Exponential smoothing in the telecommunications data

Everette S. Gardner, Jr.*, Joaquin Diaz-Saiz

University of Houston, Bauer College of Business, Houston, TX 77204-6021

*Corresponding author: Tel. 713-743-4744, Fax 713-743-4940

E-mail addresses: egardner@uh.edu (E. Gardner, Jr.), jdiaz-saiz@uh.edu (J. Diaz-Saiz)

Revised August 7, 2006

Exponential smoothing in the telecommunications data

Exponential smoothing in the telecommunications data

Abstract

Exponential smoothing methods gave poor forecast accuracy in Fildes et al.'s study of telecommunications time series. We reexamine this study and show that parameter optimization improves the accuracy of the Holt and damped trend methods. Further improvement occurs when the time series are trimmed to eliminate irrelevant early data, and when the methods are fitted to minimize the MAD rather than the MSE. Contrary to Fildes et al., we show that the damped trend is more accurate than Holt's method. Because most of the telecommunications series display steady trends, we test the Theta method of forecasting and its derivative, simple exponential smoothing with drift. The Theta method proves disappointing, but simple exponential smoothing with drift is the best smoothing method for this data, giving about the same accuracy as the robust trend.

Key words

Comparative methods – evaluation; Time series – exponential smoothing, robust trend, theta method

Exponential smoothing in the telecommunications data

Introduction

With only a few exceptions, exponential smoothing has performed well in numerous empirical studies of forecast accuracy (Gardner, 2006). Perhaps the most notable exception is the study of telecommunications data by Fildes et al. (1998), who found that the robust trend method was more accurate than Holt's additive trend or the Gardner-McKenzie (1985) damped-additive trend. Fildes et al. also found that Holt's method was more accurate than the damped trend, a conclusion so surprising that Armstrong (2006) recommended that a replication be performed.

We attempt to replicate the exponential smoothing results in Fildes et al. We also test several ideas to improve the accuracy of the Holt and damped trend methods, and we test two additional smoothing methods that should be better suited to the data, the Theta method of forecasting (Assimakopoulos & Nikolopoulos, 2000) and its derivative, simple exponential smoothing (SES) with drift (Hyndman & Billah, 2003).

In the next few sections, we review the characteristics of the telecommunications series and give brief explanations of the forecasting methods. Next, we explain how the methods were fitted. Finally, we present new empirical comparisons of the exponential smoothing methods and the robust trend.

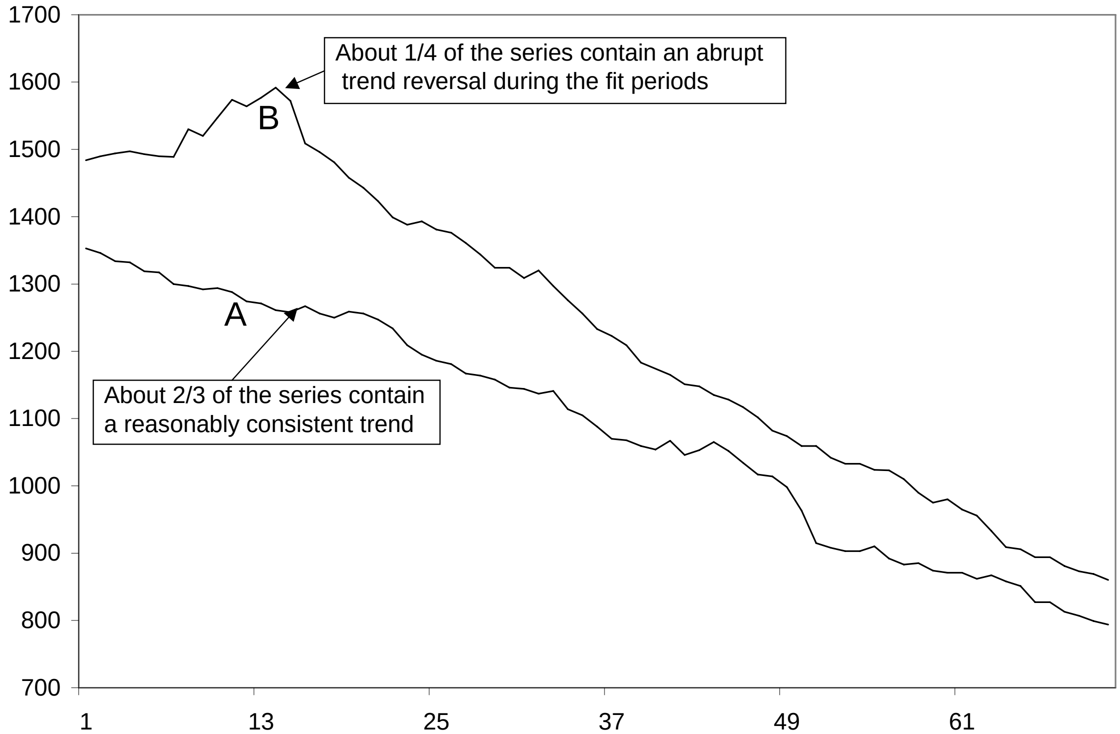
The telecommunications series and the robust trend

There are 261 telecommunications series, each with 71 monthly, nonseasonal observations of the number of a particular type of telephone circuit in service by locality within a single U.S. state. Compared to the series used in the M1 competition, Fildes et al. claim that the telecommunications series are much more homogeneous. We agree. Although outliers contaminate nearly every series, about two thirds of them are not especially difficult to forecast because they display steady downward trends, like Series A in Figure 1 (next page).

The remaining series are more challenging. In about a quarter of the series, an abrupt trend reversal occurs in the early fit periods. An example is given in Series B where the data have a positive slope for the first 14 periods, with a negative slope thereafter. This kind of behavior confounds estimation of the trend component in any exponential smoothing method. Domain knowledge for the telecommunications data, discussed in Fildes (1989, 1992), calls for a negative slope in the forecast periods, which begin at period 24. Thus we should expect to improve forecast accuracy by trimming the fit periods to delete irrelevant early data, although this was not done in Fildes et al. The remaining series (about 25) are characterized by jump shifts in level and trend and other kinds of foul discontinuities that have a major impact on average forecast accuracy for all series.

For series characterized by consistent trends with outliers, Grambsch and Stahel (1990) developed the robust trend method, easily the best method tested in Fildes et al.'s study. The model that underlies the robust trend is a random walk, or an ARIMA (0, 1, 0), with drift. The method aims at robustness by estimating the drift as the median rather than the mean of the differenced data, subject to some complex adjustments (see Grambsch and Stahel for details).

Figure 1. Examples of the telecommunications series



SES with drift

Fildes et al. tested two exponential smoothing methods, Holt's additive trend and the damped-additive trend. Given the steady trends in most of the telecommunications series, another method, SES with drift, seems more appropriate. The idea for SES with drift originated in the "Theta" method of forecasting by Assimakopoulos and Nikolopoulos (2000). In the M3 competition (Makridakis & Hibon, 2000), the Theta method performed well, although the authors' description of the method is complex. Hyndman and Billah (2003) demonstrated that the Theta method is overly complex because the same forecasts can be obtained by using SES with a fixed drift term equal to half the slope of a straight line fitted to the data.

Hyndman and Billah derive several equivalent forms of the SES with drift method. Using their notation, the simplest form is as follows:

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

$$\hat{X}_t(h) = \ell_t + hb \, , \quad (2)$$

where ℓ is the level, b is the drift, and $\hat{X}_t(h)$ is the h-step forecast. Hyndman and Billah argue that the drift term should be optimized in equations (1) and (2), rather than fixed at a predetermined value like the Theta method. Both alternatives are tested below.

Model-fitting

For each series, Fildes et al. made forecasts through 18 steps ahead, using five irregularly spaced time origins at months 23, 31, 38, 45, and 53. At the same origins, we used the Excel Solver to fit the Holt, damped trend, SES with drift, and Theta methods. For each method, we compared MSE and MAD fit criteria. The MSE has been used in almost all empirical research in exponential smoothing, including Fildes et al., although Gardner (1999) showed that the MAD criterion often produces better *ex ante* accuracy in series contaminated by outliers.

The intercept and slope of a straight line fitted to the data were used to initialize all methods. In the Holt and damped trend methods, we compared optimization of parameters alone to simultaneous optimization of parameters and initial values. In SES with drift, we optimized the initial level and drift simultaneously with the smoothing parameter. In the Theta method, we optimized the initial level and smoothing parameter simultaneously, while keeping the drift component fixed at half the slope of the fit data. For all methods, we compared parameter selection from the usual [0,1] interval to selection from the complete range of invertibility of the underlying ARIMA model. This was done because we found that the optimal level parameter from the [0,1] interval was frequently equal to 1.0 for all methods.

Finally, we compared the use of two sets of data to fit each method. In the first fit, we used all data through the forecast origins. In the second fit, we trimmed the data by discarding any observations prior to an early trend reversal. This was done by dividing the first set of fit periods roughly in half. We compared the slope of the first 12 observations to the next 11; if the slope changed from positive to negative, we simply started the fit at the maximum observation value during the first 23 periods.

Accuracy comparisons

MAPE results are given in Table 1, along with Fildes et al.'s original results for the robust trend, Holt, and damped trend methods. Fildes et al. also tested ARIMA and ARARMA methods, but the results are not repeated here because these methods performed poorly. Our results are voluminous, and we made several decisions to reduce the size of Table 1. For all methods, we found little difference in forecast accuracy between optimization of parameters over the [0,1] interval vs. the complete range of invertibility, so only the first option is reported. In the Holt and damped trend methods, we found little difference between optimization of parameters alone vs. simultaneous optimization of parameters and initial values, so only the first option is reported to make the results as comparable as possible to Fildes et al. (who did not optimize initial values). Finally, we do not report median APEs because the differences among methods are similar to MAPE comparisons.

At all forecast origins, our damped trend results are better than those reported by Fildes et al. For example, at origin 23, Fildes et al. reported a damped trend MAPE of 9.25% for the average of horizons 1-18. This result is found in the Damped 1 row of the first section in Table 1. Using an MSE criterion to fit the original data (Damped 2), we reduced the MAPE to 7.31%. If we continue with the MSE and trim the irrelevant early data (Damped 3), the MAPE falls to 6.32%. If we minimize the MAD and trim the data (Damped 4), the MAPE falls to 5.71%. Although the improvements in damped trend results vary somewhat by forecast origin and horizon, they are consistent.

Table 1. MAPEs for the telecommunications data (261 series)

Origin	Method	Fit data	Fit criterion	Horizon 1	6	12	18	1-6	1-12	1-18
23	Robust trend	Original	MSE	**0.68**	**3.26**	6.09	10.27	**2.06**	**3.48**	5.14
	Holt 1 (Fildes et al.)	Original	MSE	1.46	6.79	11.30	16.20	4.43	6.99	9.37
	Holt 2	Original	MSE	1.94	5.76	9.42	14.11	4.05	6.01	8.04
	Holt 3	Trimmed	MSE	1.83	5.08	8.47	12.76	3.62	5.37	7.24
	Holt 4	Trimmed	MAD	1.00	4.39	8.27	12.87	2.73	4.75	6.82
	Damped 1 (Fildes et al.)	Original	MSE	1.25	6.46	11.05	17.14	4.27	6.67	9.25
	Damped 2	Original	MSE	1.97	5.19	8.42	12.96	3.85	5.45	7.31
	Damped 3	Trimmed	MSE	1.76	4.40	7.31	11.38	3.21	4.66	6.32
	Damped 4	Trimmed	MAD	1.08	3.63	6.77	11.16	2.39	3.92	5.71
	SES with drift 1	Original	MSE	1.67	4.11	7.13	11.31	2.96	4.42	6.13
	SES with drift 2	Trimmed	MSE	1.78	4.02	6.77	10.41	2.97	4.32	5.83
	SES with drift 3	Trimmed	MAD	0.92	3.29	**5.88**	**9.59**	2.18	3.52	**5.02**
	Theta 1	Original	MSE	0.91	3.68	7.24	12.54	2.32	3.93	6.01
	Theta 2	Trimmed	MSE	1.82	4.05	6.95	10.75	3.03	4.39	5.97
	Theta 3	Trimmed	MAD	0.88	3.61	7.07	11.29	2.34	4.02	5.85
31	Robust trend	Original	MSE	1.65	4.19	8.40	12.80	2.97	4.83	6.88
	Holt 1 (Fildes et al.)	Original	MSE	1.79	5.30	10.77	16.66	3.59	6.09	8.80
	Holt 2	Original	MSE	1.72	5.02	10.03	15.13	3.38	5.68	8.11
	Holt 3	Trimmed	MSE	1.71	4.93	9.84	14.88	3.34	5.58	7.99
	Holt 4	Trimmed	MAD	1.72	4.99	9.81	14.61	3.39	5.62	7.96
	Damped 1 (Fildes et al.)	Original	MSE	1.97	6.36	13.23	20.53	4.15	7.25	10.65
	Damped 2	Original	MSE	1.78	5.17	10.54	15.85	3.48	5.90	8.50
	Damped 3	Trimmed	MSE	1.74	4.82	9.77	14.69	3.32	5.53	7.92
	Damped 4	Trimmed	MAD	1.71	4.66	9.64	15.09	3.22	5.40	7.89
	SES with drift 1	Original	MSE	1.75	4.77	9.55	14.20	3.30	5.43	7.73
	SES with drift 2	Trimmed	MSE	1.69	4.45	8.94	13.33	3.12	5.12	7.27
	SES with drift 3	Trimmed	MAD	**1.63**	**4.13**	**8.25**	**12.30**	**2.93**	**4.74**	**6.71**
	Theta 1	Original	MSE	1.49	5.86	11.40	17.77	3.65	6.27	9.24
	Theta 2	Trimmed	MSE	1.68	4.73	9.45	14.20	3.22	5.37	7.66
	Theta 3	Trimmed	MAD	1.69	4.74	9.49	14.22	3.24	5.40	7.68

Table 1 (continued)

Origin	Method	Fit data	Fit criterion	Horizon 1	6	12	18	1-6	1-12	1-18
38	Robust trend	Original	MSE	1.25	4.32	8.15	12.11	2.74	4.56	6.56
	Holt 1 (Fildes et al.)	Original	MSE	**1.19**	**4.10**	**7.78**	**11.91**	**2.67**	**4.40**	**6.31**
	Holt 2	Original	MSE	1.35	4.75	9.10	13.78	3.06	5.11	7.37
	Holt 3	Trimmed	MSE	1.37	4.83	9.13	13.90	3.11	5.16	7.42
	Holt 4	Trimmed	MAD	1.32	4.52	8.48	12.77	2.95	4.83	6.91
	Damped 1 (Fildes et al.)	Original	MSE	1.34	5.74	11.84	18.83	3.45	5.25	9.51
	Damped 2	Original	MSE	1.33	4.92	9.63	14.84	3.08	5.30	7.78
	Damped 3	Trimmed	MSE	1.33	4.71	9.01	13.84	2.99	5.03	7.33
	Damped 4	Trimmed	MAD	1.31	4.70	9.01	13.77	2.97	5.01	7.30
	SES with drift 1	Original	MSE	1.39	4.82	9.32	14.26	3.09	5.18	7.56
	SES with drift 2	Trimmed	MSE	1.35	4.64	8.88	13.62	2.99	4.98	7.24
	SES with drift 3	Trimmed	MAD	1.30	4.34	8.20	12.11	2.79	4.62	6.61
	Theta 1	Original	MSE	1.49	5.86	11.40	17.77	3.65	6.27	9.24
	Theta 2	Trimmed	MSE	1.43	5.14	9.93	14.92	3.26	5.53	8.02
	Theta 3	Trimmed	MAD	1.41	5.12	9.91	14.89	3.24	5.51	7.99
45	Robust trend	Original	MSE	**1.02**	**3.80**	**7.35**	**12.26**	**2.45**	**4.25**	**6.14**
	Holt 1 (Fildes et al.)	Original	MSE	1.10	4.10	8.09	12.92	2.61	4.50	6.63
	Holt 2	Original	MSE	1.20	4.87	9.78	15.90	3.05	5.43	8.03
	Holt 3	Trimmed	MSE	1.22	4.83	9.56	15.59	3.02	5.33	7.86
	Holt 4	Trimmed	MAD	1.13	4.44	8.37	13.54	2.79	4.82	6.95
	Damped 1 (Fildes et al.)	Original	MSE	1.11	4.72	10.65	17.57	2.86	5.55	8.53
	Damped 2	Original	MSE	1.13	4.60	9.33	15.55	2.84	5.17	7.70
	Damped 3	Trimmed	MSE	1.13	4.35	8.52	14.27	2.73	4.82	7.10
	Damped 4	Trimmed	MAD	1.05	4.01	7.82	12.55	2.49	4.41	6.41
	SES with drift 1	Original	MSE	1.18	4.62	9.10	14.65	2.90	5.14	7.48
	SES with drift 2	Trimmed	MSE	1.18	4.56	8.86	14.27	2.87	5.04	7.31
	SES with drift 3	Trimmed	MAD	1.07	4.00	7.54	12.56	2.55	4.38	6.32
	Theta 1	Original	MSE	1.19	5.26	10.75	18.32	3.20	5.95	8.91
	Theta 2	Trimmed	MSE	1.24	5.02	9.77	15.62	3.10	5.54	8.04
	Theta 3	Trimmed	MAD	1.22	5.00	9.75	15.60	3.08	5.52	8.02

Table 1 (continued)

Origin	Method	Fit data	Fit criterion	Horizon 1	6	12	18	1-6	1-12	1-18
53	Robust trend	Original	MSE	**0.94**	**4.18**	**7.71**	**11.55**	**2.46**	**4.27**	**6.25**
	Holt 1 (Fildes et al.)	Original	MSE	1.25	6.09	11.16	17.54	3.54	6.32	9.15
	Holt 2	Original	MSE	1.15	5.78	11.08	16.80	3.37	6.15	8.97
	Holt 3	Trimmed	MSE	1.19	5.85	11.24	17.07	3.42	6.25	9.10
	Holt 4	Trimmed	MAD	1.15	5.48	10.07	15.34	3.21	5.71	8.22
	Damped 1 (Fildes et al.)	Original	MSE	1.06	5.58	14.69	20.95	3.15	6.31	10.64
	Damped 2	Original	MSE	1.05	5.18	9.51	14.76	2.99	5.32	7.83
	Damped 3	Trimmed	MSE	1.06	5.07	9.09	14.09	2.95	5.18	7.53
	Damped 4	Trimmed	MAD	1.05	4.76	8.43	12.30	2.78	4.86	6.88
	SES with drift 1	Original	MSE	1.01	5.19	9.91	14.64	3.00	5.47	7.96
	SES with drift 2	Trimmed	MSE	1.05	5.17	9.74	14.39	3.02	5.45	7.85
	SES with drift 3	Trimmed	MAD	**0.94**	4.31	7.95	12.10	2.54	4.43	6.51
	Theta 1	Original	MSE	0.71	5.03	13.19	15.12	2.72	5.60	8.93
	Theta 2	Trimmed	MSE	1.06	5.34	10.37	15.23	3.11	5.72	8.29
	Theta 3	Trimmed	MAD	1.05	5.33	10.37	15.23	3.10	5.71	8.28
Avg.	Robust trend	Original	MSE	**1.11**	**3.95**	**7.54**	11.80	**2.54**	**4.28**	**6.19**
	Holt 1 (Fildes et al.)	Original	MSE	1.36	5.28	9.82	15.05	3.37	5.66	8.05
	Holt 2	Original	MSE	1.47	5.24	9.88	15.14	3.38	5.67	8.10
	Holt 3	Trimmed	MSE	1.46	5.10	9.65	14.84	3.30	5.54	7.92
	Holt 4	Trimmed	MAD	1.26	4.76	9.00	13.83	3.01	5.15	7.37
	Damped 1 (Fildes et al.)	Original	MSE	1.35	5.77	12.29	19.00	3.58	6.21	9.72
	Damped 2	Original	MSE	1.45	5.01	9.49	14.79	3.25	5.43	7.82
	Damped 3	Trimmed	MSE	1.40	4.67	8.74	13.65	3.04	5.04	7.24
	Damped 4	Trimmed	MAD	1.24	4.35	8.33	12.98	2.77	4.72	6.84
	SES with drift 1	Original	MSE	1.40	4.70	9.00	13.81	3.05	5.13	7.37
	SES with drift 2	Trimmed	MSE	1.41	4.57	8.64	13.20	2.99	4.98	7.10
	SES with drift 3	Trimmed	MAD	1.17	4.01	7.56	**11.73**	2.60	4.34	6.23
	Theta 1	Original	MSE	1.16	5.14	10.80	16.30	3.11	5.60	8.47
	Theta 2	Trimmed	MSE	1.45	4.86	9.29	14.14	3.15	5.31	7.60
	Theta 3	Trimmed	MAD	1.25	4.76	9.32	14.25	3.00	5.23	7.57

Why did these improvements occur? The answers are straightforward. For the MSE fit using original data, the difference is due to the use of optimal smoothing parameters rather than heuristic parameters from a grid search. We experimented with several programs, and found that we were able to come very close to Fildes et al.'s damped trend results at all origins and horizons using Gardner's (1983) Autocast software, which employs an heuristic grid search procedure to minimize MSE after initial values are determined by least-squares regression. For the MSE fit with trimmed data, we made further improvements by avoiding excessive damping caused by trend reversals like that in Figure 1. Finally, the MAD fit minimized additional parameter distortion caused by outliers.

For the Holt method, we obtained better forecast accuracy than Fildes et al. at origins 23, 31, and 53, but not at origins 38 and 45. In fact, our Holt results are much worse than Fildes et al. at origins 38 and 45. Using Autocast, we were able to replicate Fildes et al.'s Holt results at origins 23, 31, and 53, but we could not do so at origins 38 and 45. At origin 38, the Autocast average MAPE over all horizons was 7.14%, compared to 6.31% in Fildes et al. At origin 45, Autocast gave an average MAPE over all horizons of 7.62%, compared to 6.63% in Fildes et al. We cannot explain why we could replicate some, but not all of Fildes et al.'s results.

SES with drift performed extremely well at all forecast origins. Detailed inspection of the results showed that this method was particularly sensitive to the fit criterion, and the MAD fit consistently produced better estimates of the fixed drift component. SES with drift beat all methods for the average of horizons 1-18 at origins 23 and 31, and was a close second to the robust trend at the other origins. For the average of all origins and horizons (last section of Table 1), SES with drift gave an MAPE of 6.23%, only a bit worse than the robust trend at 6.19%. As predicted from the work of Hyndman and Billah (2003), the Theta method performed poorly compared to SES with drift, giving an MAPE over all origins and horizons of 7.57%.

Why did SES with drift do so well? One simple explanation is that the trends in most of these series are linear and so consistent that there is no need to change the initial estimates obtained by least squares regression. Another explanation is more subtle, that SES with drift imitates the robust trend in many series. This is because the smoothing parameter was fitted at 1.0 about 40% of the time, which creates a method equivalent to the underlying model for the robust trend, an ARIMA (0, 1, 0) with drift.

Conclusions

SES with drift was clearly the best smoothing method overall in the telecommunications series. This method is simpler than the robust trend, but gives about the same forecast accuracy. Given the steady trends in most of the telecommunications series, the performance of the Theta method of forecasting was disappointing. If a fixed drift term is used with SES, we agree with Hyndman and Billah that it should be optimized.

With better model-fitting, the Holt and damped trend methods are much more competitive in the telecommunications data. In many other empirical studies in the literature, parameter searches for exponential smoothing methods have been carried out with heuristic procedures. Our results suggest that the smoothing methods in these studies should be re-fitted with optimal parameters, which may well change the conclusions.

Contrary to Fildes et al., we show that the average forecast accuracy of the damped trend method is better than that of the Holt method. This finding is consistent with theory (see Gardner, 2006) and with all other empirical comparisons in the literature.

The telecommunications series contain numerous outliers, and we found that a MAD fit made significant improvements in forecast accuracy compared to a conventional MSE fit. See also Gardner (1999) for a similar conclusion in an analysis of the annual time series from the M1 competition. Like the use of optimal parameters, a MAD fit could change the conclusions in other empirical studies involving exponential smoothing.

Our results suggest that one should trim irrelevant data before fitting exponential smoothing methods, but it is difficult to make general recommendations about how this should be done. Our trimming procedure is ad hoc and depends on domain knowledge that is often unavailable in time series forecasting. The only other research on trimming irrelevant data appears to be that of Collopy and Armstrong (1992) and Gardner (1999); in both papers, judgmental methods were used. The development of an automatic trimming algorithm for time series awaits further research.

References

Armstrong, J.S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error, *International Journal of Forecasting*, 22, 583-598.

Assimakopoulos, V. & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting, *International Journal of Forecasting*, 16, 521-530.

Collopy, F., & Armstrong, J.S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations, *Management Science*, 38, 1394-1414.

Fildes, R. (1989). Evaluation of aggregate and individual forecast method selection rules. *Management Science*, 35, 1056-1065.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, 8, 81-98.

Fildes, R., Hibon, M., Makridakis, S., & Meade, N. (1998). Generalizing about univariate forecasting methods: Further empirical evidence, *International Journal of Forecasting*, 14, 339-358.

Gardner, E.S., Jr. (1983). *Autocast Exponential Smoothing System*, Morristown, NJ: Core Analytic.

Gardner, E.S., Jr. (1999). Note: Rule-based forecasting vs. damped-trend exponential smoothing, *Management Science*, 45, 1169-1176.

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

Gardner, E.S., Jr., & McKenzie, E. (1985). Forecasting trends in time series, *Management Science*, 31, 1237-1246.

Grambsch, P. & Stahel, W.A. (1990). Forecasting demand for special services, *International Journal of Forecasting*, 6, 53-64.

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

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