Applying SSA.Boot Procedure To Forecast Time Series

Paula Medina Maçaira
Fernando Luiz Cyrino Oliveira
Reinaldo Castro Souza

Pontifical Catholic University of Rio de Janeiro, Brazil

23 June, 2015

ISF 2015, Riverside, June 21–24
1 Introduction

2 Methodologies’ overview

3 SSA.Boot procedure

4 Applications

5 Final remarks
Motivation

SSA motivation:
- Is a technique that decompose a data set into signal and noise;
- Has recent applications in time series analysis;

Bootstrap motivation:
- Resampling technique that has application in many fields;
- Very popular methodology regarding the simplicity;

ets and auto.arima motivation:
- Automatic model selections;
- Widely used forecasting methods;
The main idea is to join the tree approaches mentioned before to obtain:

- forecasts; and
- confidence interval for the forecasts (in progress).

What to use?

- Use SSA to separate signal and noise of the data set;
- Use Bootstrap to resample the noise and simulate synthetics series;
- Use ets and auto.arima to fit the best model for each synthetic series and forecast; and
- Use R software R Core Team (2015) and the packages Rssa, boot, tseries, forecast to build the SSA.Boot procedure.
To construct the SSA.Boot procedure, the authors were inspired by the works of:

- Maçaira et al. (2015);
- Cordeiro (2009);
- Dantas and Oliveira (2014); and
- Bergmeir et al. (2014).
Objective of the study

- Combine SSA to Bootstrap to simulate synthetic series (SSA.Boot);
- Use ets or auto.arima to forecast;
- Evaluate the forecast performance using some accuracy measures and compare the results with the forecasts obtained by applying ets or auto.arima directly to the original series.
**Singular Spectrum Analysis (SSA)**

- Time series is a combination of a signal (trend, periodic, etc.) and some random error;
- The main objective of SSA is find the perfect separation between this components;
- The input parameters needed are the window length and the significant numbers of eigentriples to be extracted;
- **Difficult:** find the collection of parameter that perfectly separate signal and noise;
- In this work, the suggestion is test a few combinations and check via ACF and PACF which one produces white noise;
- A complete description of the theory and practice of SSA can be found in Hassani (2007) and Hassani and Mahmoudvand (2013).
Bootstrap

- First developed in Efron (1979), is a classic and widely used method of resampling data;
- Consists of a random sample with replacement of the elements of a random sample;
- In the context of time series, there is two basically ways to apply Bootstrap:
  - when the data has some dependence is necessary to resampling blocks of observations, so the technique name Moving Block Bootstrap Kuensch (1989); and
  - resample the residuals (white noise), which will be used in this paper, due to the fact that from the white noise extracted by SSA it is possible to ensure the hypotheses of independence of them, the required condition for application of the method.
Automatic procedures: ets and auto.arima

Both ets and auto.arima belong to the forecast package for the R system for statistical computing R Core Team (2015), where the main reference are the works of Hyndman and Khandakar (2008) and Hyndman (2015);

The ets function selects the best exponential smoothing model among all appropriate, according to AIC, and forecast the time series with the chosen parameters collection;

The auto.arima function is similar to ets, selecting the best ARIMA model.
SSA.Boot step-by-step

1. Decompose the original series into signal and white noise using SSA technique;

2. Apply the Bootstrap procedure to the noise series, generating new $P$ noise series;

3. Add to the $P$ new noise series the signal, resulting in $P$ synthetic series;

4. Fit, to each one of the $P$ synthetic series, the automatics procedures `ets` and `auto.arima`;

5. Generate $h$-steps-ahead forecasts for each synthetic series with the respective fitted models; and

6. Compute the simple average of the $P$ forecasts and assume that is the final forecast.
In order to validate the approach, it was used eight real time series available in R software.

- **co2**: monthly atmospheric concentrations of CO2, found in R package datasets;
- **MotorVehicle**: monthly series containing total domestic and foreign car sales in the USA, found in R package Rssa;
- **elec**: Australian monthly electricity production, found in R package fma;
- **UKDriverDeaths**: monthly totals of car drivers in Great Britain killed or seriously injured, found in R package datasets;
- **gas**: Australian monthly gas production, found in R package forecast;
- **uselec**: monthly total generation of electricity by the U.S. electric industry, found in R package fma;
- **ukcars**: quarterly UK passenger car production, found in R package expsmooth;
- **usgdp**: quarterly US GDP, found in R package expsmooth.
The data

c02 dataset

MotorVehicle dataset

elec dataset

UKDriverDeaths dataset

gas dataset

uselec dataset

ukcars dataset

usgdp dataset
for each time series the extraction via SSA is satisfactory when the ACF and PACF shows that the residuals are white noise;

- Bootstrap is applied in the residuals to obtain synthetic series from each time series;
- obtain the forecast for each one of the synthetic series and made the average as the final prediction;
- evaluate the performance of SSA.Boot procedure using the accuracy measures:

- **RMSE:** \[ \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_{obs,i} - X_{pred,i})^2} \]
- **MAE:** \[ \frac{1}{n} \sum_{i=1}^{n} |X_{obs,i} - X_{pred,i}| \]
- **MAPE:** \[ \frac{100}{n} \sum_{i=1}^{n} \left| \frac{X_{obs,i} - X_{pred,i}}{X_{obs,i}} \right| \]
Forecast results
Accuracy results

<table>
<thead>
<tr>
<th>Time series</th>
<th>n</th>
<th>s</th>
<th>h</th>
<th>Model</th>
<th>R function</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>co2</td>
<td>240</td>
<td>12</td>
<td>60</td>
<td>(M,Ad,M)</td>
<td>ets</td>
<td>3.43</td>
<td>3.43</td>
<td>4.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(1,1,2)(2,1,2)[12]</td>
<td>auto.arima</td>
<td>0.99</td>
<td>1.04</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=240,L=120,k=19,P=100</td>
<td>SSA.BOOT+ets</td>
<td>0.89</td>
<td>0.96</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>1.14</td>
<td>1.20</td>
<td>1.66</td>
</tr>
<tr>
<td>MotorVehicle</td>
<td>240</td>
<td>12</td>
<td>60</td>
<td>(M,Ad,A)</td>
<td>ets</td>
<td>114.56</td>
<td>121.33</td>
<td>131.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1,0,1)(2,1,2)[12]</td>
<td>auto.arima</td>
<td>123.39</td>
<td>128.16</td>
<td>138.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=240,L=120,k=26,P=100</td>
<td>SSA.BOOT+ets</td>
<td>117.51</td>
<td>123.80</td>
<td>133.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>102.20</td>
<td>110.55</td>
<td>120.17</td>
</tr>
<tr>
<td>elec</td>
<td>120</td>
<td>12</td>
<td>12</td>
<td>(M,M,M)</td>
<td>ets</td>
<td>144.23</td>
<td>202.35</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(0,1,1)(0,1,2)[12]</td>
<td>auto.arima</td>
<td>73.52</td>
<td>163.84</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=120,L=60,k=24,P=30</td>
<td>SSA.BOOT+ets</td>
<td>115.29</td>
<td>185.57</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>69.74</td>
<td>162.95</td>
<td>1.16</td>
</tr>
<tr>
<td>UKDriverDeaths</td>
<td>180</td>
<td>12</td>
<td>12</td>
<td>(M,N,A)</td>
<td>ets</td>
<td>196.00</td>
<td>196.00</td>
<td>14.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(3,1,2)(2,0,2)[12]</td>
<td>auto.arima</td>
<td>295.83</td>
<td>295.83</td>
<td>20.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=180,L=90,k=16,P=30</td>
<td>SSA.BOOT+ets</td>
<td>148.44</td>
<td>148.44</td>
<td>10.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>210.91</td>
<td>210.91</td>
<td>14.55</td>
</tr>
</tbody>
</table>

Where $N$ is the series length, $L$ the window length, $k$ the numbers of eigentriples selected and $P$ the quantity of synthetic series generate.
<table>
<thead>
<tr>
<th>Time series</th>
<th>n</th>
<th>s</th>
<th>h</th>
<th>Model</th>
<th>R function</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>gas</td>
<td>240</td>
<td>12</td>
<td>12</td>
<td>ETS(M,A,M)</td>
<td>ets</td>
<td>2027.48</td>
<td>2666.66</td>
<td>5.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(1,1,2)(0,1,1)[12]</td>
<td>auto.arima</td>
<td>401.79</td>
<td>2252.06</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=240,L=120,k=30,P=30</td>
<td>SSA.BOOT+ets</td>
<td>182.25</td>
<td>1725.51</td>
<td>3.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>1414.86</td>
<td>2186.95</td>
<td>4.81</td>
</tr>
<tr>
<td>uselec</td>
<td>130</td>
<td>12</td>
<td>12</td>
<td>(M,N,M)</td>
<td>ets</td>
<td>4.25</td>
<td>4.30</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(1,0,0)(0,1,2)[12]</td>
<td>auto.arima</td>
<td>3.28</td>
<td>4.17</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=130,L=75,k=13,P=30</td>
<td>SSA.BOOT+ets</td>
<td>3.71</td>
<td>3.93</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>3.11</td>
<td>3.95</td>
<td>1.54</td>
</tr>
<tr>
<td>ukcars</td>
<td>105</td>
<td>4</td>
<td>8</td>
<td>(A,N,A)</td>
<td>ets</td>
<td>15.22</td>
<td>16.24</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(3,1,0)(2,0,1)[4]</td>
<td>auto.arima</td>
<td>20.90</td>
<td>20.90</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=105,L=52,k=14,P=30</td>
<td>SSA.BOOT+ets</td>
<td>14.52</td>
<td>15.67</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>11.97</td>
<td>14.08</td>
<td>2.26</td>
</tr>
<tr>
<td>usgdp</td>
<td>229</td>
<td>4</td>
<td>8</td>
<td>(A,Ad,N)</td>
<td>ets</td>
<td>42.36</td>
<td>43.37</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIMA(2,2,2)</td>
<td>auto.arima</td>
<td>113.77</td>
<td>113.77</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N=229,L=114,k=50,P=30</td>
<td>SSA.BOOT+ets</td>
<td>42.95</td>
<td>43.77</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSA.BOOT+auto.arima</td>
<td>108.38</td>
<td>108.38</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Where $N$ is the series length, $L$ the window length, $k$ the numbers of eigentriples selected and $P$ the quantity of synthetic series generate.
This study shows that:

- for each one of the tested series the best error measures were produced with the proposed approach (SSA.Boot);
- in 75% of cases there was a significantly improvement in the errors measures when compared to the obtained by applying the ets or auto.arima straight to the original series; and
- the proposed methodology enhances the errors measures of others techniques applied in series without simulation when the forecast period is up to 10% the length of the training period.

There is a work in progress that is searching a method to extract automatically the window size and the numbers of eigentriple inside SSA technique, in order to improve the SSA.Boot procedure.
Thank you!

Complete references on the next slide.


