Discussiones Mathematicae
Probability and Statistics 31 (2011) 121–139
doi:10.7151/dmps.1143

COMPUTATIONAL INTENSIVE METHODS FOR PREDICTION AND IMPUTATION IN TIME SERIES ANALYSIS

Maria Manuela Neves

CEAUL and Instituto Superior de Agronomia
Technical University of Lisbon
Tapada da Ajuda, 1349–017 Lisboa, Portugal
e-mail: manela@isa.utl.pt

and

Clara Cordeiro

Mathematics Department, Faculty of Science and Technology,
University of Algarve
Campus Gambelas, 8005–139 Faro, Portugal
e-mail: ccordei@ualg.pt

To Professor M. Ivette Gomes, an Extremal Friendship.

Abstract

One of the main goals in times series analysis is to forecast future values. Many forecasting methods have been developed and the most successful are based on the concept of exponential smoothing, based on the principle of obtaining forecasts as weighted combinations of past observations. Classical procedures to obtain forecast intervals assume a known distribution for the error process, what is not true in many situations. A bootstrap methodology can be used to compute distribution free forecast intervals. First an adequately chosen model is fitted to the data series. Afterwards, and inspired on sieve bootstrap, an AR(p) is used to filter the series of the random component, under the stationarity hypothesis. The centered residuals are then resampled and the initial series is reconstructed. This methodology will be used to obtain forecasting intervals and for treating missing data, which often appear in a real time series. An automatic procedure was developed in R language and will be applied in simulation studies as well as in real examples.

Keywords: bootstrap, forecast intervals, missing data, time series analysis.

2010 Mathematics Subject Classification: 62G32, 62E20, 65C05.

1. Motivation and scope of the paper

Time series analysis deals with records collected over time. One distinguishing feature in time series data is that time order is important and the records are usually dependent. Depending on the application, data may be collected hourly, daily, weekly, monthly or yearly, etc. Time series arise in many different contexts. Its impact on scientific, economic and social applications is well recognized by the large list of fields in which important time series problems may arise. Just to refer a few we can mention economics (daily stock market, monthly unemployment figures,...), social sciences (populations series of birthrates, school enrollments,...), medicine (blood pressure measurements,...), physical sciences (meteorological data, geophysics data,...), environmental sciences (global warming data, levels of pollution,...), etc. Time series can show different displays, see Figure 1 for some examples.

In time series analysis many challenging topics can be pointed out:

- Obtain point and interval forecasting, i.e., consider the time series to gain some insight into the future. This is one of the main objectives in time series analysis.

- Deal with the existence of missing values which causes difficulties in producing reliable and sound statements about the variables concerned. It happens in many environmental situations (e.g. time series on water quality data are sometimes interrupted due to several causes: changes in analytical methodology, miscommunication, (temporary) financial cuts, etc.).

- Predict extreme or even rare events that can occur beyond the available data. This is crucial in many environmental situations (e.g. daily levels of a river, hourly ozone concentration, etc). Here we are mainly

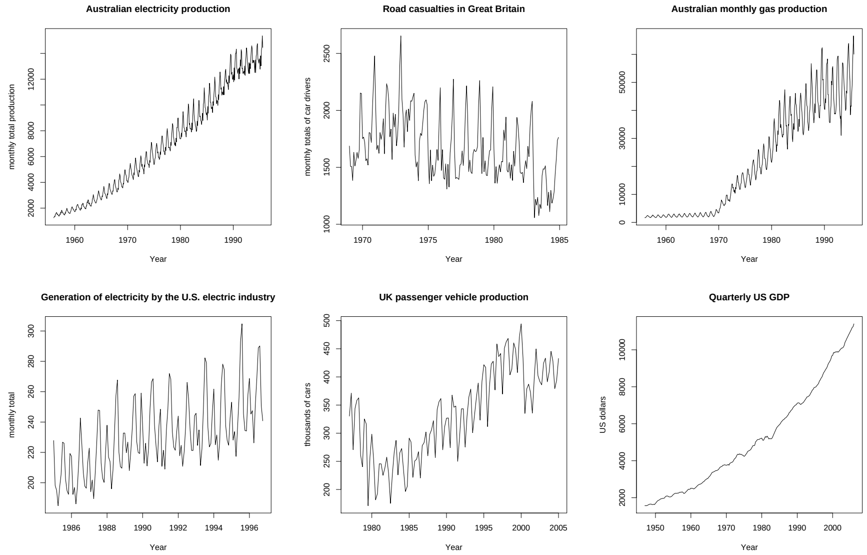


Figure 1. Some examples of time series from R packages fma, datasets and expsmooth, showing different behaviors.

interested in modeling and predicting the behavior of extreme (often maximum) values of the time series (e.g. for security reasons).

- Consider ways of dealing with "short" time series. Nonparametric and semi-parametric techniques constitute recent areas of research helped by the increasing possibilities of computers to give answers in situations where classical methods cannot be applied.
- Develop procedures for multivariate time series (that appear, for example, in climatology or meteorology, where the phenomena under study are very complex and several variables and several scales are sometimes involved), where less research has been developed.
- Develop adequate procedures for time series recorded at not equally spaced points.

After these general ideas on several important questions that arise linked to time series, basic concepts in time series analysis will be remembered in Section 2. Exponential smoothing methods will be briefly explained in Section 3 and resampling methods in time series will appear in Section 4.

Finally a computational procedure for prediction and also to detect and to impute missing data in time series will be explained in Section 5 and some comparative studies will be done in Section 6. Final Comments and the References will conclude this work.

2. Basic concepts in time series analysis

A time series is a set of observations $\{y_{t_1}, y_{t_2}, \ldots, y_{t_N}\}$ each one recorded at a specific time $t_1, t_2, \ldots, t_N$. A time series is said to be discrete (the case we are going to consider here) if the set T_0 of times at which observations are made is a discrete set. Usually the records are done at equally spaced times and the time series is then represented by

$$y_t, \ t \in T_0 = \{1, 2, \ldots, N\} \text{ or } T_0 = \boldsymbol{N} \text{ or } T_0 = \boldsymbol{Z}.$$

If the observations are made continuously in time, the time series is said to be continuous.

The analysis of data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. Indeed, most standard statistical techniques assume that the available data can be regarded, at least approximately, as an independent random sample from a population of interest. This is a critical assumption for the construction of standard hypothesis tests and confidence intervals. One distinguishing feature in time series is that the records are usually dependent. Dependence between successive observations in a time series is referred to as "autocorrelation". Time series analysis deals with methods specially designed for autocorrelated data.

The main objectives in a time series analysis, see Chatfield (2004) for a complete description, are: description, explanation, prediction and control.

- Description refers to the first step in the analysis. It begins by looking at the data and involves a variety of graphical displays. Graphical representation of a time series allows to look for some patterns that the time series exhibits, such as upward or downward movement (trend) or a pattern that repeats (seasonal variation). The calculation of simple descriptive measures of the main properties is another important step.

- Explanation intends to understand and interpret the mechanisms that generated the data. To develop mathematical models that provide plausible descriptions for sample data is one of the primary objectives of a time series analysis. *"Different purposes of the analysis may also dictate the use of different models. For example, a model that provides a good fitting and admits nice interpretation is not necessarily good for forecasting"*, Bickel *et al.* (2003).

- Prediction deals with the extrapolation for the future. These extrapolations are often used to assess the risk of future adverse events or to justify changing of policies, for example.
- Control is an important objective mainly, for example, when the time series is measuring the "quality" of manufacturing processes.

A time series is a realization of a stochastic process $\{Y_t,\ t \in \mathcal{T}\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and with values in $(\boldsymbol{R}^n, \mathcal{B}(\boldsymbol{R}^n))$. If $n = 1$ it is a univariate stochastic process, if $n > 1$ it is a multivariate stochastic process.

A fundamental task is to unveil the probabilistic law that governs the observed time series such that we can understand the underlying dynamics. Many stochastic processes have been developed in order to explain that underlying dynamics. Let us refer to some examples: White noise, Moving Averages, Random Walk, Autoregressive Processes, ARMA processes, ARIMA processes, are perhaps the most well known.

A time series can be thought as a combination of some components: trend (T), the long term direction of the series; the seasonal component, (S) that is a pattern that repeats with a known periodicity; the cycle (C) is a pattern that repeats with some regularity but with unknown and changing periodicity and the error (ϵ) that is the unpredictable component of the series. Those components can be combined in several ways, giving different models, for example:

- A purely additive model, that can be expressed as:

$$y_t = T_t + S_t + C_t + \epsilon_t.$$

- A purely multiplivative model, that can be expressed as:

$$y_t = T_t \times S_t \times C_t \times \epsilon_t.$$

- A mixed model, that can be, for example:

$$y_t = (T_t + S_t) \times C_t + \epsilon_t.$$

Another point to be careful with is that most of the probability theory of time series is concerned with stationary time series and for this reason many procedures require to turn a non-stationary series to a stationary one.

3. Exponential smoothing methods

Forecasting future values of a time series is one of the main objectives in the analysis. Forecasting methods have been developed based on well known models: AR, ARMA, ARIMA, SARIMA, etc.

In the decade of 1950 another class of forecasting methods appeared. These methods are based on the concept of exponential smoothing, i.e., methods having the property that forecasts are weighted combinations of past observations, with recent observations given relatively more weight than older observations. The name "exponential smoothing" reflects the fact that the weights decrease exponentially as the observations get older.

Exponential smoothing (EXPOS) refers then to a set of methods that, in a versatile way, can be used to model and to obtain forecasts.

The best known exponential smoothing methods, Hyndman *et al.* (2008), are:

- Simple exponential smoothing — Suppose we have observed data up to and including time $t-1$, and we wish to forecast the next value of our time series, $\widehat{y}_t$. The method of simple exponential smoothing, due to Brown (1959) takes the forecast for the previous period and adjusts it using the forecast error. So, with α a constant between 0 and 1, the forecast for the next period is

  $$\widehat{y}_{t+1} = \widehat{y}_t + \alpha(y_t - \hat{y}_t) \Longleftrightarrow \widehat{y}_{t+1} = \alpha y_t + (1-\alpha)\widehat{y}_t.$$

  By developing the relation above it is easy to see that $\widehat{y}_{t+1}$ represents a weighted moving average of all past observations with the weights decreasing exponentially.

- Holt's linear trend — Holt (1957) extended the simple exponential smoothing procedure to linear exponential smoothing to allow forecasting of data with trends. The forecast for this method is found using two smoothing constants, α and β (with values between 0 and 1) and three equations:

- Level $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$;
- Growth $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1})$;
- Forecast $\hat{y}_{t+h|t} = l_t + b_t h$.

Here l_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope (growth) of the series at time t. This procedure needs the parameters initialization and estimation, see Hyndman *et al.* (2008) for suggestions.

- Holt-Winters Trend and Seasonality Method — Holt (1957) proposed a method for seasonal data. Later, Winters (1960) improved it. The method is based on three smoothing equations for level, trend and seasonality. For additive seasonality the equations are:

  - Level $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$;
  - Growth $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1})$;
  - Seasonal $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m})$;
  - Forecast $\hat{y}_{t+h|t} = l_t + b_t h + s_{t-m+h_m^+}$;

$h_m^+ = [(h-1)modm] + 1$ and parameters (α, β, γ) are usually restricted to lie between 0 and 1.

Gardner and Mackenzie (1985) proposed a modification of Holt's linear and Holt-Winters to allow the "damping" of trends, i.e., the growth is dampened by a factor of ϕ for each additional future time period. For example, in Holt's linear, the level will become

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1}).$$

Pegel's (1969) classified exponential smoothing methods regarding the trend and seasonal pattern that a series reveals as: none, additive (linear) or multiplicative (nonlinear). Since then, many researchers such as Gardner (1985), Hyndman *et al.* (2002), Taylor (2003) have investigated and developed the EXPOS models. Table 1 resumes the fifteen possibilities of exponential smoothing (ignoring the error component).

For example (N, N) stands for the simple exponential smoothing and (A, N) stands for Holts linear method. Hyndman *et al.* (2008) provided a state space formulation for all models in the classification of Table 1. For each method in the framework, additive error and multiplicative error

Table 1. The exponential smoothing models.

	Seasonal Component		
Trend	N	A	M
Component	(None)	(Additive)	(Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
Ad (Additive damped)	Ad,N	Ad,A	Ad,M
M (Multiplicative)	M,N	M,A	M,M
Md (Multiplicative damped)	Md,N	Md,A	Md,M

versions are considered. The state space model usually consists of two sets of equations: the observation equation (1) and the state equation (2),

$$y_t = \mathbf{w}'\mathbf{x}_{t-1} + \varepsilon_t, \tag{1}$$

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + \mathbf{g}\varepsilon_t, \tag{2}$$

with $t = 1, 2, \ldots$, where y_t is the observation in time t, $\mathbf{x}_t$ is a "state vector" containing unobserved components (level, trend and seasonality), $\{\varepsilon_t\}$ is a white noise series and F, $\mathbf{g}$ and $\mathbf{w}$ are coefficients. The first equation (1) relates the observable time series value y_t to a random k-vector $\mathbf{x}_{t-1}$ of unobservable components from the previous period. $\mathbf{w}$ is a fixed k-vector. F is a fixed $k \times k$ matrix and $\mathbf{g}$ is a k-vector of smoothing parameters. For more details see Hyndman *et al.* (2008). The estimates of the exponential smoothing parameters are obtained by minimizing the mean squared error (MSE) of the one-step-ahead forecasts errors over the fitted period. The model selection is made using the Akaike's criterion (AIC). This model selection criterion is preferable when compared to other criteria because of the parsimonious model penalty, see Hyndman *et al.* (2002) for more details.

4. Resampling methods in time series

Among resampling techniques, bootstrap is perhaps the most popular one. It is a computational method for estimating the distribution of an estimator or test statistic by resampling from the data. Under conditions that hold in a wide variety of applications, the bootstrap provides approximations to distributions of statistics, coverage probabilities of confidence intervals and accurate rejection probabilities of tests. The procedure was devised for an i.i.d. situation and it usually fails for dependent observations.

In context of stationary time series two different bootstrap methods have been proposed. Perhaps the best-known for time-series data is the block bootstrap. It consists of dividing the data into blocks of observations and sampling the blocks randomly with replacement, as in the independent case. The blocks may be non-overlapping, Hall (1985) and Carlstein (1986) or overlapping, Hall (1985), Knsch (1989) and Politis and Romano (1992). Afterwards the resampled blocks are joined in order to reconstruct the series.

However, if the time series process is driven from i.i.d. innovations another way of resampling can be considered. The classical bootstrap derived for i.i.d. samples can easily be extended to the dependent case.

Another procedure, the sieve bootstrap, was proposed by Bühlmann (1997) for dependent observations and extended by Alonso *et al.* (2002, 2003) for constructing prediction intervals in stationary time series. The scheme of the sieve approach is the following:

Step 1. Fit an AR(p) using the AIC criterion;

Step 2. Obtain the AR residuals;

For B replicates

Step 3. Resample the centered residuals;

Step 4. Obtain a new series by recursion using the resampled series and the autoregressive coefficients from **Step 1**;

Step 5. Fit an AR(p) to the new series;

Step 6. Obtain the forecasts using the previous model.

In previous works, Cordeiro and Neves (2006, 2007, 2008) studied and analyzed the possibility of joining EXPOS methods and the bootstrap methodology. From those studies the idea behind the sieve bootstrap, Bühlmann (1997), suggested the connection of those two procedures. In a few words, the sieve bootstrap considers first an autoregressive process that is fitted to a stationary time series. Considering a model-based approach, which resamples from approximately i.i.d. residuals, the classical bootstrap methodology was applied to the centered residuals. The bootstrap proposed by Bühlmann (1997) was extended for obtaining prediction intervals in stationary time series, Alonso *et al.* (2002, 2003). Following Bühlmann (1997) and Lahiri (2003), validity and accuracy of IID-innovation bootstrap is well studied.

5. Computational procedure for prediction and imputation

A first computational algorithm was constructed using four models for fitting to the time series: single exponential smoothing, Holts linear and Holt-Winters with additive and multiplicative seasonality. Nowadays it considers thirty exponential smoothing methods and it consists of an automatic procedure in R language. This procedure was named Boot.EXPOS. The idea is to select the most adequate EXPOS model by using the AIC criterion and obtain the residuals. The error component is isolated and investigated regarding its stationarity using the Augmented Dickey-Fuller test. If it is not compatible with this hypothesis, data transformation is required. If there is some stationarity evidence, the residual sequence is filtered by an autoregressive model, autoregressive coefficients are estimated and innovations are obtained. In the context of AR models the bootstrap can be conducted by resampling the centered residuals and then generating a data set, using the estimated coefficients and the resampled residuals. The EXPOS fitted values and the reconstructed series are used to obtain a sample path of the data. Forecasts are obtained using the initial EXPOS model. The bootstrap process is repeated B times and information is kept into a matrix. An "optimal" point forecast is obtained by taking the average of each column. The procedure also includes testing for stationarity and Box-Cox transformations. The performance of our procedure was evaluated through the forecasts obtained for a given period in a very large set of time series.

5.1. A sketck of the algorithm

For a given time series $\{y_1, \ldots, y_n\}$ select the "best" EXPOS model (Table 1) using the AIC criterion. Any good model should yield residuals that do not show a significant pattern. It is rare to discuss white noise in this context because there is frequently some pattern left in the residuals, see DeLurgio (1998). In order to model such left-over patterns and in case of stationarity, an autoregressive model is used to filter the EXPOS residuals series. Thus, in order to apply the residual-based bootstrap discussed in Section 4, a stationary series is required. The algorithm that joins the EXPOS methods with the bootstrap approach is summarized as follows:

Step 0. Select an EXPOS model by AIC criterion, $\theta_0 = (\alpha, \beta, \gamma, \phi)$, $\hat{\boldsymbol{y}} = \{\hat{y}_1, \dots, \hat{y}_n\}$ and the residuals $\{r_1, \dots, r_n\}$;

Boot.EXPOS

Step 1. Fit an AR(p) to the residual sequence using the AIC criterion;

Step 2. Obtain the AR residuals;

For B replicates

Step 3. Resample the centered residuals;

Step 4. Obtain a new series by recursion using the resampld series and the autoregressive coefficients from **Step 1**;

Step 5. Join the fitted values $\hat{\boldsymbol{y}}$ (**Step 0**) to the previous series;

Step 6. Forecast the initial series using the selected model and θ_0 estimated in **Step 0**.

Statistical tests, transformations and differentiation are prepared for analysis of stationarity of the random part before the AR ajustment is done (**Step 1** of Boot.EXPOS). All the intensive computational work is performed in R software. Some R packages: **car**, **forecast**, **tseries** are used. New functions in R environment were constructed.

5.2. Measuring Forecast Errors

Large forecasting errors occur if the random component is very large or the forecasting technique is not capable of accurately predicting the trend, seasonal or cyclic components. The forecast performance is evaluated using some accuracy measures. For each value y_t of the variable of interest in time period t, the forecast error for a particular forecast $\hat{y}_t$ is $e_t = y_t - \hat{y}_t$. Several measures can be considered:

Table 2. The Accuracy measures.

Acronyms	Definition	Formula
RMSE	Root Mean Squared Error	$\sqrt{mean((y_t - \hat{y}_t)^2)}$
MAE	Mean Absolute Error	$mean(\lvert y_t - \hat{y}_t \rvert)$
MAPE	Mean Absolute Percentage Error	$mean(100 \left\lvert \frac{y_t - \hat{y}_t}{y_t} \right\rvert)$
sMAPE	Symmetric Mean Absolute P.E.	$mean(200 \frac{\lvert y_t - \hat{y}_t \rvert}{y_t + \hat{y}_t})$

6. Comparative study

In this Section, the forecasting effectiveness of the proposed procedure is evaluated. The Boot.EXPOS is used to produce forecasts on some well-known data sets and its performance is compared with other forecasting methods.

Some functions already existing in R environment, for example for the exponential smoothing selection `ets()` are used. The selection is made using the AIC criterion and during the process there is no user intervention, that is, it runs in an automatic way. For more R details see Hyndman and Khandakar(2008).

6.1. In forecasting

All time series are separated into two parts: the fitting set and the validation set. The fitting set $\{y_1, \ldots, y_{n-h}\}$ is used to find the appropriate EXPOS model and the exponential smoothing parameters estimates. The validation set $\{y_{n-h+1}, \ldots, y_n\}$ is used to evaluate the forecasting capacity using some accuracy measures. The forecasts are computed for a hold-out period $\hat{y}_n(1), \ldots, \hat{y}_n(h)$ and compared with the true values (the validation set) using criteria given in Table 2.

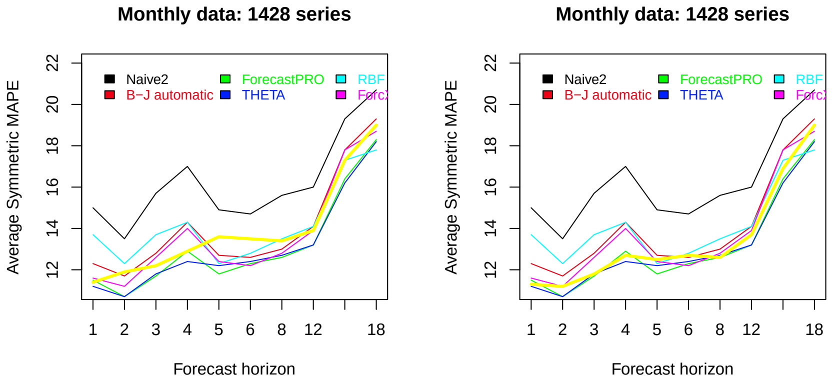
The M3 competition is a large set of 3003 series (Table 3) that is commonly used for evaluation the performance of a forecasting procedure. A different number of forecasts, depending on the categories, are requested: 6 for yearly; 8 for quarterly and "other"; 18 for monthly.

Makridakis and Hibon (2000) gives the 24 forecasting methods used in the M3 competition and the best 6 methods were: Naive2, Box-Jenkins automatic, ForecastPro, THETA, RBH and ForecastX.

Table 3. The M3 competition time series.

	Type of times series data						
Period	Demographic	Finance	Industry	Macro	Micro	OTHER	Total
Monthly	111	145	334	312	474	52	1428
OTHER	0	29	0	0	4	141	174
Quarterly	57	76	83	336	204	0	756
Yearly	245	58	102	83	146	11	645
Total	413	308	519	731	828	204	3003

Recently, Hyndman (2008) included in R the function **ets()**, that chooses the model (among those fifteen showed before, with additive and multiplicative errors for each model) that better fits the data and that makes forecasts. Boot.EXPOS procedure was then extended considering all the fifteen exponential smoothing models presented before, with additive and multiplicative errors for each model. For illustration see Figure 2 where the Symmetric Mean Absolute Percentage Error (sMAPE) is plotted for those best six methods and Boot.EXPOS (in yellow) for the M3 competition time series.



(a) First approach.

(b) Recent approach

Figure 2. Boot.EXPOS with the selection among (a) 4 EXPOS methods, (b) all the EXPOS methods.

It is visible the progress of the Boot.EXPOS. In Figure 2 (a) the EXPOS selection was among the simple exponential smoothing, the Holt's linear and Holt-Winters, with additive and multiplicative seasonal component

(Cordeiro and Neves (2009)). Recently, Figure 2 (b), the EXPOS selection was augmented by incorporating more models and the Boot.EXPOS revealed improvement in terms of forecast accuracy.

For time series in Figure 1, forecasts using the ets and the Boot.EXPOS are obtained and the accuracy measures are presented in Table 4. Once again the Boot.EXPOS showed a better performance in forecasting.

Table 4. Accuracy measures for time series in Figure 1.

						Accuracy measures		
Time series	n-h	s	h	ets	R function	RMSE	MAE	MAPE
elec	464	12	12	(M,Ad,M)	ets	348.87	305.88	2.19
					Boot.EXPOS	333.90	300.85	2.17
UKDriverDeaths	180	12	12	(M,N,A)	ets	205.63	198.49	14.68
					Boot.EXPOS	84.93	67.79	4.88
gas	464	12	12	(M,Md,M)	ets	2773.72	2097.73	4.22
					Boot.EXPOS	2354.81	1929.19	3.88
uselec	130	12	12	(M,N,M)	ets	5.68	4.35	1.72
					Boot.EXPOS	4.03	3.04	1.20
ukcars	105	4	8	(A,N,A)	ets	19.46	16.05	3.95
					Boot.EXPOS	15.58	11.56	2.88
usgdp	229	4	8	(A,Ad,N)	ets	59.08	43.12	0.38
					Boot.EXPOS	38.70	24.98	0.22

6.2. Forecast Intervals

Let F_h be the empirical distribution function of the $\{\hat{y}^*_{bh}, b = 1, \ldots, B\}$. The $(1-\alpha) \times 100\%$ confidence intervals are given by

$$[F_h^{-1}(\alpha/2), F_h^{-1}(1-\alpha/2)].$$

For a 95% confidence interval and $B = 1000$ replications, the percentiles are $F_h^{-1}(0.025) = \hat{y}^{*(25)}_{bh}$, $F_h^{-1}(0.925) = \hat{y}^{*(975)}_{bh}$. So in what concerns forecast intervals, they are obtained with the percentile bootstrap method with 1000 replicas. For the time series in Figure 1 the forecast intervals obtained are plotted in Figure 3.

As it can be seen the forecasting intervals using the proposed procedure are narrower than those obtained with the ets.

6.3. Missing data imputation

Another way of Boot.EXPOS application is in time series with missing data. So this procedure was extended to deal with non-observable data: it detects, estimates and replaces. It is named NABoot.EXPOS. How does it work?

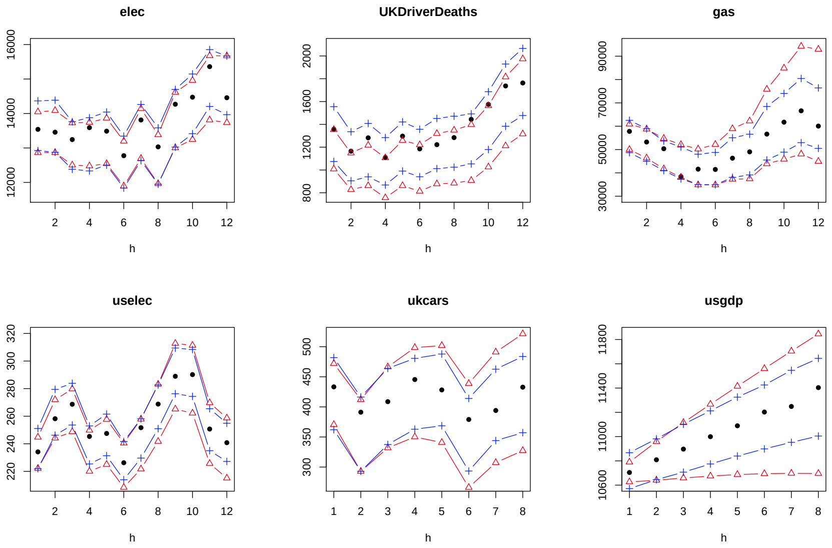


Figure 3. Forecast intervals for the time series in Figure 1.

Step 1. It starts by detecting the first missing observation;

Step 2. If the ith observation is missing (and also for the consecutive observations) the Boot.EXPOS will estimate (predict) the ith observation (and the following);

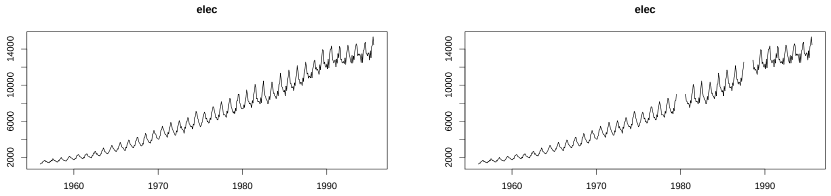
Step 3. The approach generates one or more forecasts to impute the missing value in i position and following missing values;

Step 4. Detect the next missing observation(s). If TRUE go to **Step 2**;

Step 5. The procedure finishes when the time series is complete.

In order to compare the performance of our procedure, we selected two well known R functions devised for inputing missing values. We chose na.interp(), that uses linear interpolation and amelia(), that uses the bootstrap with the EM algorithm. Figure 4 shows a complete time series (a) and the same times series after being randomly removed some observations (b).

The missing data were estimated using our method and the two methods, available in R, just mentioned. The imputed values can be observed in Figure 5.



(a) Time series elec. (b) Time series elec with missing data.

Figure 4. (a) The complete data and (b) the 24 (5%) missing observations.

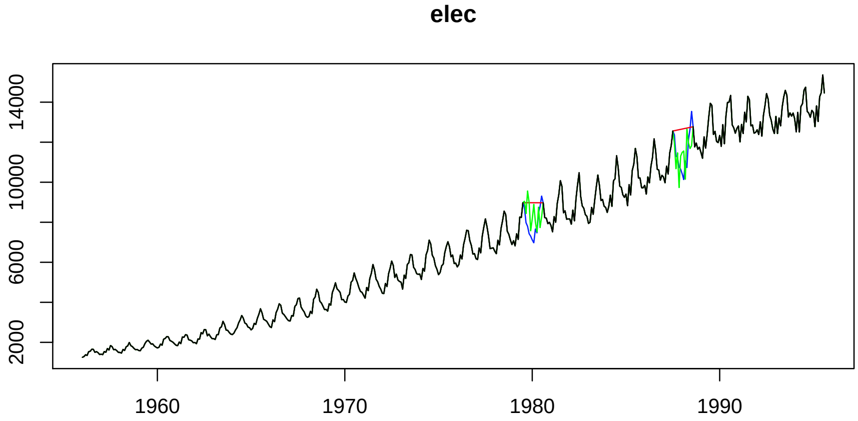


Figure 5. The na.interp, amelia and NABoot.EXPOS imputation.

Table 5 shows the correspondent accuracy measures.

Table 5. Figure 5 accuracy measures.

Funes	RMSE	MAE	MAPE
na.interp	286.34	57.73	0.62
amelia	247.14	47.25	0.50
NABoot.EXPOS	76.32	13.06	0.13

7. Closing comments

In forecasting the Boot.EXPOS has revealed a good procedure for obtaining forecasts. So the “optimal” combination of EXPOS methods and bootstrap resampling can provide more accurate forecasts. As a consequence it also produces narrower intervals when compared to the forecasting intervals achieved through the exponential smoothing models.

In missing data, an initial interpretation of the results suggests that using NABoot.EXPOS to estimate missing data can be a good option. Also the “optimal” combination of EXPOS methods and bootstrap resampling seems to provide here more accurate imputed values than the two other considered methods.

References

[1] A.M. Alonso, D. Peña and J. Romo, *Forecasting time series with sieve bootstrap*, Journal of Statistical Planning and Inference **100** (2002) 1–11.

[2] A.M. Alonso, D. Peña and J. Romo, *On sieve bootstrap prediction intervals*, Statistics & Probability Letters **65** (2003) 13–20.

[3] P. Bickel, P. Diggle, S. Fienberg and K. Krickeberg, Nonlinear Time Series (Springer Series in Statistics, New York, Springer, 2003).

[4] R.G. Brown, Nonlinear Time Series (Statistical Forecasting for inventory control, New York, McGraw-Hill, 1959).

[5] P. Bühlmann, *Sieve bootstrap for time series*, Bernoulli **3** (1997) 123–148.

[6] E. Carlstein, *The use of subseries values for estimating the variance of a general statistic from a stationary sequence*, Annals of Statistics **14** (1986) 1171–1179.

[7] C. Chatfield, The Analysis of Time Series. An Introduction (6^{th} ed. Chapman & Hall, 2004).

[8] C. Cordeiro and M.M. Neves, *The Bootstrap methodology in time series forecasting*, in: “Proceedings of CompStat2006”, J. Black and A. White (Ed(s)), (Springer Verlag, 2006) 1067–1073.

[9] C. Cordeiro and M.M. Neves, *The Bootstrap prediction intervals: a case-study*, in: ”Proceedings of the 22nd International Workshop on Statistical Modelling (IWSM2007)”, J. Castillo, A. Espinal and P. Puig (Ed(s)), (Springer Verlag, 2007) 191–194.

[10] C. Cordeiro and M.M. Neves, *Bootstrap and exponential smoothing working together in forecasting time series*, in: "Proceedings in Computational Statistics (COMPSTAT 2008)", Paula Brito (Ed(s)), (Physica-Verlag, 2008) 891–899.

[11] C. Cordeiro and M.M. Neves, *Forecasting time series with Boot.EXPOS procedures*, REVSTAT **7** (2009) 135–149.

[12] S.A. DeLurgio, Forecasting Principles And Applications (McGraw-Hill International Editions, 1998).

[13] E.S. Gardner, *Exponential smoothing: the state of the art*, J. of Forecasting **4** (1985) 1–38.

[14] E.S. Gardner and E. Mckenzie, *Forecasting trends in time series*, Management Science **31** (1985) 1237–1246.

[15] P. Hall, *Resampling a coverage pattern*, Stochastic Processes and their Applications **20** (1985) 231–246.

[16] C. Holt, Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum 52/1957, Carnegie Institute of Technology, 1957).

[17] R. Hyndman, forecast: Forecasting functions for time series (software available at http://www.robjhyndman.com/Rlibrary/forecast/, 2011).

[18] R. Hyndman and Y. Khandakar, *Automatic Time Series Forecasting: The forecast Package for Rh*, Journal of Statistical Software **27** (2008).

[19] R. Hyndman, A. Koehler, R. Snyder and S. Grose, *A state framework for automatic forecasting using exponential smoothing methods*, International Journal of Forecasting **18** (2002) 439–454.

[20] R. Hyndman, A. Koehler, J. Ord and R. Snyder, Forecasting with Exponential Smoothing: The State Space Approach (Springer-Verlag Inc, 2008).

[21] H. Künsch, *The Jackknife and the Bootstrap for General Stationary Observations*, The Annals of Statistics **17** (1989) 1217–1241.

[22] S.N. Lahiri, Resampling Methods for Dependente Data (Springer Verlag Inc, 2003).

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

[24] C.C. Pegels, *Exponential smoothing: some new variations*, Management Science **12** (1969) 311–315.

[25] D. Politis and J. Romano, *A circular block-resampling procedure for stationary data*, in: Exploring the limits of bootstrap, Lepage, R. e Billard, L. (Ed(s)), (Wiley, 1992) 263–270.

[26] R Develpment core team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/, 2011).

[27] J.W. Taylor, *Exponential smoothing with a damped multiplicative trend*, International Journal of Forecasting Management Science **19** (2003) 273–289.

[28] A. Trapletti, datasets: The R Datasets Package by A. Trapletti (package version 0.10, URL http://CRAN.R-project.org/package=datasets, 2008).

[29] A. Trapletti and K. Hornik, tseries: Time Series Analysis and Computational Finance (R package version 0.10-18, 2009).

[30] P.R. Winters, *Forecasting sales by exponentially weighted moving averages*, Management Science **6** (1960) 349–362.

Received 14 September 2011