Multi-item Sales Forecasting with

Total and Split Exponential Smoothing

James W. Taylor

Saïd Business School

University of Oxford

Journal of the Operational Research Society, 2011, Vol. 62, pp. 555–563.

Address for Correspondence:

James W. Taylor
Saïd Business School
University of Oxford
Park End Street
Oxford OX1 1HP, UK

Tel: +44 (0)1865 288927
Fax: +44 (0)1865 288805
Email: james.taylor@sbs.ox.ac.uk

Multi-item Sales Forecasting with Total and Split Exponential Smoothing

Abstract

Efficient supply chain management relies on accurate demand forecasting. Typically, forecasts are required at frequent intervals for many items. Forecasting methods suitable for this application are those that can be relied upon to produce robust and accurate predictions when implemented within an automated procedure. Exponential smoothing methods are a common choice. In this empirical case study paper, we evaluate a recently proposed seasonal exponential smoothing method that has previously been considered only for forecasting daily supermarket sales. We term this method 'total and split' exponential smoothing, and apply it to monthly sales data from a publishing company. The resulting forecasts are compared against a variety of methods, including several available in the software currently used by the company. Our results show total and split exponential smoothing outperforming the other methods considered. The results were also impressive for a method that trims outliers and then applies simple exponential smoothing.

Keywords: forecasting; exponential smoothing; supply chain; robust methods.

Introduction

In production and inventory control systems, forecasts are often required, at frequent intervals, for many different products or parts. In this situation, there is strong motivation to automate the forecasting process. The robustness and pragmatic appeal of smoothing methods has led to their extensive use in forecasting applications where a large number of series necessitates an automated procedure (Fildes *et al*, 2008; Syntetos *et al*, 2009). Exponential smoothing is particularly popular due to its impressive performance in empirical studies (Gardner, 2006). In their survey of sales forecasting practitioners, McCarthy *et al* (2006) report a greater level of satisfaction with exponential smoothing than any other method used.

The main aim of this paper is to provide further insight into the usefulness of a seasonal exponential smoothing method that was devised by a supermarket company, and presented by Taylor (2007). The method was developed for daily sales forecasting, and involves smoothing both the total weekly sales and the split of the total sales for each day of the week. We term the method 'total and split' exponential smoothing. In Taylor's empirical analysis, the method performed well for daily supermarket sales forecasting, particularly for the early lead times considered. In this paper, we apply the method to a dataset consisting of a large number of monthly sales series provided by a publishing company. We compare the accuracy of the method against a variety of alternatives, including several of the methods available in the demand planning module of SAP ERP, which is the enterprise resource planning software used by the publishing company. One of the less standard methods available in this software involves outlier correction prior to exponential smoothing. Inclusion of this method in our study is of particular interest because, in Taylor's analysis, similar outlier correction methods outperformed total and split exponential smoothing beyond the early lead times. As we discuss in the next section, outliers are certainly a feature of our monthly sales data.

The issue of outliers prompts consideration of a robust approach to parameter estimation. Given that absolute error is a more robust measure of error magnitude than squared error, it may be preferable to optimise exponential smoothing parameters by minimising the sum of absolute estimation sample errors, rather than the standard use of the sum of squared errors (see Gardner and Diaz-Saiz, 2008). We consider this in our analysis.

When forecasting a large number of series, one can either use a common method for all series or, using an automated procedure, separately select a method for each series, or perhaps for each cluster of series. There is no established procedure in the forecasting literature for automated method selection (Fildes *et al*, 2008). Procedures have been proposed that use information criteria, but these require statistical models estimated using maximum likelihood (see, for example, Hyndman *et al*, 2008, Chapter 7), and the more accurate methods considered in this study do not fall into this category. Gardner (2006, Section 5) categorises the other automated method selection procedures as being either based on time series characteristics or rule-based expert systems. To select from among exponential smoothing methods, typically, such procedures involve ad hoc trend and seasonality tests, which are hard to justify. Importantly, empirical studies have not provided convincing evidence of the benefit of using automated method selection approaches (Gardner, 2006; Fildes *et al*, 2008). A possible exception to this is the expert system incorporated in the forecasting package ForecastPro® (see Makridakis and Hibon, 2000). In this paper, we briefly consider a simple 'cross-validation' approach to method selection for each series. This proceeds by estimating methods using all available data except a final hold-out sample. The best performing method for the hold-out sample is chosen as the method to use to forecast future values. Billah *et al* (2006) describe this as a common approach to method selection.

An alternative to selecting separately a method for each series is to identify clusters of series for which a common method would be suitable. Fildes *et al* (2008) suggest that this is a potentially valuable, although challenging, area of research. Our interest in this was prompted by personnel at our collaborating company who suggested that we cluster all titles within a given publishing category, such as GMAT study aids or foreign language dictionaries.

In the next section, we describe the dataset, the structure of our empirical analysis, and the accuracy measures that we use to compare forecasting methods. The section that follows describes each of the methods that we consider. The rest of the paper describes a series of empirical 'experiments' that we conducted. We evaluate which methods are preferable, if the aim is to choose a single method to apply to all series. The next section briefly considers a cross-validation approach that selects a method separately for each series. We also briefly evaluate the benefit in assigning a common forecasting method to all series belonging to a particular cluster of series. The final section summarises our findings, and provides concluding comments.

Data, Structure of Study and Error Measures

The initial dataset consisted of 3880 monthly time series. The observations in each series corresponded to the sales of an individual title, classified by its ISBN. The data had been aggregated across markets from a number of different countries. As a result, the sales volumes were not low, and so we did not consider forecasting methods for intermittent demand (see Syntetos *et al*, 2009). The titles included books, CDs, DVDs and videos. As shown in Figure 1, the series varied greatly in length, ranging from just one observation to series of length 127 months. Four of the series are presented in Figures 2 to 5. These plots show some of the typical features present in the data. Most series seemed to possess high volatility; many series contained outliers with occasional extremely large outliers; few series possessed clear trend; and many series seemed to have some degree of seasonality, but the seasonality often changed or disappeared over time. Figure 4 shows a series that has demand steadily fading to zero. In one part of our empirical analysis, we consider the implications for our ranking of methods if we attempt to remove such 'fading' series. Figure 5 seems to show signs of seasonality, but there is barely sufficient data to be able to model the seasonality. In this paper, we follow the practice of the SAP ERP demand planning module by requiring at least 36 months of data in order to fit a seasonal forecasting method.

---------- Figures 1-5 ----------

We considered prediction from 1 to 18 months ahead. Lead times of between 6 to 18 months are of greatest relevance to the company, but we included the earlier lead times for completeness. As the focus of this paper is total and split exponential smoothing, which is a seasonal method, we only considered series for which at least 36 months were available for the estimation sample. In view of this, the shortest series that we considered consisted of 36+18=54 months of data. The resulting dataset contained 1849 series.

In each empirical experiment that we performed, for each series and for each lead time, we generated just one forecast from each method. A similar approach has been used in several high-profile forecasting competitions (see, for example, Markidakis and Hibon, 2000). An alternative is to use a rolling origin for each series, which would lead to more than one post-sample forecast for each lead time. To summarise a method's performance for each lead time, we averaged the accuracy across the series by calculating the root mean square error (RMSE) and the mean absolute error (MAE), which has the appeal of robustness. When averaging these measures across series, it is likely that they will be dominated by the higher volume series because errors for these series are likely to be larger than those from the lower volume series. Although higher volume series may be considered to be of greater importance, it is also interesting to consider error measures that are not dominated by these series. An appeal of percentage error measures is that they control for the level of the series. However, we could not use them in this study because the post-sample period for many of the series contained one or more observations close to or equal to zero. As an alternative, we used a relative absolute error measure, which, for each series and each lead time, was calculated as the absolute error for each method divided by the average of the absolute errors for all the methods. This was then averaged across all the series to deliver the mean relative absolute error (MRelAE). We also computed the average of each method's ranking in terms of absolute error (MRankAE). For simplicity in the reminder of this report, we present results in terms of only the MAE and MRankAE. The relative performances of the methods were broadly similar when compared using the RMSE and MRelAE.

Forecasting Methods

Total and Split Exponential Smoothing

The total and split exponential smoothing method is presented by Taylor (2007), who provides results that show the method performing well for daily supermarket sales data, particularly for short lead times. We are not aware of any other empirical evidence regarding the method, and we do not know of commercial software in which the method is implemented. In this paper, we apply the method to monthly data. For a series of monthly observations, y_t, the method involves smoothing the total yearly sales, T_t, and the split, S_t, of the yearly sales across the months of the year. The method has the following formulation:

$$T_t = \alpha \sum_{i=0}^{11} y_{t-i} + (1-\alpha) T_{t-1}$$

$$S_t = \frac{\gamma\, y_t}{\sum_{i=0}^{11} y_{t-i}} + (1-\gamma) S_{t-12}$$

where α and γ are smoothing parameters. The forecasts are given by $\hat{y}_t(k) = T_t\ S_{t+k-12}$ for lead times k = 1 to 12, and by $\hat{y}_t(k) = T_t\ S_{t+k-24}$ for k = 13 to 18. The method can be viewed as a hybrid of the ratio-to-moving average seasonal adjustment procedure (see Makridakis and Hibon, 2000) and Holt-Winters exponential smoothing with multiplicative seasonality and no trend (which, as we explain below, is termed 'N-M' exponential smoothing). The total and split method replaces the Holt-Winters smoothing of the level by smoothing of the yearly total. In the ratio-to-moving average seasonal adjustment approach, the total and split method can be viewed as replacing simple averages by exponentially weighted moving averages.

We used simple averages of the first few observations to calculate initial values for the smoothed components. Taylor (2007) describes how the supermarket company use the same subjectively chosen parameter values for all series; α=0.7 and γ=0.1. For our publishing sales data, we consider these as well as optimised values. For all exponential smoothing methods in this paper, we optimised parameters by minimising the sum of squared one step-ahead estimation sample forecast errors, as well as minimising the sum of absolute errors. Fildes *et al* (1998) note that using commonly occurring exponential smoothing parameters for all series can be preferable to using the values optimised for each series. We implemented the total and split method for all series with each parameter set as the median of all the optimised values of that parameter.

Non-Seasonal Methods

As our collaborating publishing company currently uses the demand planning module of SAP ERP to produce its forecasts, it seemed natural to use the more attractive methods available in this module as benchmarks against which to compare the total and split exponential smoothing method. As many of the series did not show clear seasonality, we included non-seasonal as well as seasonal methods in our forecast comparison. The nature of demand is that it obviously cannot attain a negative value. Therefore, for all methods considered in this paper, we imposed the constraint that if a forecast is produced that is negative, that forecast is set to zero. We implemented the following non-seasonal methods:

Naïve - The value at the forecast origin is used as the forecast for all future periods.

Linear trend - This method uses regression with time as the single regressor.

Simple exponential smoothing (N-N) - This method involves smoothing the level of the series through the use of a single smoothing parameter, α. The notation 'N-N' indicates no trend and no seasonal component in the classification of exponential smoothing methods introduced by Hyndman et al. (2002) and extended by Taylor (2003).

Trend exponential smoothing (A-N) - This method involves smoothing the level and trend of the series through the use of two parameters. The notation 'A-N' indicates an additive trend and no seasonal component.

Damped trend exponential smoothing (DA-N) - This method smoothes the level and additive trend of the series, and dampens the trend in the forecast function. It is an attractive candidate method because of its impressive performance in a number of empirical studies (Gardner, 2006). The notation 'DA-N' indicates a damped additive trend and no seasonal component.

Methods Using Seasonal Adjustment

A common approach to forecasting seasonal time series is to seasonally adjust the data, then apply a non-seasonal method to produce forecasts, and finally incorporate seasonality into the forecasts (see, for example, Makridakis and Hibon, 2000). We used the seasonal adjustment method based on ratio-to-moving averages. The SAP ERP demand planning module enables only the linear trend method to be used as the non-seasonal method. In our analysis, we considered this, and the non-seasonal exponential smoothing methods: simple (N-N), trend (A-N) and damped trend (DA-N).

Simple exponential smoothing (N-N) with outlier correction - Many of the series contain outliers, so there is strong appeal in using a method that is robust to outliers. Taylor's (2007) results for daily supermarket sales data showed that the total and split exponential smoothing method was outperformed beyond about six days ahead by a number of robust methods based on quantile forecasting. Several of these methods applied simple exponential smoothing to series that had been winsorised. This involves replacing all in-sample observations below a chosen estimated lower quantile by the value of this estimated quantile, and all in-sample observations above a chosen estimated upper quantile by this value. For our publishing data, we implemented a similar outlier correction method that is available in the demand planning module of the SAP ERP software. This method proceeds by fitting an exponential smoothing method, and then, for each in-sample period, a tolerance interval is constructed as the one-step-ahead forecast plus or minus a ‘correction factor’ multiplied by the MAE of the estimation sample one-step-ahead forecast errors. Starting from the beginning of the sample, if an observation falls outside the interval, it is replaced by the nearest bound of the interval. When this occurs, the exponential smoothing method is refitted, and the process is repeated. This continues until the end of the estimation sample. The software documentation provided no guidance as to how to choose the correction factor, and so, for simplicity, we subjectively selected factors of 1 and 2. As we found benefit in using the outlier correction procedure with only simple exponential smoothing, in the remainder of this paper, we report results for only this form of the method. We report the results for the method applied to seasonally adjusted data, as this led to improved accuracy.

Other Seasonal Methods

Seasonal naïve - The forecast for a future month is the most recently observed value for the same month of the year.

Seasonal (no trend) exponential smoothing (N-M) - This method smoothes the level and seasonality through the use of two smoothing parameters. We considered both additive and multiplicative seasonality formulations. We found multiplicative seasonality to be superior, and so in the remainder of the paper we report only the results for this version. The notation ‘N-M’ indicates no trend and the use of multiplicative seasonality. In our analysis, the results for this method were very sensitive to the approach used to initialise the seasonal indices. Poor results were produced when we used the initialisation expressions that, according to the online manual, are used in the SAP ERP demand planning module.

Therefore, in this paper, as with all other exponential smoothing methods, we opted to initialise the method using simple averages based on the early observations in each series.

Seasonal trend exponential smoothing (A-M) - This method smoothes the level, trend and seasonality through the use of three smoothing parameters. The notation 'A-M' indicates additive trend and multiplicative seasonality.

Seasonal damped trend exponential smoothing (DA-M) - This method smoothes the level, trend and seasonality, and dampens the trend in the forecast function. The notation 'DA-M' indicates damped additive trend and multiplicative seasonality. The method involves three smoothing parameters and a dampening parameter.

Comparing the Accuracy of Individual Methods

Experiment 1 - Series with at least 36 months in the estimation sample

Table 1 presents the MAE and MRankAE post-sample results for methods that have each been applied to the 1849 series with at least 36+18=54 months of data. The table includes just one exponential smoothing method that had parameters estimated by minimising the sum of squared errors. For the other exponential smoothing methods shown, the parameters were estimated by minimising the sum of absolute errors. We have the following comments regarding the results of Table 1:

(i) The accuracies of the linear trend method and the non-seasonal exponential smoothing methods were each improved by applying the methods with seasonal adjustment.

(ii) The results for damped trend exponential smoothing (DA-N) show that dampening the additive trend delivers improved accuracy, which is consistent with other similar studies.

(iii) The inclusion of the outlier correction in simple exponential smoothing (N-N) led to noticeable improvement in the accuracy of this method. Using a correction factor of 2 was more successful than a factor of 1.

(iv) Given its simplicity, the naïve seasonal method was surprisingly competitive. Its superiority over the naïve method indicates that there is seasonality in many of the series.

(v) The seasonal (no trend) exponential smoothing method (N-M) was more successful than the similar method with trend term included (A-M). The N-M method delivered accuracy comparable with the better of the simple exponential smoothing methods with outlier correction and seasonal adjustment.

(vi) It is interesting to compare the results for the three traditional seasonal exponential smoothing methods, N-M, A-M and DA-M, with the results for the corresponding three methods that involve seasonal adjustment instead of seasonal smoothing, N-N, A-N, DA-N. In terms of MRankAE, smoothing seasonality was preferable to seasonal adjustment for all three cases. In terms of MAE, using seasonal adjustment was a little better for the damped trend method, but smoothing seasonality was more accurate for simple and trend exponential smoothing. These results are, perhaps, a little surprising because it tends to be assumed in the literature that using seasonal adjustment with N-M, A-M and DA-M is preferable to using the corresponding seasonal smoothing methods (see, for example, Makridakis and Hibon, 2000).

(vii) Estimating the parameters of total and split exponential smoothing by minimising the sum of absolute errors was preferable to minimising the sum of squared errors. This was a consistent result across all of the exponential smoothing methods.

(viii) The best results across all lead times were achieved by total and split exponential smoothing. We considered outlier correction for this method, but it did not lead to greater accuracy. Outliers are difficult to identify in the presence of seasonality, and so it is perhaps not surprising that we found that outlier correction was more useful for the non-seasonal simple exponential smoothing method than for the seasonal total and split exponential smoothing method.

(ix) Total and split exponential smoothing with parameters optimised separately for each series was slightly outperformed by the method with α=0.48 and γ=0.24 for all 1849 series.

These values were the median of the optimised values for all the series, where the optimisation was performed by minimising the sum of absolute errors. Interestingly, beyond 6 months ahead, total and split exponential smoothing was most accurate when the subjectively chosen values of α=0.7 and γ=0.1 were used. The reason for the optimised parameters being more useful at the early lead times may be due to the parameter optimisation using one step-ahead errors. This suggests that it may be useful to optimise parameters separately for each lead time using in-sample forecast errors corresponding to that lead time. We note that optimising parameters separately for each lead time was considered unappealing by those responsible for forecasting at the publishing company.

---------- Table 1 ----------

Experiment 2 - Series with between 36 and 72 months in the estimation sample

Inspection of some of the longer series revealed that the demand level was fading to very low values towards the end of the series. An example of this is given in Figure 4. These series correspond to titles that are coming to the end of their life cycle. The company were curious to know whether the relative performances of methods in Table 1 would be repeated if we were to try to eliminate the terminating life cycle feature from the dataset. Although this feature was certainly not present in all of the longer series, in our second experiment, we decided to repeat the analysis of Experiment 1 using the same 1849 series and a maximum of the first 90 observations from each series. This meant that no more than 72 months of each series would be used in the estimation sample, with the remaining 18 months used as the evaluation sample. We do not present the detailed results of Experiment 2, as they were generally consistent with the results of Experiment 1. However, we note the following two findings from Experiment 2 that differed from Experiment 1:

(i) For total and split exponential smoothing, optimising the parameters by minimising the sum of absolute errors gave substantially better accuracy than using the subjectively chosen values of α=0.7 and γ=0.1.

(ii) The total and split method, with parameters optimised by minimising the sum of absolute errors, delivered similar accuracy to that of seasonal (no trend) exponential smoothing (N-M) and simple exponential smoothing (N-N) with outlier correction factor of 2.

Automated Method Selection

Experiment 3 - Simple Cross-Validation

In this section, we consider a simple cross-validation approach to method selection for each series. This proceeds by estimating methods using all except a final hold-out sample of the estimation sample. The method found to be most accurate at forecasting for the hold-out sample is chosen as the method to use for that series. We considered method selection from the following five methods that performed reasonably well in the earlier analysis: simple exponential smoothing (N-N) with seasonal adjustment and outlier correction with a correction factor of 2; seasonal naïve; seasonal (no trend) exponential smoothing (N-M); seasonal damped trend exponential smoothing (DA-M); and total and split exponential smoothing. For each exponential smoothing method, the parameters were estimated by minimising the sum of absolute errors. We used 24 months in the cross-validation period, and judged forecasting performance in this period using 12 month-ahead MAE, calculated for months 12 to 24 in the cross-validation period. Other lead times could have been considered, but we chose 12 month-ahead prediction due to its relevance for the company. Having selected an individual method in this way, before producing forecasts for the post-sample period, we re-estimated method parameters with the cross-validation period included in the estimation sample. With at least 36 months of data needed in each estimation sample, 24 months required for cross-validation, and a further 18 for post-sample evaluation, we restricted our attention to series of length at least 36+24+18=78 months. This resulted in the use of 1416 series.

Post-sample results are presented in Table 2. We have included the performance for 12 months ahead, as this was the lead time used to select methods. For this lead time, the method selection approach is outperformed by the seasonal naïve method. Looking at the results across all lead times, we see the seasonal naïve method was the poorest. Overall, the most accurate method was the total and split exponential smoothing method. The results suggest that cross-validation method selection may well not be able to improve on the use of the best individual method used for all series. We reached the same conclusion when repeating the experiment with total and split exponential smoothing removed.

---------- Table 2 ----------

Automated Selection of a Method for a Cluster of Time Series

We now consider a method selection approach that allocates the same method to all series within a given cluster. The clustering of series is defined by the publishing category within which each title lies. These categories are defined by the publishing company, and two examples are dictionaries for a particular language and a series of examination study aids.

Experiment 4 - Using post-sample results (cheating) to allocate a method to each cluster

In this experiment, we essentially cheated by allocating a method to a cluster based on post-sample performance of forecasting methods for all series in that cluster. The purpose of this analysis is to establish whether finding a common method for all series in a cluster is an idea worth pursuing. In this experiment, we used the same 1849 series from Experiment 1, and the five individual methods considered in Experiment 3. The experiment proceeded by noting the rankings of methods, in terms of 12 month-ahead MAE in the post-sample period of 18 months. For each cluster, we averaged the rankings across all series in that cluster, and noted the method with highest average ranking. This method was assigned to all series in that cluster. We then used this assignment of methods in a method selection procedure. The results are given in Table 3. The underlined entries show that for 12-month-ahead prediction, the results for the method selection approach are better than any other method. This indicates that using a common method for all series in a cluster is an idea with some potential. However, an approach is needed for assigning a method to each cluster using just the estimation sample, and we consider this in Experiment 5. It is disappointing to see that the method selection approach performed poorly at lead times other than 12 months. This suggests that the assignment of a single method to each cluster may only be useful if a different method is permitted for different forecast lead times. We note that this was viewed as unattractive by those responsible for forecasting at the publishing company.

---------- Table 3 ----------

Experiment 5 - Using cross-validation (not cheating) to allocate a method to each cluster

This experiment involved no cheating, but instead used cross-validation to allocate a method to each cluster. For each series, we ranked each method in terms of its 12 month-ahead MAE for a 24 month cross-validation period. The method allocated to a cluster was the one with highest average ranking for the series in that cluster. As in Experiment 3, we used the 1416 series that were composed of at least 78 monthly observations. We used the same five individual methods considered in Experiments 3 and 4. The results in Table 4 show that, overall, the method selection procedure was not able to outperform total and split exponential smoothing.

---------- Table 4 ----------

Summary and Concluding Comments

Overall, the best results in our study were achieved by the total and split exponential smoothing method. Other methods that produced competitive results were seasonal (no trend) exponential smoothing (N-M), and simple exponential smoothing (N-N) with seasonal adjustment and outlier correction. Indeed, it was interesting to see that the inclusion of the outlier correction in simple exponential smoothing led to noticeable improvement in the accuracy of this method. In our empirical analysis, an outlier correction factor of 2 led to greater accuracy than a factor of 1. However, it would be interesting to consider some form of optimisation of this factor. Estimating the parameters of the exponential smoothing methods by minimising the sum of absolute errors was preferable to minimising the sum of squared errors. For total and split exponential smoothing, the results were mixed for the use of the same subjectively chosen parameters (α=0.7 and γ=0.1) for all series, and so we would recommend parameter optimisation. An approach that delivered strong results in our study is to use, for all series, parameters set as the median of the set of values optimised for each series.

We considered cross-validation for the automated selection of a forecasting method for each series. In this procedure, the forecasting method found to be most accurate for the hold-out sample is chosen as the method to use for that series. Our results showed that this method selection approach was not as accurate as the use of total and split exponential smoothing for all series. We also considered a method selection approach that allocates the same method to all series within a cluster, defined according to the publishing company's descriptive categorisation of the titles. Although our analysis indicated that using the same method for all series in a cluster can lead to improved accuracy, a major issue remains as to how to select a method for each cluster. Using cross-validation to select methods for a cluster did not lead to results that were better than the best of the individual methods.

Further work in the area of method selection would be useful. Another potential area for research would be to investigate the appropriate level of aggregation of the data. We used data aggregated across a large number of country-specific markets, but even if a forecast at the aggregate level is needed, there may be benefit in modelling and producing forecasts at the disaggregated market level for each title. In addition to point forecasting, estimates of forecast uncertainty, such as prediction intervals, are also important (Datta *et al* 2007). For total and split exponential smoothing, it is straightforward to write out the innovations state space model, and from this prediction intervals can be simulated (see Hyndman *et al*, 2008, Chapter 6). One caveat to this is that, for the prediction intervals to be valid, the model parameters should be estimated using a statistical procedure, such as maximum likelihood, rather than minimising the sum of absolute errors, which we have advocated for point forecasting in this paper. This implies a potential need to sacrifice a degree of point forecast accuracy to enable the estimation of prediction intervals.

Acknowledgments

We would like to thank the collaborating publishing company for providing the data and background information for the study. We are also grateful for the very useful comments of two anonymous referees.

References

Billah MB King ML Snyder and RD Koehler AB (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting* **22**: 239-249.

Datta S Granger CWJ Barari M and Gibbs T (2007). Management of supply chain: an alternative modelling technique for forecasting. *Journal of Operational Research* **58**: 1459 - 1469.

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

Fildes R Nikolopoulos K Crone SF and Syntetos AA (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society* **59**: 1150-1172.

Gardner ES Jr (2006). Exponential smoothing: the state of the art - Part II. *International Journal of Forecasting* **22**: 637-666.

Gardner ES Jr and Diaz-Saiz J (2008). Exponential smoothing in the telecommunications data. *International Journal of Forecasting* **24**: 170-174.

Hyndman RJ Koehler AB Ord JK and Snyder RD (2008). *Forecasting with exponential smoothing: The state space approach*. Springer-Verlag: Berlin, Heidelberg, Germany.

Hyndman RJ Koehler AB Snyder RD and Grose S (2002). A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting* **18**: 439-454.

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

McCarthy TM Davis TF Golicic and SL Mentzer JT (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting* **25**: 303-324.

Syntetos AA Boylan JE and Disney SM (2009). Forecasting for inventory planning: A 50-year review. *Journal of the Operational Research Society* **60**: S149-S160.

Taylor JW (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting* **19**: 715-725.

Taylor JW (2007). Forecasting supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research* **178**: 154-167.

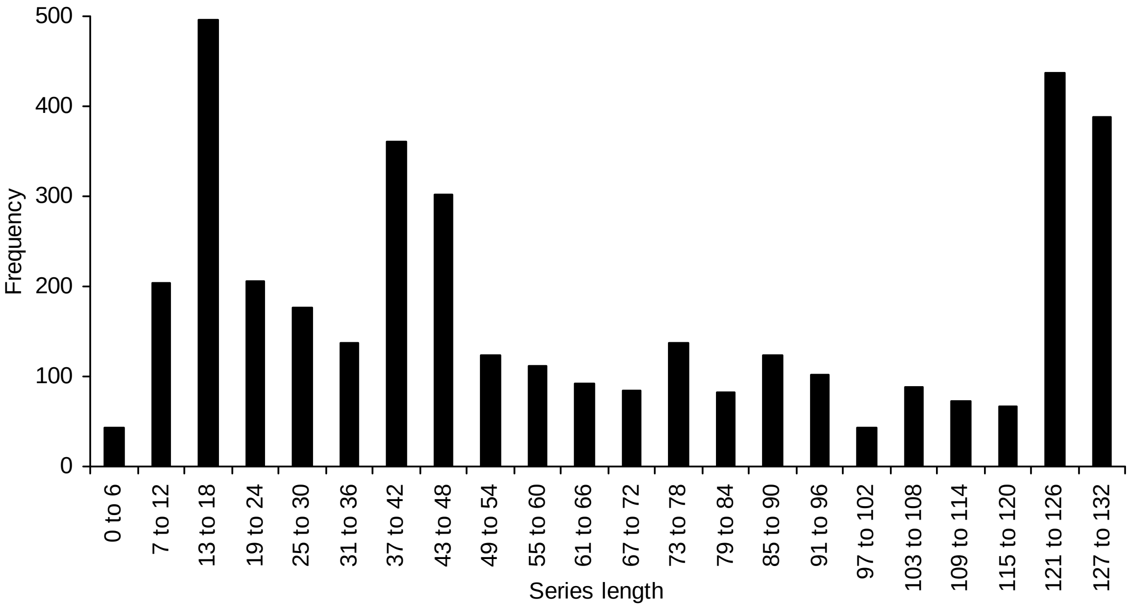

Figure 1 Histogram for the lengths of the 3880 time series.

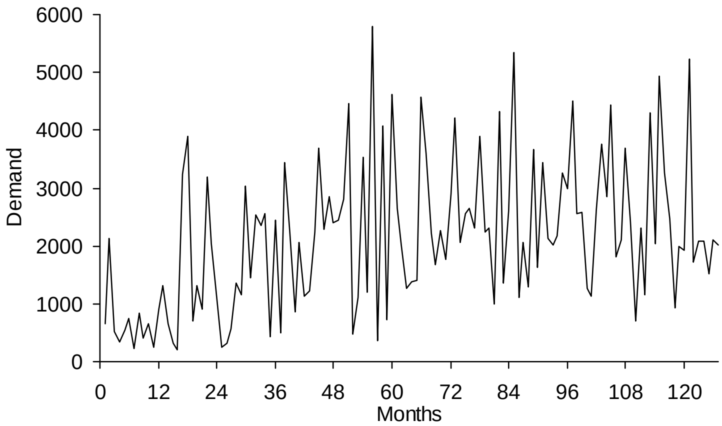

Figure 2 Time series plot for Series 1.

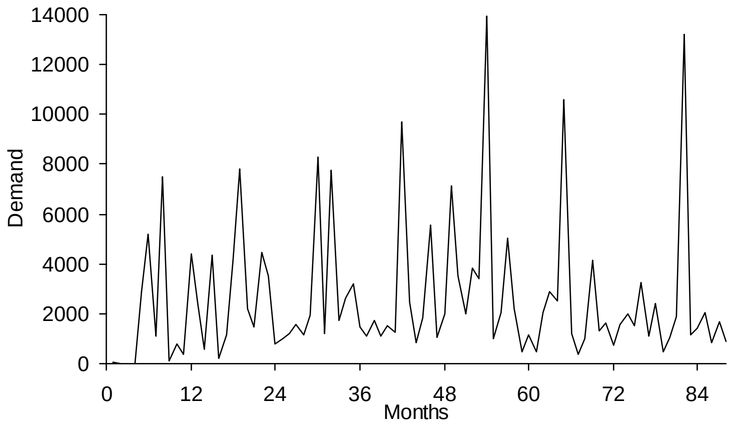

Figure 3 Time series plot for Series 2.

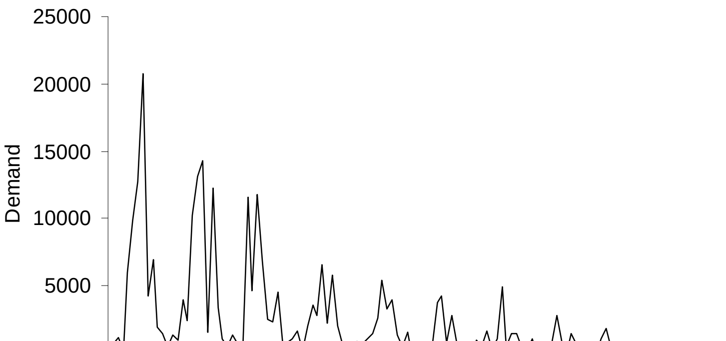

Figure 4 Time series plot for Series 3.

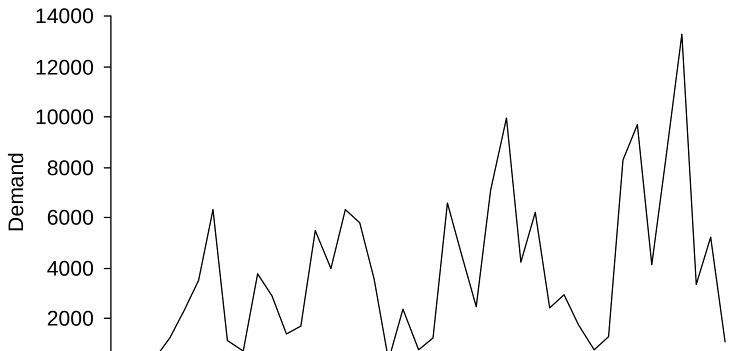

Figure 5 Time series plot for Series 4.

Table 1 Experiment 1 – Post-sample comparison of methods for 1849 series with at least 36 months in the estimation sample. (Lower values are better. Bold indicates four lowest values in each column. Unless stated to the contrary, exponential smoothing methods estimated using absolute errors.)

	MAE				MRankAE			
	Forecast lead time				Forecast lead time			
	1-6	7-12	13-18	All	1-6	7-12	13-18	All
Non-seasonal methods								
Naïve	620	586	633	613	10.8	10.2	10.6	10.6
Linear trend	487	523	572	527	11.3	10.9	11.1	11.1
Simple exp sm (N-N)	447	459	489	465	10.9	10.9	11.3	11.0
Trend exp sm (A-N)	458	492	554	501	10.1	9.8	10.3	10.1
Damped trend exp sm (DA-N)	429	439	466	445	10.1	9.7	10.1	10.0
Methods using seasonal adjustment								
Linear trend	573	710	690	658	11.0	11.5	10.9	11.1
Simple exp sm (N-N)	381	491	416	429	10.4	11.1	10.6	10.7
Simple exp sm (N-N) with outlier correction factor of 1	356	471	378	402	10.1	10.7	10.2	10.4
Simple exp sm (N-N) with outlier correction factor of 2	337	447	364	383	9.7	10.3	9.7	9.9
Trend exp sm (A-N)	391	519	480	463	9.8	10.1	10.0	10.0
Damped trend exp sm (DA-N)	373	490	410	424	9.8	10.0	**9.6**	9.8
Other seasonal methods								
Seasonal naïve	382	457	422	420	10.2	9.8	10.3	10.1
Seasonal (no trend) exp sm (N-M)	341	421	386	383	10.0	10.5	10.5	10.3
Seasonal trend exp sm (A-M)	359	440	444	414	9.6	**9.6**	9.9	9.7
Seasonal damped trend exp sm (DA-M)	389	480	452	440	9.5	**9.6**	**9.6**	9.6
Total and split exp sm	**313**	**393**	**332**	**346**	**8.8**	**8.7**	**8.7**	**8.8**
Total and split exp sm (optimised using squared errors)	**323**	**414**	**360**	**366**	**9.4**	**9.6**	**9.6**	**9.5**
Total and split exp sm ($\alpha = 0.48$, $\gamma = 0.24$, median optimised)	**306**	**378**	**326**	**337**	**9.0**	**9.0**	**8.9**	**9.0**
Total and split exp sm ($\alpha = 0.7$, $\gamma = 0.1$)	**327**	**331**	**285**	**314**	**9.4**	**8.1**	**8.1**	**8.5**

Table 2 Experiment 3 – Post-sample evaluation of the benefit of separate cross-validation method selection for each series. (Lower values are better. Bold indicates the lowest value in each column. Underlining indicates the results for the forecast lead time used in the cross-validation. Parameters optimised for each exponential smoothing method using absolute errors.)

	MAE					MRankAE				
	Forecast lead time					Forecast lead time				
	12	1-6	7-12	13-18	All	12	1-6	7-12	13-18	All
Simple exp sm (N-N) with outlier correction factor of 2 after seasonal adjustment	595	318	448	357	374	3.8	3.6	3.7	3.5	3.6
Seasonal naïve	**487**	365	464	419	416	3.3	3.7	3.5	3.7	3.6
Seasonal (no trend) exp sm (N-M)	570	333	434	395	388	3.8	3.7	3.8	3.8	3.8
Seasonal damped trend exp sm (DA-M)	554	328	437	377	381	3.5	3.5	3.4	3.5	3.5
Total and split exp sm	526	**292**	**400**	**327**	**340**	**3.1**	**3.1**	**3.1**	**3.1**	**3.1**
Method selection based on 12 month-ahead cross-validation	547	322	427	372	374	3.5	3.4	3.4	3.4	3.4

Table 3 Experiment 4 – Evaluation of the benefit of using post-sample performance to allocate a common method to all series in a cluster. (Lower values are better. Bold indicates the lowest value in each column. Underlining indicates the results for the forecast lead time used for the method selection. Parameters optimised for each exponential smoothing method using absolute errors.)

	MAE					MRankAE				
	Forecast lead time					Forecast lead time				
	12	1-6	7-12	13-18	All	12	1-6	7-12	13-18	All
Simple exp sm (N-N) with outlier correction factor of 2 after seasonal adjustment	577	324	427	342	364	3.8	3.5	3.6	3.4	3.5
Seasonal naïve	481	382	457	422	420	3.5	3.7	3.6	3.7	3.6
Seasonal (no trend) exp sm (N-M)	539	341	421	386	383	3.9	3.7	3.8	3.8	3.8
Seasonal damped trend exp sm (DA-M)	538	353	434	404	397	3.6	3.5	3.5	3.5	3.5
Total and split exp sm	513	**313**	**393**	**332**	**346**	3.4	**3.2**	**3.2**	**3.2**	**3.2**
Method selection based on 12 month-ahead post-sample results	**447**	330	407	361	366	**2.8**	3.4	3.3	3.4	3.4

Table 4 Experiment 5 – Post-sample evaluation of the benefit of using cross-validation to allocate a common method to all series in a cluster. (Lower values are better. Bold indicates the lowest value in each column. Underlining indicates the results for the forecast lead time used in the cross-validation. Parameters optimised for each exponential smoothing method using absolute errors.)

	MAE					MRankAE				
	Forecast lead time					Forecast lead time				
	<u>12</u>	1-6	7-12	13-18	All	<u>12</u>	1-6	7-12	13-18	All
Simple exp sm (N-N) with outlier correction factor of 2 after seasonal adjustment	<u>561</u>	303	424	333	353	<u>3.6</u>	3.5	3.6	3.4	3.5
Seasonal naïve	**<u>487</u>**	365	464	419	416	<u>3.4</u>	3.7	3.5	3.7	3.6
Seasonal (no trend) exp sm (N-M)	<u>570</u>	333	434	395	388	<u>3.9</u>	3.8	3.9	3.9	3.9
Seasonal damped trend exp sm (DA-M)	<u>554</u>	328	437	377	381	<u>3.5</u>	3.6	3.5	3.6	3.5
Total and split exp sm	<u>526</u>	**292**	**400**	**327**	**340**	**<u>3.2</u>**	**3.2**	**3.2**	**3.2**	**3.2**
Method selection based on 12 month-ahead cross-validation	<u>533</u>	312	405	351	356	<u>3.4</u>	3.3	3.3	**3.2**	3.3