Proceedings

ITISE 2016

International work-conference on Time Series

Granada

27-29, June, 2016

Combining simple exponential smoothing models for time series forecasting

F. Martínez, M.D. Pérez, F. Charte, and M.J. del Jesus

Department of Computer Science, University of Jaén, Spain

Abstract. Simple exponential smoothing is a well-known technique for forecasting univariate time series without trend and seasonality. Forecast combinations such as medians or means are known to improve the accuracy of point forecasts. In this paper we have experimented with combining the forecasts of several simple exponential smoothing models with different smoothing factors. Experimental results, using the M3-competition time series, show that the combined forecasts outperform the forecasts of the model that best fits the series.

Keywords: Combined forecast, Time series forecasting, Simple exponential smoothing

1 Introduction

Time series forecasting is a key tool in many areas, such as Hydrology, Business or Biology, wherein well-known statistical models such as ARIMA or exponential smoothing are used to make predictions about the future values of a time series [1]. Very often a large number of series need to be forecast, in this situation an automatic forecasting algorithm is an essential tool [2].

Ensemble learning [3] is a useful technique used by data mining practitioners to improve classification and numeric prediction accuracy. In this technique instead of learning one model from a training data set several models are learned. That is, multiple models that "fit well" the training data are selected from a candidate model space containing models that"explain" the training data. In order to make predictions an ensemble learner combines the predictions of its learned models using a combination function such as the average or the median. The predictive performance of an ensemble learner often outperforms the performance of the individual models comprising the ensemble. Why? A single model can be either too simple and not capture the essential patterns in the data, or too complex and overfit the data. In fact, there is hardly ever a true underlying model, and even if there was, selecting that model will not necessarily give the best predictions, because the parameter estimation may not be accurate. On the other hand, an ensemble of models can capture different patterns of the data and reduce the danger of overfitting or the uncertainty of choosing the wrong model.

The idea underlying ensemble learning has also been applied in time series forecasting to produce combined forecasts [4–6]. One of the four conclusions of the M3-competition [7] is that "the accuracy when various methods are being combined outperforms, on average, the individual methods being combined".

Exponential smoothing [8] comprises a set of univariate time series forecasting methods, in which forecasts are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In [9] up to thirty of these methods, depending on the series, are combined in several ways to improve forecast accuracy. In this paper we have worked with only one exponential smoothing method: simple exponential smoothing, that is suitable to forecast time series without trend and seasonality. This method has a smoothing parameter that controls how fast the weights of the past observations decay. The smoothing parameter is usually estimated from a time series and once estimated, the series can be forecast. Instead of working with only a smoothing parameter we have experimented with selecting multiple simple exponential smoothing models with different smoothing parameters. Then, we have combined the forecasts of the multiple models to assess whether the combined forecasts outperform the forecasts of the "best" model, that is, the model whose parameters are estimated from the time series.

The remainder of the paper is structured as follows. In Section 2 the simple exponential smoothing technique is described. In Section 3 the combined forecast approach that we are proposing is explained. Section 4 shows the experimental results and in Section 5 some conclusions are drawn.

2 Simple exponential smoothing

Simple exponential smoothing, also known as single exponential smoothing, is a technique for forecasting univariate time series without trend and seasonality. Given a time series whose observed values are: $x_1, x_2, \ldots x_T$, we call $\hat{x}_{T+1|T}$ to the one-step ahead prediction for the time series, that is, the prediction for horizon $T+1$ considering that the past values $x_1, x_2, \ldots x_T$ are known. In simple exponential smoothing the one-step ahead forecast is computed using the following recursive equation:

$$\hat{x}_{T+1|T} = \alpha x_T + (1-\alpha)\hat{x}_{T|T-1} \quad (1)$$

where $\alpha \in [0,1]$ is the smoothing factor or parameter. That is, the forecast is a weighted average of the last value of the time series and the one-step ahead forecast of the last value. If the expression $\hat{x}_{T|T-1}$ is substituted recursively in (1) we obtain:

$$\hat{x}_{T+1|T} = \alpha x_T + \alpha(1-\alpha)x_{T-1} + \alpha(1-\alpha)^2 x_{T-2} + \ldots + \alpha(1-\alpha)^{T-1}x_1 + (1-\alpha)^T l_0 \quad (2)$$

where l_0 is a prediction for the first value of the time series. Analyzing this last equation is easy to understand why the method is called exponential smoothing. The one-step forecast is a weighted average of the past values in the series; the weights attached to the past values decaying exponentially as we go back in time at a rate controlled by the smoothing parameter α. When α is close to 0 more weight is given to the distant past, when α is close to 1 more weight is given to the recent past. At the extreme cases, when $\alpha = 1$ we get naive forecast because $\hat{x}_{T+1|T} = x_T$, when $\alpha = 0$ we get $\hat{x}_{T+1|T} = l_0$—in this case l_0 is normally the time series average value, so this is suitable for a stationary process.

The parameters α and l_0 can be manually chosen by an expert. However, a more objective way of selecting these parameters, which is also suitable for automatic forecasting, is to estimate them from the time series. This is usually done choosing the parameters that minimize the sum of squared errors (SSE) of the in-sample one-step ahead forecasts, that is:

$$SSE = \sum_{t=1}^{T} (x_t - \hat{x}_{t|t-1})^2 \tag{3}$$

this implies a non-linear minimization problem and an optimization tool is used.

Figure 1 shows the in-sample one-step ahead forecasts for a time series using simple exponential smoothing with two different smoothing parameters. The time series is the fifth series from the M3-competition [7]: a yearly time series with observations from 1975 to 1988. In Figure 1.a the parameters l_0 and α are estimated from the data so that SSE is minimized ($SSE = 8,173,322$), α is estimated as 0.97, very close to 1, so the forecasts are similar to the predictions of the naive method. In Figure 1.b l_0 and α are manually chosen, l_0 is set to the first value of the time series (4977.18) and α is set to 0.6, the forecasts form a smoother series and the SSE obtained with these parameters is 9,013,960.

Simple exponential smoothing is suitable for time series with no trend or seasonal pattern, so that the out-of-sample k-steps ahead forecasts are equal to the out-of-sample one-step ahead forecast:

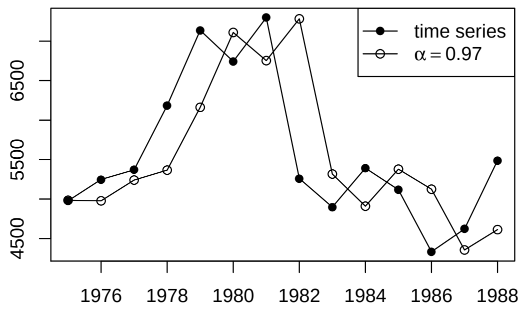
$$\hat{x}_{T+h|T} = \hat{x}_{T+1|T}, h \geq 2$$

Figure 2 shows the forecasts for the horizon $T + h$, with $h = 1, 2, \ldots 6$, for the two simple exponential smoothing models in Figure 1, the out-of-sample or test observations can also be seen.

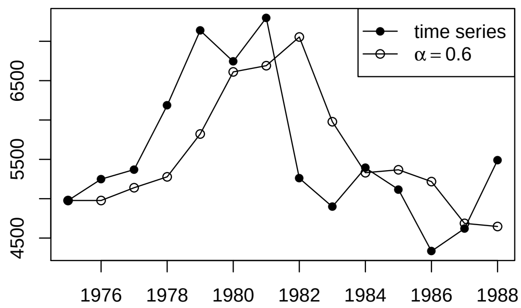
3 The combined forecast approach

In this section the combined forecast technique that we are proposing is explained. The approach has two stages: selection and combination. In the first stage n simple exponential smoothing models with different smoothing parameters are selected. In the second stage the n selected models are combined using different strategies to obtain forecasts.

Let us first analyze the selection stage. The models are selected out of a candidate model space S consisting of a set of simple exponential smoothing models with different smoothing parameters. For example:



(a) Estimated l_0 and α



(b) $l_0 = x_1$ and $\alpha = 0.6$

Fig. 1. In-sample one-step ahead forecasts.

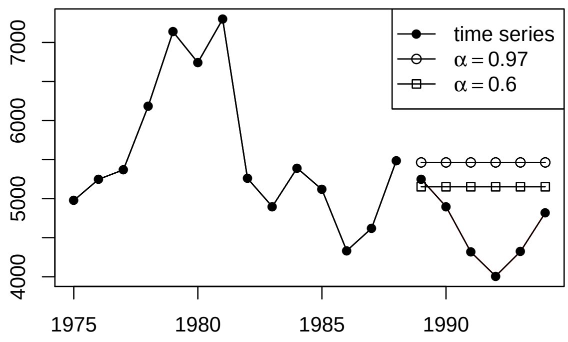


Fig. 2. Forecasts of two models with different smoothing parameters.

$$S = \{M_{0.05}, M_{0.1}, M_{0.15}, \ldots, M_{0.95}\}$$

where M_α represents the model with smoothing parameter α. Given a time series, t_s, the selection stage selects from S the n models that best fit t_s. As selection criterion, i.e. to assess how well a candidate model fits a time series, we use the SSEs. Given a model M_α and a time series t_s we compute the SSEs of the in-sample one-step ahead forecasts of model M_α as specified in equation (3). The initial prediction parameter of a model M_α, l_0, is estimated from the data so that SSE is minimized for the smoothing factor α. The SSEs for all the candidate models in S are computed and the n models with the smallest SSEs are selected. Figure 3 shows an algorithm for the selection phase.

Once the n models have been selected the forecasts can be computed. The combined forecast for horizon h is computed as a combination of the forecasts for horizon h of the n selected models. The combinations we have experimented with are: the average, the median and a weighted average. In the weighted average the weights sum to 1 and are inversely proportional to the SSEs of the models, so that models that fit worse the series receive lower weights than models that fit better the series.

Let us see an example of combination. For the time series shown in Figure 1, the models with the $n = 3$ smallest SSEs are $M_{0.95}$, $M_{0.9}$ and $M_{0.85}$. These models produce the out-of-sample one-step ahead forecasts 5444.77, 5399.9 and 5354.84 respectively. If the median is used to combine the forecasts, then the one-step ahead forecast of our combined method would be 5399.9—see Figure 4.

```
set<Model> selectModels (TimeSeries t_s, int n) {
  set<Model> M;                    // the selected models
  α = 0.05
  while (α ≤ 0.95) {
    Model m(α, t_s);
    if (M.size() < n)
      M.add (m);
    else if (m.SSE() < M.maxSSE())
      M.replace(m); // replace model in M with max SSE by m
    α = α + 0.05
  }
  return M;
}
```

Fig. 3. Function for selecting models in a C++-like syntax.

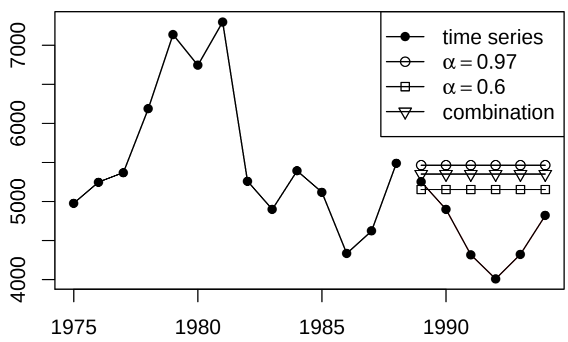


Fig. 4. The combined forecasts.

Table 1. MAPE of the benchmarking methods

method	MAPE	method	MAPE
THETA	26.25	AAM1	27.77
ForcX	26.36	AutoBox2	27.78
ForecastPro	26.67	Flors-Pearc2	28.11
SES	**26.73**	Flors-Pearc1	28.20
SMARTFCS	26.83	AutoBox1	28.76
COMB S-H-D	27.14	RBF	29.02
DAMPEN	27.29	WINTER	29.25
Auto-ANN	27.31	HOLT	29.25
PP-Autocast	27.50	AutoBox3	29.58
SINGLE	27.60	ARARMA	30.43
THETAsm	27.63	NAIVE2	33.62
AAM2	27.72	ROBUST-Trend	39.70
B-J auto	27.72		

Table 2. MAPE values of the combined models

	[0.05,0.95]	[0.025,0.975]
Median(2)	26.48	26.59
Median(3)	26.36	26.48
Median(4)	26.45	26.48
Median(5)	26.50	26.44
Mean(2)	26.48	26.59
Mean(3)	26.44	26.51
Mean(4)	26.53	26.52
Mean(5)	26.61	26.50
Weighted average(2)	26.48	26.59
Weighted average(3)	26.44	26.52
Weighted average(4)	26.53	26.52
Weighted average(5)	26.60	26.50

4 Experimentation

Our experiments have been conducted in the free statistical software R. For specific work with time series the following R packages have been used:

- The `Mcomp` package [10] that contains the 3003 time series of the M3-competition [7] and the 1001 series of the M1-competition. These series have been used to assess the predictive performance of our approach based on combined forecasts. The package also contains the forecasts of the different methods tested during the competition.
- The `forecast` package [2, 11, 12], which contains automatic time series forecasting algorithms, including the exponential smoothing family of forecasting methods.

The first step we have taken is to select the series without trend and seasonality from the 3003 series of the M3-competition. This has been done using the *ets* function from the forecast package. Given a time series, the *ets* function automatically selects one exponential smoothing method out of thirty exponential smoothing methods for forecasting the series, if the method selected is simple exponential smoothing we consider that the series has no trend or seasonality. This way 857 series out of the 3003 series of the M3-competition have been selected for assessing our approach.

The 857 series contain yearly, quarterly and monthly data from different fields such as Finance, Industry and Demographic. Every series contains in-sample observations (or training data) as well as out-of-sample observations (or test data). The forecast horizon of the out-of-sample data range from 4 to 18 depending on the series.

In order to measure the predictive performance of a forecasting technique, we have used it to predict every out-of-sample observation in each one of the 857 selected series. In order to measure forecast accuracy we have used MAPE, that is, the Mean Absolute Percentage Error of all the forecasts:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{100(x_i - \hat{x}_i)}{x_i}\right|$$

This measure is only suitable when the out-of-sample data are strictly positive, as it is the case with the series of the M3-competition.

For benchmarking purposes we have used the results of the methods tested in the M3 competition. Table 1 shows the forecast accuracy of these methods, in terms of the MAPE measure, when they are applied to forecast the out-of-sample observations of the 857 selected time series. The method labeled as SES does not belong to the M3-Competition. It corresponds to the *ses function* of the forecast package. This function fits a simple exponential smoothing model to a time series. The parameters α and l_0 are estimated from the data as suggested in Section 2—i.e., minimizing the sum of squared errors of the in-sample one-step forecasts. The goal of our combined forecast approach is to outperform this method using a combination of the forecasts of multiple simple exponential smoothing models with different smoothing factors.

Table 2 shows the forecast accuracy of the combined methods. The first column specifies what kind of combination (median, average or weighted average) has been used and, between parentheses, the number of models (n) combined. The first row specifies the smoothing parameters of the candidate models. $[0.05, 0.95]$ represents the candidate model space

$$S = \{M_{0.05}, M_{0.1}, M_{0.15}, \ldots, M_{0.95}\}$$

and $[0.025, 0.975]$ represents

$$S = \{M_{0.025}, M_{0.05}, M_{0.075}, \ldots, M_{0.975}\}$$

that is, in the first case the alphas are incremented by 0.05 and in the second case by 0.025.

The results of the combined forecasts are very encouraging. The most outstanding fact is that all combinations outperform the single exponential smoothing method with estimated parameters. In general, the results are better when $n = 3$, that is, when three models are combined. Experimental results also show that the median is the most effective way of combining the forecasts. The best result is obtained by the median of three models: a MAPE of 26.36, almost 0.4 lower than the MAPE obtained by the single exponential smoothing method with estimated parameters (26.73) and very close to the accuracy of the best method—Theta—in the M3 competition. This represents a significant improvement in forecast accuracy that can save a lot of money and/or resources.

5 Conclusions

In this paper we have experimented with using combined forecasting to improve the predictive performance of simple exponential smoothing. Instead of estimating the parameters of the model that best fits a time series, we have selected multiple models with different smoothing factors, using as selection criterion the sum of squared errors of the in-sample one-step ahead forecasts. Different combinations of forecasts: average, median and weighted average have been used. Experimental results, using the M3-competition time series, show that the combined forecasts are very effective, outperforming the results obtained by a simple exponential smoothing model whose parameters are estimated from the time series. This way we conclude that forecasting practitioners should consider using a combined forecasting when the chosen technique for forecasting is simple exponential smoothing.

Acknowledgements

This work has been partially supported by the Spanish Ministry of Science and Technology and the European Union by means of the ERDF funds, under the research Project TIN 2012-33856.

References

1. Peña, D., Tiao, C., Tsay, S.: A course in time series analysis. Wiley (2001)
2. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, 27(3), 1–22 (2008)
3. Witten, I.H., Frank, E., Hall, M.A: Data Mining: Practical Machine Learning Tools and Techniques (3rd edition), Morgan Kaufmann Publishers Inc., San Francisco CA, (2011)
4. Bates,J., Granger.,C.: The combination of forecasts, Operational Research Quarterly, 20, 451–468 (1969)
5. Hibon,M., Evgeniou,T.: To combine or not to combine: selecting among forecasts and their combinations, International Journal of Forecasting, 21, 15–24, (2005)

6. Zhang, G.: Time series forecasting using a hybrid arima and neural network model, Neurocomputing, 50, 159–175 (2003)
7. Makridakis,S., Hibon,M.: The m3-competition: results, conclusions and implications, International Journal of Forecasting, 16(4), 451–476 (2000)
8. Gardner. Jr., E.S.: Exponential smoothing: The state of the artpart {II}. International Journal of Forecasting, 22(4), 637–66 (2006)
9. Kolassa, S.: Combining exponential smoothing forecasts using Akaike weights, International Journal of Forecasting, 27(2), 238–251 (2011)
10. Hyndman, R.J. Mcomp: Data from the m-competitions. R package version 2.04, `http://CRAN.Rproject.org/package=Mcomp`
11. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice, Otexts, 2014, `http://otexts.org/fpp`
12. Hyndman, R.J.: forecast: Forecasting functions for time series and linear models. R package version 5.4., 2014 `http://CRAN.Rproject.org/package=forecast`.
13. Assimakopoulos, V., Nikolopoulos, K.: The theta model: a decomposition approach to forecasting, International Journal of Forecasting, 16(4), 521–530 (2000)