Automatic method selection in exponential smoothing

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

January 16, 2009

Automatic method selection in exponential smoothing

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 model-fitting. 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. In most special cases, the optimal smoothing parameter for trend was zero, which produces a method with a drift term; the most common special case was a new method, simple exponential smoothing with a damped drift term.

Key words: Time series—exponential smoothing, identification, method selection

Automatic method selection in exponential smoothing

1. Introduction

In exponential smoothing, it is common to apply the damped trend method to every time series. Numerous 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 often result in simpler methods than the damped trend, they have failed to produce better average 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 yet to be demonstrated and it is not clear how one should proceed.

The failure of method selection procedures to improve forecast accuracy is frustrating, but what has been overlooked is that the damped trend method contains a variety of special cases, ranging from a random walk to a deterministic trend. Fitting the method is actually a means of automatic selection from these special cases. The next section derives the special cases, including a new method of exponential smoothing. In the following section, we demonstrate the frequency with which special cases occur in the time series from the M3 competition (Makridakis and Hibon, 2000).

2. The damped trend method and its special cases

Following the standard notation of Hyndman et al. (2008), the recurrence form of the damped trend method is written:

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

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

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

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:

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

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

where e_t is the one-step-ahead error. In the trend equation (5), some forecasters delete the α parameter and smooth the error using an independent parameter β. However, we prefer to use equation (5) as shown to preserve the equivalence between recurrence and error-correction forms, both of which are commonly used in practice.

Depending on the parameter combinations chosen from the [0, 1] interval, the method described by equations (3) – (5) has numerous equivalent ARIMA models (see Gardner and McKenzie, 1988, and Gardner, 2006). Depending on the assumptions regarding the properties of the errors, the method also has numerous equivalent state space models that go beyond the scope of the ARIMA class, as discussed in Chatfield et al. (2001), Koehler et al. (2001), Hyndman et al. (2002), and McKenzie and Gardner (2008). Here we discuss only the most common *methods* that are special cases of (3) – (5).

In the empirical work below, the damped trend method is defined by the following parameter ranges: $0 \le \alpha \le 1$, $0 \le \beta \le 1$, and $0 < \phi < 1$. There are at least ten special cases of the method. The best-known special case occurs when $0 < \alpha \le 1$, $0 < \beta \le 1$, 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 from the damped trend method. 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). With the same α and β parameters and $0 < \phi < 1$, we have a new method, SES with damped drift.

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.

In equation (5), when $\alpha = 0$, three deterministic methods are possible depending on the value of ϕ. If $\phi = 1$, the method is a deterministic linear trend because parameter optimization does not change the initial values of level and trend. If $0 < \phi < 1$, the method is a deterministic modified exponential trend. Finally, if $\phi = 0$, the method reduces to a simple average of the fit data.

3. The special cases demonstrated

To demonstrate the special cases, we used the 3,003 series from the M3 competition. The damped trend method in equations (3) - (5) 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. We tested two common procedures for computing initial values for level and trend: *local* initial values were computed by fitting an OLS regression on time to the first five observations in the fit periods, and *global* initial values were 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 mean squared error in the fit periods.

Tables 1 and 2 summarize the methods identified in the M-competition series 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.

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

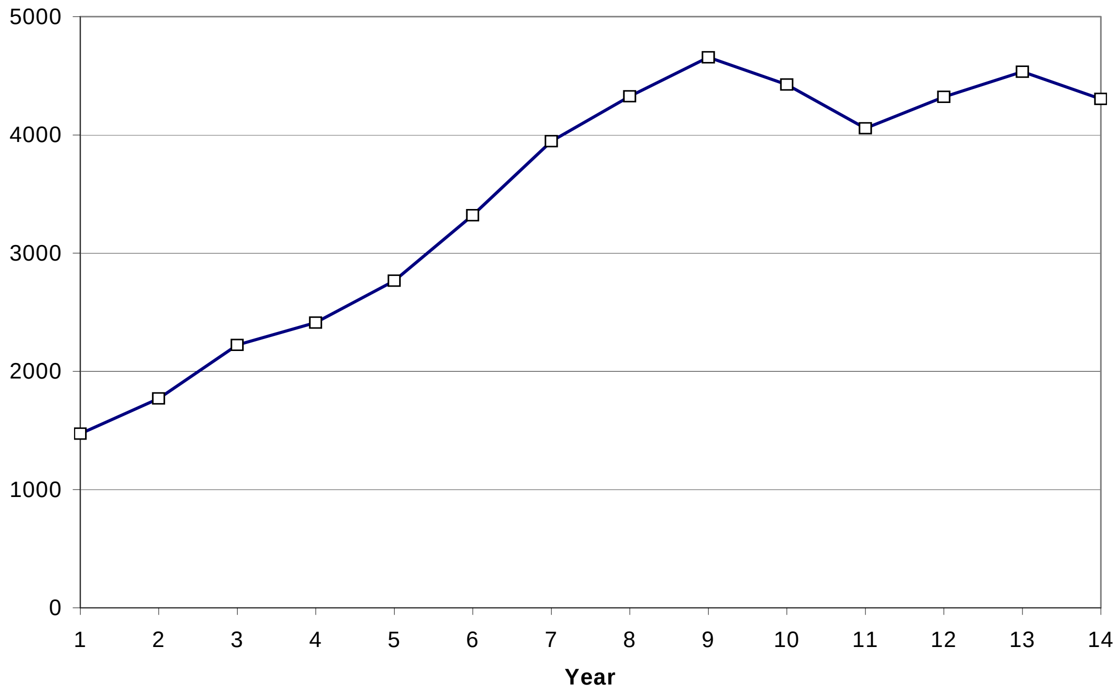
	Parameter values				Percent of series				
Case	Level	Trend	Damping	Method	Ann.	Qtr.	Mon.	Other	All
1	$0 < \alpha \leq 1$	$0 < \gamma \leq 1$	$0 < \phi < 1$	Damped trend	25.9	47.1	47.5	51.1	43.0
2	$0 < \alpha \leq 1$	$0 < \gamma \leq 1$	1	Holt	17.4	14.2	3.6	17.2	10.0
3	$0 < \alpha < 1$	0	$0 < \phi < 1$	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	NA	$0 < \phi < 1$	Modified exponential trend	9.1	6.0	10.1	0.0	8.3
10	0	NA	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. Methods identified in the M3 series using *global* initial values.

	Parameter values				Percent of series				
Case	Level	Trend	Damping	Method	Ann.	Qtr.	Mon.	Other	All
1	$0 < \alpha \leq 1$	$0 < \gamma \leq 1$	$0 < \phi < 1$	Damped trend	11.8	32.0	32.6	29.3	27.8
2	$0 < \alpha \leq 1$	$0 < \gamma \leq 1$	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	NA	$0 < \phi < 1$	Modified exponential trend	8.1	5.7	11.7	0.0	8.7
10	0	NA	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

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.

Figure 1. Fit periods for M3 annual series YB067



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.

How do our forecast accuracy results compare to the Makridakis and Hibon implementation of the damped trend? Table 3 summarizes the average symmetric MAPE for Makridakis and Hibon's "dampen" method and the damped trend fitted with local and global initial values. The comparisons vary somewhat by type of data, but for the average over all data and horizons, there is no significant difference in accuracy between the dampen method and the damped trend with local initial values. We also note that, except in the quarterly data, local initial values produced better accuracy than global.

Table 3. Average symmetric APE by forecast horizon

Data	Nbr. series	Method	Initial values	Forecast horizon 1	2	3	4	5	6	8	12	15	18	All
Ann.	645	MH Dampen		**8.0**	12.4	17.0	19.3	22.3	24.0					17.2
		Damped trend	Local	8.1	**11.9**	**16.4**	**18.9**	**21.7**	**23.5**					**16.7**
		Damped trend	Global	8.4	12.4	17.1	19.4	22.0	23.7					17.2
Qtr.	756	MH Dampen		5.1	6.8	7.7	9.1	**9.7**	11.3	**12.8**				9.3
		Damped trend	Local	5.0	6.7	7.8	9.4	10.1	11.6	13.5				9.6
		Damped trend	Global	**4.8**	**6.5**	**7.5**	**9.0**	**9.7**	**11.0**	**12.8**				**9.2**
Mon.	1,428	MH Dampen		11.9	11.4	13.0	14.2	12.9	12.6	**13.0**	13.9	17.5	**18.9**	14.6
		Damped trend	Local	11.9	**10.9**	12.7	13.5	**12.3**	**12.4**	13.5	**13.8**	**17.2**	**18.9**	**14.5**
		Damped trend	Global	**11.6**	11.1	**12.2**	**13.1**	12.8	12.8	13.4	14.0	18.0	19.8	14.8
Other	174	MH Dampen		**1.8**	**2.7**	3.9	4.7	5.8	5.4	6.6				4.6
		Damped trend	Local	**1.8**	**2.7**	**3.7**	**4.4**	**5.5**	**5.1**	**6.2**				**4.4**
		Damped trend	Global	**1.8**	2.8	3.9	4.6	5.7	5.3	6.5				4.6
All	3,003	MH Dampen		8.8	10.0	12.0	13.5	13.7	14.3	**12.5**	13.9	17.5	**18.9**	13.6
		Damped trend	Local	8.7	**9.6**	11.7	13.1	**13.4**	**14.2**	12.9	**13.8**	**17.2**	**18.9**	**13.5**
		Damped trend	Global	**8.6**	9.7	**11.6**	**12.9**	13.6	**14.2**	12.7	14.0	18.0	19.8	13.8

4. Conclusions

The damped trend method of exponential smoothing has performed well in empirical studies of forecast accuracy. One explanation for this performance is the flexibility of the method, which contains a variety of special cases that are automatically selected during model-fitting. If for some reason the forecaster wishes to avoid any particular special case, the method parameters must be constrained accordingly. To avoid all special cases, that is to guarantee that the forecasts will be made with the original damped trend method, all parameters must be kept off the 0-1 boundaries.

The trend smoothing parameter is especially problematic and was fitted at zero about half the time using local initial values in the M3 series; with global initial values, the trend parameter turned out to be zero about 70% of the time. When the trend parameter is zero, the method almost always includes a drift term, which may be fixed or damped.

This paper introduces a new variant of exponential smoothing, SES with a damped drift term. This may seem an unlikely method, but in my opinion it is no more unlikely than any of the other time series methods that contain a fixed drift term. Given that SES with damped drift was identified so often in the M3 series, this method should receive some consideration in empirical research.

References

Chatfield, C., Koehler, A.B., Ord, J.K., & Snyder, R.D. (2001). A new look at models for exponential smoothing, *Journal of the Royal Statistical Society*, *Series D*, 50, 147-159.

Fildes, R. (2001). Beyond forecasting competitions, *International Journal of Forecasting*, 17, 556-560.

Flores, B.E. & Pearce, S.L. (2000). The use of an expert system in the M3 competition, International Journal of Forecasting, 16, 485-496.

Gardner Jr., E. S. (2006). Exponential smoothing: The state of the art – Part II. *International Journal of Forecasting*, 22, 637-666.

Gardner Jr., E. S., & 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.

Hyndman, R., Koehler, A., Ord, J.K., & Snyder, R.D. (2008) Forecasting with exponential smoothing: The state space approach, Springer-Verlag: Berlin.

Hyndman, R.J., & Billah, B. (2003). Unmasking the Theta method, *International Journal of Forecasting*, 19, 287-290.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, 18, 439-454.

Koehler, A.B., Snyder, R.D., & Ord, J.K. (2001). Forecasting models and prediction intervals for the multiplicative Holt-Winters method, *International Journal of Forecasting*, 17, 269-286.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

McKenzie, E., & Gardner Jr., E. S. (2008). Damped trend exponential smoothing: A modelling viewpoint. Working paper available from the authors.