Forecasting time series with SSA.Boot procedure

Paula Medina Maçaira*
Industrial Engineering Department, Pontifical Catholic University of Rio de Janeiro, Brazil - paulamacaira@gmail.com

Fernando Luiz Cyrino Oliveira
Industrial Engineering Department , Pontifical Catholic University of Rio de Janeiro, Brazil - cyrino@puc-rio.br

Abstract

Produce reliable and accurate forecasts is essential in any method of time series. To develop such forecasts, the proposed approach relies on the decomposition signal and noise, using the non-parametric technique Singular Spectrum Analysis (SSA). The second step is the generation of synthetic series from the noise, using the Bootstrap procedure, produce forecast *h* steps ahead for each series, using the automatic algorithms *ets* and *auto.arima*, from R software, and taking the simple average as the final prediction. This approach, called here as SSA.Boot, is applied in six real time series, available in R, and the results obtained are very satisfactory when compared to those methods applied directly to the original data.

Keywords: Singular Spectrum Analysis; Bootstrap; Time Series Analysis; Forecast.

1. Introduction

Develop a method that results in more reliable and accurate forecasts is the main goal of this paper.To do that, we propose the joint use of Singular Spectrum Analysis (SSA), Bootstrap and automatic time series procedures as an alternative forecast approach. The software used is the R (R Development Core Team 2008) and the packages *fma*, *datasets* and *expsmooth*.

The paper is organized as follows: section two describes the material and methods used for the forecasting exercise; in section three it is presented the results and discussions from the application of the proposed methodology and, finally, in section four are drawn some conclusions and directions for further researches on this topic.

2. Material and Methods

As mentioned in the Introduction section, to achieve the proposed objectives, the following techniques are used in the paper: Singular Spectrum Analysis (SSA), Bootstrap and the automatic procedures *ets* and *auto.arima*. This approach was inspired in the works of Bergmeir et al. (2014) and Dantas & Oliveira (2014), where the last one uses the STL decomposition (Seasonal-Trend Decomposition Using Loess) to model a wind power time series and the Moving Blocks Bootstrap technique applied to the noise from the fitted model. Twenty nine new series were generated and 24-step-ahead forecasts were obtained for each one of these simulated series plus the original one. The final 24-step-ahead forecast were obtained by averaging these forecasts generated by the 30 series. Such procedure produced more accurate forecasts than those obtained by modelling only the original series. Some modifications were made in the methodology described in the implementation of the present approach. The algorithm to implement the proposed procedure, called SSA.Boot, are:

1. Decompose the original series into signal and noise using SSA technique;
2. Apply the Bootstrap to the noise series, generating new 2 noise series;
3. Add to the 2 new noise series the signal, resulting in 2 synthetic series;
4. Fit, to each one of the 2 series, the automatic procedures *ets* or *auto.arima* and forecast D steps ahead; and
5. Use the simple average of the 2 forecasts and compare with the ones obtained by the model fitted to the original series.

ets

Exponential smoothing methods were originally classified by Pegels (1969) taxonomy, later extended by Gardner (1985), modified by Hyndman et al. (2002) and extended again by Taylor (2003), giving a total of fifteen methods. For an easy understand regarding all the possible methods, the triplet :' á6á5; refers to the three components: error, trend and seasonality. So the model '65:#á#á0; has additive errors, additive trend and no seasonality, similarly, '65:/á/×á/; refers to a model with multiplicative errors, a damped multiplicative trend and multiplicative seasonality. The steps involved in the automatic forecasting algorithm are summarized below and a complete description can be see in Hyndman & Khandakar (2008).

1. For each series, apply all models that are appropriate, optimizing the parameters of the model in each case;
2. Select the best of the models according to the AIC; and
3. Produce point forecasts using the best model for h steps ahead.

auto.arima

ARIMA models or Box & Jenkins methodology are so named for George Box and Gwilym Jenkins (Box et al., 2013), while exponential smoothing models are based on a description of trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data. The *auto.arima* function in R combines unit root tests, minimization of the AICc and MLE to obtain an ARIMA model :Lá@áM,T:2á&á3;>à ?where Lis the order of the autoregressive part, @the degree of first differencing involved and Mis the order of the moving average part, the respective capitals has the same meaning, but for seasonal part, if exist, of the series. The algorithm follows these steps:

1. The number of differences @is determined using repeated KPSS tests;
2. The values of Land Mare then chosen by minimizing the AICc after differencing the data @ times. Rather than considering every possible combination of Land M,the algorithm uses a stepwise search to traverse the model space.
    a. The best model (with smallest AICc) is selected from the following four: #4+/#tá@át;, #4+/#rá@ár;, #4+/#sá@ár;and #4+/#rá@ás;. If @Lr then the constant ?is included; if @Rs then the constant ?is set to zero;
    b. Variations on the current model are considered: vary Land/or Mfrom the current model by Gsand include/exclude ?from the current model; and
    c. Repeat Step 2(b) until no lower AICc can be found.

For a full explanation see Hyndman & Athanasopoulos (2013).

3. Results and discussions

To validate the SSA.Boot procedure, we use the series *elec*, *UKDriverDeaths*, *gas*, *uselec*, *ukcars* and *usgdp*, from the R packages *fma*, *datasets* and *expsmooth*, see the respective grapichs in Figure 1.

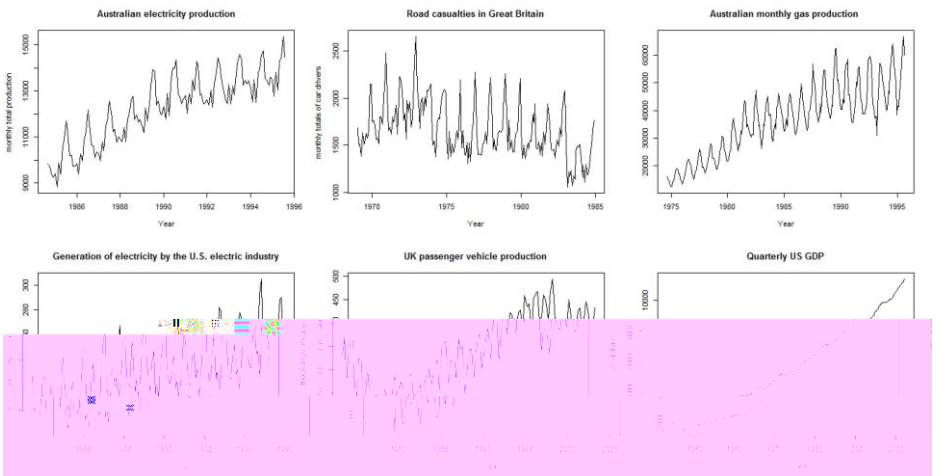

Figure 1. Time series from R packages ***fma***, ***datasets*** and ***expsmooth***.

From now on, will be carried out the steps from the approach suggested applied to each of the six series listed above. The series were split into the training set, in which the suggested approach is applied, and the validation set, where the comparisons between the forecast and actuals are made. The first step involves the application of SSA to extract the signal and the noise components, to do that we set the window length (.) as half of the series length and performed tests to determine the quantity of components (G that provides white noise. For four of the six series we were able to extract white noise with SSA technique without making any change in the series, but for the series *elec* and *gas* to obtain white noise according to the BDS test (Brock et al., 1987), size range cuts had to be performed. This may be related to the fact that these series in question have originnaly a very long history (464 observations), with very different patterns within it.

After obtain white noise in the previous step, the Bootstrap technique is applied to create the synthetic series that will be forecast using *ets* and *auto.arima*. Since there is simulation techniques involved in the calculation, the finals accuracy measures RMSE, MAE and MAPE are given as the average of ten rounds for each series, in Table 1 presented the results found using the suggested approach, where in (1) it was used the *ets* function to forecast each of the synthetic series and in the column (2) the *auto.arima* function.

Table 1. Accuracy measures of ten rounds out-of-sample

Time series		elec		UKDriverDeaths		gas		uselec		ukcars		usgdp	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Round 1	RMSE	92,74	78,97	150,67	240,43	46,91	1934,49	3,23	3,44	14,05	14,74	41,63	108,10
	MAE	172,89	164,68	150,67	240,43	1731,69	2382,67	3,47	4,27	15,40	15,18	42,52	108,10
	MAPE	1,23	1,17	11,13	16,84	3,56	5,24	1,36	1,66	2,53	2,44	0,25	0,64
Round 2	RMSE	129,67	55,38	157,85	205,34	173,53	1261,62	4,52	3,02	14,35	9,11	43,74	114,90
	MAE	193,94	156,95	157,85	205,34	1713,07	2185,47	4,56	3,89	15,57	11,79	44,41	114,90
	MAPE	1,38	1,12	11,63	14,16	3,56	4,78	1,79	1,51	2,56	1,89	0,26	0,69
Round 3	RMSE	113,27	64,42	147,47	238,83	209,92	1189,85	3,56	3,13	15,09	13,39	41,86	106,91
	MAE	184,66	160,22	147,47	238,83	1736,73	2051,30	3,79	3,75	16,07	14,43	42,82	106,91
	MAPE	1,32	1,14	10,82	16,69	3,62	4,52	1,48	1,46	2,63	2,32	0,25	0,64
Round 4	RMSE	112,79	84,19	142,46	165,55	356,44	1431,06	3,44	2,69	15,41	10,54	43,17	110,96
	MAE	186,02	168,24	142,46	165,55	1712,45	2113,37	3,64	3,80	16,29	12,40	43,88	110,96
	MAPE	1,33	1,19	10,50	11,09	3,58	4,68	1,43	1,47	2,67	1,99	0,26	0,66
Round 5	RMSE	120,98	61,15	142,31	203,02	176,66	1339,24	5,22	3,16	14,04	14,40	41,46	110,19
	MAE	186,63	164,62	142,31	203,02	1720,82	2106,50	5,22	4,04	15,28	14,89	42,44	110,19
	MAPE	1,33	1,17	10,47	14,01	3,58	4,62	2,05	1,58	2,51	2,38	0,25	0,66
Round 6	RMSE	130,30	71,22	147,31	218,76	84,80	1389,08	2,79	3,22	15,56	12,16	39,65	106,78
	MAE	192,87	162,70	147,31	218,76	1773,84	2232,82	3,32	4,10	16,46	13,88	40,96	106,78
	MAPE	1,38	1,15	10,85	15,08	3,63	4,87	1,30	1,59	2,70	2,23	0,24	0,64
Round 7	RMSE	111,48	83,64	149,65	187,71	55,59	1332,33	3,46	2,67	14,78	14,69	43,39	105,40
	MAE	184,03	170,61	149,65	187,71	1762,01	2147,43	3,78	3,78	15,83	15,49	44,25	105,40
	MAPE	1,31	1,21	11,02	12,82	3,64	4,73	1,48	1,47	2,60	2,49	0,26	0,63
Round 8	RMSE	117,59	59,60	140,41	197,17	206,44	1353,47	3,05	3,78	13,25	14,28	43,86	105,97
	MAE	186,76	156,29	140,41	197,17	1690,71	2229,36	3,48	4,04	14,69	14,88	44,58	105,97
	MAPE	1,33	1,11	10,33	13,53	3,52	4,89	1,37	1,58	2,42	2,39	0,26	0,63
Round 9	RMSE	123,53	69,34	144,33	251,55	256,49	1337,52	4,37	3,29	15,09	14,30	45,96	109,99
	MAE	189,52	162,34	144,33	251,55	1710,90	2198,41	4,38	3,75	16,12	15,01	46,48	109,99
	MAPE	1,35	1,15	10,63	17,54	3,57	4,84	1,72	1,47	2,64	2,42	0,28	0,66
Round 10	RMSE	100,61	69,52	161,92	200,71	255,73	1579,96	3,49	2,72	13,60	2,14	44,76	104,58
	MAE	178,41	163,39	161,92	200,71	1702,92	2222,20	3,67	4,12	15,02	12,86	45,38	104,58
	MAPE	1,27	1,16	11,94	13,76	3,55	4,90	1,44	1,59	2,47	2,10	0,27	0,62
Average	**RMSE**	**115,29**	**69,74**	**148,44**	**210,91**	**182,25**	**1414,86**	**3,71**	**3,11**	**14,52**	**11,97**	**42,95**	**108,38**
	MAE	**185,57**	**163,00**	**148,44**	**210,91**	**1725,51**	**2186,95**	**3,93**	**3,95**	**15,67**	**14,08**	**43,77**	**108,38**
	MAPE	**1,32**	**1,16**	**10,93**	**14,55**	**3,58**	**4,81**	**1,54**	**1,54**	**2,57**	**2,26**	**0,26**	**0,65**

Table 2 shows the accuracy measures obtained with *SSA.Boot+ets* or *SSA.Boot+auto.arima* and with the automatic methods (*ets* and *auto.arima*) applied only to the original series. The proposed approach showed that in all cases there is a gain in generating 2synthetic time series, forecast each of one and made the average as the final forecast. For the series *UKDriverDeaths*, *gas*, *uselec* and *ukcars* the suggested approach produce smallest errors in all measures, while for the series *elec* using *ets* to forecast the syntethic series there was a gain when compared to the method used directly in the original series. Finally, looking to the accuracy measures obtain for *usgdp*, note that the errors values are already very small only applying to the original series, so the proposed method obtained similar results. That is, in any case the proposed method decreased performance, shown to be efficient.

Table 2. Comparison of methodologies with the accuracy measures RMSE, MAE and MAPE.

Time series	n-h	s	h	Model	R function	RMSE	MAE	MAPE
elec	120	12	12	(M,M,M)	ets	144,23	202,35	1,45
				ARIMA(0,1,1)(0,1,2)[12]	auto.arima	73,52	163,84	1,16
				N=120, L=60, k=24, P=29+1	**SSA.BOOT+auto.arima**	**69,74**	**162,95**	**1,16**
				N=120, L=60, k=24, P=29+1	**SSA.BOOT+ets**	**115,29**	**185,57**	**1,32**
UKDriverDeaths	180	12	12	(M,N,A)	ets	196,00	196,00	14,51
				ARIMA(3,1,2)(2,0,2)[12]	auto.arima	295,83	295,83	20,89
				N=180, L=90, k=16, P=29+1	**SSA.BOOT+auto.arima**	**210,91**	**210,91**	**14,55**
				N=180, L=90, k=16, P=29+1	**SSA.BOOT+ets**	**148,44**	**148,44**	**10,93**
gas	240	12	12	ETS(M,A,M)	ets	2027,48	2666,66	5,44
				ARIMA(1,1,2)(0,1,1)[12]	auto.arima	401,79	2252,06	4,77
				N=240, L=120, k=30, P=29+1	**SSA.BOOT+auto.arima**	**1414,86**	**2186,95**	**4,81**
				N=240, L=120, k=30, P=29+1	**SSA.BOOT+ets**	**182,25**	**1725,51**	**3,58**
uselec	130	12	12	(M,N,M)	ets	4,25	4,30	1,69
				ARIMA(1,0,0)(0,1,2)[12]	auto.arima	3,28	4,17	1,62
				N=130, L=75, k=13, P=29+1	**SSA.BOOT+auto.arima**	**3,11**	**3,95**	**1,54**
				N=130, L=75, k=13, P=29+1	**SSA.BOOT+ets**	**3,71**	**3,93**	**1,54**
ukcars	105	4	8	(A,N,A)	ets	15,22	16,24	2,66
				ARIMA(3,1,0)(2,0,1)[4]	auto.arima	20,90	20,90	3,37
				N=105, L=52, k=14, P=29+1	**SSA.BOOT+auto.arima**	**11,97**	**14,08**	**2,26**
				N=105, L=52, k=14, P=29+1	**SSA.BOOT+ets**	**14,52**	**15,67**	**2,57**
usgdp	229	4	8	(A,Ad,N)	ets	42,36	43,37	0,26
				ARIMA(2,2,2)	auto.arima	113,77	113,77	0,68
				N=229, L=114, k=50, P=29+1	**SSA.BOOT+auto.arima**	**108,38**	**108,38**	**0,65**
				N=229, L=114, k=50, P=29+1	**SSA.BOOT+ets**	**42,95**	**43,77**	**0,26**

5. Conclusions

The joint use of the methods SSA and Bootstrap, called SSA.Boot procedure, with the both automatic algorithms *ets* and *auto.arima* proved to be a promising way for obtaining forecasts, since provided better accuracy measures when compared to the automatic methods applied into the original series.

As an extension of this work, the authors intend to use the proposed approach in other time series, such as the M3 competition.

Acknowledgements

The authors would like to thank CNPq (research project 443595/2014-3), FAPERJ (research project E-26/111.621/2014) and CAPES, which are governmental agencies that have also given support to this research.

References

Bergmeir, C., Hyndman R. J., & Benitez J. M. (2014) Bagging Exponential Smoothing Methods using STL Decomposition and Box-Cox Transformation. Working paper 11/14.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2013). Time series analysis: forecasting and control. John Wiley & Sons.

Brock, W. A., Dechert, W. D. & Scheinkman, J. A. (1987). A test for independence based on the correlation dimension. Departament of Economics, University of Wisconsin at Madison, University of Houston and University of Chicago.

Broomhead, D. S., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. Physica D, 20, 217-236.

Dantas, T. M., & Oliveira, F. C. (2014). Prediction of Wind Speed: An approach using Bagging Holt Winters with decomposition STL. XLVI Simpósio Brasileiro de Pesquisa Operacional (SBPO), (in Portuguese).

Efron, B. (1979). Bootstrap Methods: Another Look at the Jacknife. The Annals of Statistics, 7, 1-26.

Gardner Jr E. S. (1985). Exponential Smoothing: The State of the Art. Journal of Forecasting, 4, 1-28.

Golyandina, N., & Korobeynikov, A. (2013). Basic Singular Spectrum Analysis and Forecasting with R. Computational Statistics & Data Analysis, 71, 934-954.

Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. Journal of Data Science, 5, 239-257.

Hassani, H., & Mahmoudvand R. (2013). Multivariate Singular Spectrum Analysis: A general view and new vector forecasting approach. International Journal of Energy and Statistics, 1 (1), 55-83.

Hyndman, R. J., Koehler, A. B., Snyder R. D., & Grose, S. (2002). A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods. International Journal of Forecasting, 18(3), 439-454.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27 (3).

Hyndman, R. J., & Athanasopoulos, G. (2013). Forecasting: principles and practice. OTexts.

Kuensch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. Annals of Statistics, 17 (3), 1217-1241.

Pegels, C. C. (1969). Exponential Forecasting: Some New Variations. Management Science, 15 (5), 311-315.

R Development Core Team (2008). 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/.

Taylor, J. W. (2003). Exponential Smoothing with a Damped Multiplicative Trend. International Journal of Forecasting, 19, 715-725.