B3 - Time Series Analysis - Fall 2017
Home Work Assignment 2

Lars Forsberg
Uppsala University, Department of Statistics

November 8, 2017

Abstract

The task is to identify a good model, estimate it, evaluate, forecast it, and evaluate the forecasts for five simulated and one real time series.

1 General Instructions

For general instructions about the HWA, such as group formation, deadlines etc, see the section Home Work Assignments in the document 'B3 - Time Series - Fall 2017 - Schedule - and General Information.pdf.'

2 Introduction

The assignment consists of **two tasks**, in both tasks, you will apply the **Box-Jenkins approach** to time series modelling. In Task 1 you do that on five simulated series. In Task 2 you do that on a real data set.

3 Task 1

The data set for the first task is in the file B3_HWA2.csv downloadable from studentporatlen (folder HWA2). The file contains five series labeled Y1-Y2. This is **simulated data**. These series are generated using one of the stochastic processes in HWA1, using one of the parameter combinations used in that assignment. Your task is to use the Box-Jenkins approach to time series modelling, that is, go through the following steps

- Model specification/ Identification
- Estimation
- Model diagnostics
- Forecasting

Below some theory on how to do that.

3.1 Identification

Identification of the process should be done using mainly two tools:

- Time series plot
- The correlogram

Using these two tools you make an initial "guess" of what stochastic process/model that might suit the data.[1]

3.2 Estimation

Estimation is straightforward since you use the build in standard/default procedures in R

[1] In this particular task, one could say that your task is to find the process that actually generated the data. This, however, is true only in this constructed task, since in reality, no data is generated by a stochastic process, but a stochastic process can mimic the data, that is, provide a good approximation of the "true data generating process".

3.3 Evaluation

Evaluate the estimated model using the following tools:

- Time series plot of the residuals, do they look uncorrelated, that is, have the model succeded to capture the systematic variation in the data?
- Correlogram of residuals, comment on the autocorrelation structure, does the data seem uncorrelated , would they pass a formal test of being uncorrelated, if so, which test? (Hint: It is a test of whether several autocorrelations are simultaneously zero.) Perform that test. If the residuals are *not* uncorrelated, how could you change the model to capture that (remaining) systematic variation. What does the correlogram suggest?

3.4 Forecasts

The last step is to **forecast** the future. To be able to evaluate the forecasts, you must "save" a few observations, (not to be used in the estimation) for the forecast evaluation.

- The data Y1-Y5 consistes of **404** observations
- Use the first **400** to find and estimate the model
- Make forcasts for time points **401-404**
- Use the **actual observations** for time points 401-404 to evaluate the forecasts

Evaluate the forecasts from the different models using the following evaluation measures, (which you need to calculate **by hand**)

MSE

$$MSE = \frac{\Sigma \left(Y_t - \widehat{Y}_t\right)^2}{n},$$

RMSE

$$RMSE = \sqrt{\frac{\Sigma \left(Y_t - \widehat{Y}_t\right)^2}{n}},$$

and **MAPE**

$$MAPE = \frac{1}{n}\Sigma \left| \frac{Y_t - \widehat{Y}_t}{Y_t} \right| 100,$$

where Y_t is the true value, $\widehat{Y}_t$ is your prediction and n is the number of predictions, in this case 4.

Report the forecast error measures, from the different models/datasets on the following format:

	Choosen	Evaluation measure		
Data	Model	MSE	RMSE	MAPE
Y_1	Model A			
Y_2	Model B			
Y_3	Model C			
Y_4	Model D			
Y_5	Model E			

4 Task 2

In this task you will analyze real data. There are lots of economic data on the web, for instance

- http://ec.europa.eu/eurostat/data/database
- http://research.stlouisfed.org/fred2/

The task here is straightforward

1. Choose one dataset
2. Apply the Box-Jenkins approach to analyze and forecast that dataset, as above when you analyzed the simulated series, i.e.
   (a) identify the process/model,
   (b) Estimate the model.
   (c) Do the diagnostics
   (d) Forecast

Thus, like before, you need to "save" some observations for forecast evaluation, choose the number of observations to forecast that you find appropriate.

Note: You can, and are encouraged to, do the full modelling-cycle for more than one data set. Different kinds of data can have very different statistical properties, and it is very instructive to analyse more than one series.

In any event: In the report that you hand in - you should hand in the analysis of *one* data-set only.

Optional
For the real data set, ask yourself, could it actually be the case that a some model will provide *better* forecasts (than the one choosen in the identification/specifcation stage), even though it will **not** be the model of first choice,

using the criteria used in the identification/specification stage. If so compare two different models according to the different forecast *evaluation measures*. Present the results in a table like the one *below*. Note now, for this result presentation, that here there is only *one* dataset per table, and the comparision is across *different models* (for that particular dataset, for example GDP_t).

Data is GDP_t	Evaluation measure		
Model	MSE	RMSE	MAPE
AR(1)			
AR(2)			
Model C			
etc			

Analyze the results along the lines: "We see that according to MSE model X is best while according to the MAPE the model Y is better, however in a real life application we need to consider..."