \le 1$	1	Holt	2.9	3.2	0.8	0.0	1.8
3	$0 < \alpha < 1$	0	$0 < \phi < 1$	SES with damped drift	15.8	18.3	30.5	17.8	23.5
4	$0 < \alpha < 1$	0	1	SES with drift	7.1	17.3	10.4	13.2	11.6
5	$0 < \alpha < 1$	0	0	SES	0.0	0.4	1.0	0.0	0.6
6	1	0	$0 < \phi < 1$	Random walk with damped drift	21.4	10.1	2.8	19.0	9.6
7	1	0	1	Random walk with drift	24.3	6.3	0.9	20.1	8.4
8	1	0	0	Random walk	0.0	0.0	0.1	0.0	0.0
9	0	0	$0 < \phi < 1$	Modified exponential trend	8.1	5.7	11.7	0.0	8.7
10	0	0	1	Linear trend	8.5	6.7	9.2	0.6	7.9
11	0	0	0	Simple average	0.0	0.0	0.0	0.0	0.0
				Total	100.0	100.0	100.0	100.0	100.0

Captions for Figures and Tables

Figure 1. Fit periods for M3 annual series YB067, an apt series for a method with a damped drift term.

Table 1. Methods identified in the M3 series using *local* initial values.

Table 2. Methods identified in the M3 series using *global* initial values.