Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 24 (2008) 170 – 174
www.elsevier.com/locate/ijforecast

Exponential smoothing in the telecommunications data

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

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

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 the accuracy of the Holt and damped trend methods can be improved by trimming the time series to eliminate irrelevant early data, fitting the methods to minimize the MAD rather than the MSE, and optimizing the parameters. 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 a closely related method, 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.

Keywords: Comparative methods — evaluation; Time series — exponential smoothing; Robust trend; Theta method

1. 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, Hibon, Makridakis, and Meade (1998), who found that the robust trend method was more accurate than Holt's additive trend or the Gardner and 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.

This paper attempts to replicate the exponential smoothing results in Fildes et al. We also test several ideas for improving 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 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, new empirical comparisons of the exponential smoothing methods and the robust trend are presented.

* Corresponding author. Tel.: +1 713 743 4744; fax: +1 713 743 4940.

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

0169-2070/$ - see front matter © 2007 Published by Elsevier B.V. on behalf of International Institute of Forecasters.
doi:10.1016/j.ijforecast.2007.05.002

2. The telecommunications series and the robust trend

The Fildes et al. collection includes 263 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. Fildes et al. dropped two series that contained numerous zeroes, leaving 261 for analysis, and we did the same. Fildes et al. claim that compared to the series used in the M competition (Makridakis et al., 1982), the telecommunications series are much more homogeneous. We agree. Although outliers contaminate nearly every series, about two thirds of the series are not especially difficult to forecast because they display steady downward trends, like Series A in Fig. 1.

The remaining series are more challenging. In about a quarter of the series, an abrupt trend reversal occurs in the early part of the series. 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 makes it difficult to estimate 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 series 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 discontinuities that have a major impact on average forecast accuracy for all series.

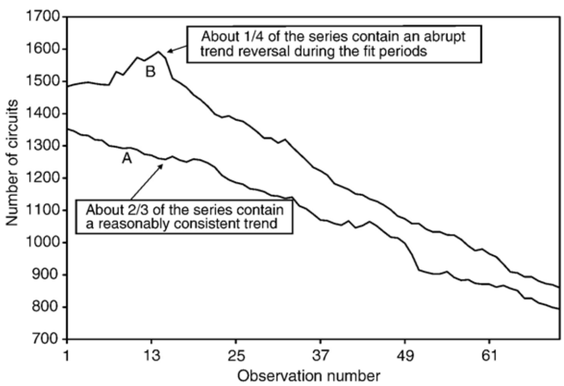

Fig. 1. Examples of the telecommunications series.

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 & Stahel for details).

3. 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:

$$\ell_t = \ell_{t-1} + b + \alpha\varepsilon_t \tag{1}$$

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

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 Eqs. (1) and (2), rather than fixed at a predetermined value like the Theta method. Both alternatives are tested below.

4. 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. Two sets of data were used to fit each method. The first fit used all data from period one through each forecast origin. In the second fit, the data were trimmed by discarding any observations prior to an early trend reversal. This was done by dividing the first set of fit periods (1–23) roughly in half. The slope of the first 12 observations was compared to the next 11; if the slope changed from positive to negative, the fit for all five forecast origins was started at the maximum observation value during the first 23 periods.

To initialize all methods, the intercept and slope of a classical linear trend were used. Trend lines were computed from the first fit period, as determined in the previous paragraph, through each forecast origin. One of the referees for this paper commented that it is good practice to fit a trend line only to the first part of the time series, say the first five observations, to obtain estimates of the beginning local intercept and slope. We agree with the referee if changes in trends are expected, but this was not the case in the telecommunications series. The trends in most series were so consistent (after adjusting for early trend reversals) that we chose to use all available observations for initialization.

For each set of fit data, we compared the mean squared error (MSE) and mean absolute deviation (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.

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.

5. Accuracy comparisons

Mean absolute percentage error (MAPE) results for the average of all forecast origins are given in Table 1, which can be replicated using data and computer code available on the IJF web site. Table 1 also contains

Table 1
Average MAPE over all forecast origins for the telecommunications data (261 series)

	Fit	Fit	Horizon						
Method	Data	Criterion	1	6	12	18	1–6	1–12	1–18
Robust trend	Original	MSE	**1.11**	**3.95**	**7.54**	11.80	**2.54**	**4.28**	**6.19**
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
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
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

Numbers in bold face indicate the minimum number in each column.

Fildes et al.'s original results for the robust trend, damped trend, and Holt methods. Fildes et al. also tested ARIMA and ARARMA methods, but the results are not repeated here because these methods performed poorly. We do not report median APEs because the differences among methods are similar to MAPE comparisons. For all methods, little difference in forecast accuracy was found between optimization of parameters over the [0,1] interval vs. the complete range of invertibility, so only the first option was used in Table 1. In the Holt and damped trend methods, little difference was found 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).

Our damped trend results are significantly better than those reported by Fildes et al. In Table 1, Fildes et al.'s MAPE (Damped 1) averaged over horizons 1–18 is 9.72%. Using an MSE criterion to fit the original data (Damped 2) reduces the MAPE to 7.82%. If we continue with the MSE criterion and trim the irrelevant early data (Damped 3), the MAPE falls to 7.24%. Minimizing the MAD with trimmed data (Damped 4) gives the best results, a MAPE of 6.84%. The improvements over the Fildes et al. damped trend results are consistent at all forecast origins and horizons.

Why did the improvements in the damped trend results occur? For the MSE fit using original data, the difference is due to the use of optimal smoothing parameters. We experimented with several programs, and found that we were able to obtain damped trend MAPEs approximately the same as those in Fildes et al. using Gardner's (1983) Autocast software. Parameters in Autocast are selected by a grid search procedure to minimize the MSE after initial values are determined by least-squares regression. The Autocast grid is rather coarse, and there are significant differences between the optimal Solver parameters and Autocast parameters in many series. For the MSE fit with trimmed data, further improvements were made by avoiding excessive damping caused by trend reversals like that in Fig. 1. Finally, the MAD fit minimized additional parameter distortion caused by outliers.

For the Holt method, Fildes et al. obtained results (Holt 1) better than our results using an MSE fit criterion with original data (Holt 2). We cannot explain this, because we were unable to replicate some of Fildes et al.'s results. Using Autocast, we obtained approximately the same Holt MAPEs as Fildes et al. at origins 23, 31, and 53. However, at origin 38, Autocast gave an average MAPE over all horizons of 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. Nevertheless, Fildes et al.'s Holt results can be improved, and the best option is to minimize the MAD using trimmed data. For the average of horizons 1–18, this strategy (Holt 4) produces an MAPE of 7.37%, compared to 8.05% for Fildes et al. (Holt 1).

SES with drift performed well at all forecast origins and horizons. A 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 1–18 in Table 1, SES with drift gave an MAPE of 6.23%, only slightly worse than the robust trend at 6.19%. As predicted by 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 1–18 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, creating a method equivalent to the underlying model for the robust trend, an ARIMA (0, 1, 0) with drift.

6. Conclusions

Chatfield (1978) demonstrated that automatic forecasting with exponential smoothing can often be improved through subjective judgment. The basic idea is simple – one should plot the data, and tailor method selection and model fitting to the features of the time series. In later work, Chatfield and Yar (1988) expanded on this strategy and called it a "thoughtful" use of exponential smoothing. For a discussion and additional references, see Gardner (2006, Section 5.5).

We attempted to use exponential smoothing thoughtfully in the telecommunications series. Plots displayed irrelevant early data in many series, and trimming such data improved forecast accuracy for all methods. In other applications, it is difficult to make general recommendations about how trimming should be done. Our trimming procedure is necessarily ad hoc, and depends on plotting the data as well as domain knowledge. There appear to be only two other papers on trimming time series, by Collopy and Armstrong (1992) and Gardner (1999); in both papers, judgmental methods were used. We agree with one of the referees for this paper, who commented that it is debatable whether an automatic trimming algorithm for time series ever could, or should, be developed.

To cope with outliers, we fitted all methods to minimize the MAD, which improved forecast accuracy over the conventional MSE criterion. It may be that a MAD fit would change the conclusions in other empirical studies involving exponential smoothing. The parameters for all methods were optimized during fitting, which in itself improved forecast accuracy. In many other empirical studies in the literature, parameter searches have been carried out with coarse grid search routines like those in Autocast. Our results suggest that the smoothing methods in these studies should be re-fitted with optimal parameters, which may change the conclusions.

Graphical analysis suggested that SES with drift would perform well, and it proved to be the most accurate smoothing method at every forecast horizon. Compared to the robust trend, SES with drift is a simpler method that gives about the same forecast accuracy. We expected that the Theta method of forecasting would also perform well, but its performance was disappointing. For the average of all origins and horizons 1–18, the Theta method was the worst method tested, regardless of fit data or criterion. If a fixed drift term is used with SES, Hyndman and Billah are correct that it should be optimized.

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

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.

Chatfield, C. (1978). The Holt-Winters forecasting procedure. *Applied Statistics*, *27*, 264−279.

Chatfield, C., & Yar, M. (1988). Holt-Winters forecasting: Some practical issues. *Journal of the Royal Statistical Society. Series D*, *37*, 129−140.

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*, *22*, 637−666.

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., 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.