International Journal of Forecasting 32 (2016) 303–312

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation

Christoph Bergmeir[a,*], Rob J. Hyndman[b], José M. Benítez[c]

[a] Faculty of Information Technology, Monash University, Melbourne, Australia

[b] Department of Econometrics & Business Statistics, Monash University, Melbourne, Australia

[c] Department of Computer Science and Artificial Intelligence, E.T.S. de Ingenierías Informática y de Telecomunicación, University of Granada, Spain

A R T I C L E   I N F O

Keywords:
Bagging
Bootstrapping
Exponential smoothing
STL decomposition

A B S T R A C T

Exponential smoothing is one of the most popular forecasting methods. We present a technique for the bootstrap aggregation (bagging) of exponential smoothing methods, which results in significant improvements in the forecasts. The bagging uses a Box–Cox transformation followed by an STL decomposition to separate the time series into the trend, seasonal part, and remainder. The remainder is then bootstrapped using a moving block bootstrap, and a new series is assembled using this bootstrapped remainder. An ensemble of exponential smoothing models is then estimated on the bootstrapped series, and the resulting point forecasts are combined. We evaluate this new method on the M3 data set, and show that it outperforms the original exponential smoothing models consistently. On the monthly data, we achieve better results than any of the original M3 participants.

1. Introduction

After more than 50 years of widespread use, exponential smoothing is still one of the most practically relevant forecasting methods available (Goodwin, 2010). This is because of its simplicity and transparency, as well as its ability to adapt to many different situations. It also has a solid theoretical foundation in ETS state space models (Hyndman & Athanasopoulos, 2013; Hyndman, Koehler, Ord, & Snyder, 2008; Hyndman, Koehler, Snyder, & Grose, 2002). Here, the acronym *ETS* stands both for ExponenTial Smoothing and for Error, Trend, and Seasonality, which are the three components that define a model within the ETS family.

Exponential smoothing methods obtained competitive results in the M3 forecasting competition (Koning, Franses, Hibon, & Stekler, 2005; Makridakis & Hibon, 2000), and the `forecast` package (Hyndman, 2014; Hyndman & Khandakar, 2008) in the programming language R (R Core Team, 2014) means that a fully automated software for fitting ETS models is available. Thus, ETS models are both usable and highly relevant in practice, and have a solid theoretical foundation, which makes any attempts to improve their forecast accuracy a worthwhile endeavour.

Bootstrap aggregating (bagging), as proposed by Breiman (1996), is a popular method in machine learning for improving the accuracy of predictors (Hastie, Tibshirani, & Friedman, 2009) by addressing potential instabilities. These instabilities typically stem from sources such as data uncertainty, parameter uncertainty, and model selection uncertainty. An ensemble of predictors is estimated on bootstrapped versions of the input data, and the output of the ensemble is calculated by combining (using

* Correspondence to: Faculty of Information Technology, P.O. Box 63 Monash University, Victoria 3800, Australia. Tel.: +61 3 990 59555.
E-mail address: christoph.bergmeir@monash.edu (C. Bergmeir).

http://dx.doi.org/10.1016/j.ijforecast.2015.07.002
0169-2070/

the median, mean, trimmed mean, or weighted mean, for example), often yielding better point predictions. In this work, we propose a bagging methodology for exponential smoothing methods, and evaluate it on the M3 data. As our input data are non-stationary time series, both serial dependence and non-stationarity have to be taken into account. We resolve these issues by applying a seasonal-trend decomposition based on loess (STL, Cleveland, Cleveland, McRae, & Terpenning, 1990) and a moving block bootstrap (MBB, see, e.g., Lahiri, 2003) to the residuals of the decomposition.

Specifically, our proposed method of bagging is as follows. After applying a Box–Cox transformation to the data, the series is decomposed into trend, seasonal and remainder components. The remainder component is then bootstrapped using the MBB, the trend and seasonal components are added back in, and the Box–Cox transformation is inverted. In this way, we generate a random pool of similar bootstrapped time series. For each of these bootstrapped time series, we choose a model from among several exponential smoothing models, using the bias-corrected AIC. Then, point forecasts are calculated using each of the different models, and the resulting forecasts are combined using the median.

The only related work that we are aware of is the study by Cordeiro and Neves (2009), who use a sieve bootstrap to perform bagging with ETS models. They use ETS to decompose the data, then fit an AR model to the residuals, and generate new residuals from this AR process. Finally, they fit the ETS model that was used for the decomposition to all of the bootstrapped series. They also test their method on the M3 dataset, and have some success for quarterly and monthly data, but overall, the results are not promising. In fact, the bagged forecasts are often not as good as the original forecasts applied to the original time series. Our bootstrapping procedure works differently, and yields better results. We use STL for the time series decomposition, MBB to bootstrap the remainder, and choose an ETS model for each bootstrapped series. Using this procedure, we are able to outperform the original M3 methods for monthly data in particular.

The rest of the paper is organized as follows. In Section 2, we discuss the proposed methodology in detail. Section 3 presents the experimental setup and the results, and Section 4 concludes the paper.

2. Methods

In this section, we provide a detailed description of the different parts of our proposed methodology, namely exponential smoothing, and the novel bootstrapping procedure involving a Box–Cox transformation, STL decomposition, and the MBB. We illustrate the steps using series M495 from the M3 dataset, which is a monthly series.

2.1. Exponential smoothing

The general idea of exponential smoothing is that recent observations are more relevant for forecasting than older observations, meaning that they should be weighted more highly. Accordingly, simple exponential smoothing, for example, uses a weighted moving average with weights that decrease exponentially.

Starting from this basic idea, exponential smoothing has been expanded to the modelling of different components of a series, such as the trend, seasonality, and remainder components, where the trend captures the long-term direction of the series, the seasonal part captures repeating components of a series with a known periodicity, and the remainder captures unpredictable components. The trend component is a combination of a level term and a growth term. For example, the Holt–Winters purely additive model (i.e., with additive trend and additive seasonality) is defined by the following recursive equations:

$$\begin{aligned}
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-m+h_m^+}.
\end{aligned}$$

Here, ℓ_t denotes the series level at time t, b_t denotes the slope at time t, s_t denotes the seasonal component of the series at time t, and m denotes the number of seasons in a year. The constants α, β^*, and γ are smoothing parameters in the $[0, 1]$-interval, h is the forecast horizon, and $h_m^+ = \lfloor (h-1) \bmod m \rfloor + 1$.

There is a whole family of ETS models, which can be distinguished by the type of error, trend, and seasonality each uses. In general, the trend can be non-existent, additive, multiplicative, damped additive, or damped multiplicative. The seasonality can be non-existent, additive, or multiplicative. The error can be additive or multiplicative; however, distinguishing between these two options is only relevant for prediction intervals, not point forecasts. Thus, there are a total of 30 models with different combinations of error, trend and seasonality. The different combinations of trend and seasonality are shown in Table 1. For more detailed descriptions, we refer to Hyndman and Athanasopoulos (2013), Hyndman et al. (2008), and Hyndman et al. (2002).

In R, exponential smoothing is implemented in the `ets` function from the `forecast` package (Hyndman, 2014; Hyndman & Khandakar, 2008). The different models are fitted to the data automatically; i.e., the smoothing parameters and initial conditions are optimized using maximum likelihood with a simplex optimizer (Nelder & Mead, 1965). Then, the best model is chosen using the bias-corrected AIC. We note that, of the 30 possible models, 11 can lead to numerical instabilities, and are therefore not used by the `ets` function (see Hyndman & Athanasopoulos, 2013, Section 7.7, for details). Thus, `ets`, as it is used within our bagging procedure, chooses from among 19 different models.

2.2. The Box–Cox transformation

This is a popular transformation for stabilizing the variance of a time series, and was originally proposed by Box and Cox (1964). It is defined as follows:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$