Forecasting: Exam assignment
IÉSEG School of Management

Practical issues

1. Please hand in the paper, the `R` scripts and - for the last question - the data set.

2. Deadline

   - The deadline for handing in your assignment is **2019-04-26** (midnight). Papers that are submitted after the deadline will not be accepted. In case of discussion, the time stamp in the email will be used as a criterion (unless grave circumstances can be argued).
   - Send the paper to f.vandenbossche@ieseg.fr.

Data sets

The data sets for the first and second exercise are in the Excel spreadsheet `DataSets.xlsx`, to be found on IÉSEG Online. The "content" tab in this file contains a description of the characteristics of the data sets.

   Once you've assigned a working directory, you can load the data sets. These are stored in the other tabs of the Excel spreadsheet. For example, to load the `Fatalities` data set, you can use the following code:

```
setwd("C:/Temp") # Specify you own working directory here.
data <- read_excel("DataSets.xlsx", sheet="Fatalities")
Fat <- ts(data[,2], frequency = 1, start = 1965)
```

   Other data sets in the Excel spreadsheet can be loaded in a similar way, by referring to the correct tab in the `read_excel` command.

Exercise 1

The data set `Turnover` contains the Belgian industry turnover index for manufacturing of beverages from January 2000 to December 2018.

Split the data in a training set up to December 2015 and a test set from January 2016 up to December 2018. Use the training set for estimation of the methods/models, and use the test set for assessing the forecast accuracy.

In each step of the exercise, discuss your results and explain your choices. Use additional tables and graphs if they clarify your answer.

1. Explore the data using relevant graphs.

2. Create forecasts using the seasonal naive method. Check the residual diagnostics and the forecast accuracy.

3. Use an STL decomposition to forecast the turnover index. Use the appropriate underlying methods to do so. Check the residual diagnostics and the forecast accuracy.

4. Generate forecasts using Holt-Winters' method. Check the residual diagnostics and the forecast accuracy.

5. Generate forecasts using ETS. First select the appropriate model(s) yourself and discuss their performance. Compare these models with the results of the automated ETS procedure. Check the residual diagnostics and the forecast accuracy for the various ETS models you've considered.

6. Generate forecasts using ARIMA. First select the appropriate model(s) yourself and discuss their performance. Compare these models with the results of the `auto.arima` procedure. Check the residual diagnostics and the forecast accuracy for the various ARIMA models you've considered.

7. Compare the different models in terms of residual diagnostics, model fit, and forecast accuracy. Analyse your results and select your final model.

8. Finally, generate out of sample forecasts up to December 2020, based on the complete time series. Discuss your results.

Exercise 2

The data set `Houses` shows the yearly number of family houses sold in Belgium (Houses) and the total price of transactions (Houseprices) from 1973 up to 2017.

First, create the *average house price* by dividing the total price of transactions by the yearly number of family houses sold. This new variable will be analysed.

Split the data in a training set up to 2010 and a test set from 2011 up to 2017. Use the training set for estimation of the methods/models, and use the test set for assessing the forecast accuracy.

In each step of the exercise, discuss your results and explain your choices. Use additional tables and graphs wherever they clarify your answer.

1. Explore the data using relevant graphs.

2. Create forecasts using the most appropriate naive method. Check the residual diagnostics and the forecast accuracy.

3. Generate forecasts using the relevant exponential smoothing methods. Check the residual diagnostics and the forecast accuracy.

4. Generate forecasts using ETS. First select the appropriate model(s) yourself and discuss their performance. Compare these models with the results of the automated ETS procedure. Check the residual diagnostics and the forecast accuracy for the various ETS models you've considered.

5. Generate forecasts using ARIMA. First select the appropriate model(s) yourself and discuss their performance. Compare these models with the results of the `auto.arima` procedure. Check the residual diagnostics and the forecast accuracy for the various ARIMA models you've considered.

6. Compare the different models in terms of residual diagnostics, model fit, and forecast accuracy. Analyse your results and select your final model.

7. Finally, generate out of sample forecasts up to 2020, based on the complete time series. Discuss your results.

Exercise 3

For this exercise, find a recent and relevant time series to forecast, using the techniques that have been discussed during the lectures. In your analysis, set up a carefully selected forecasting process, taking data considerations and implementation issues into account. Describe your approach, and motivate your choices.

The data set should be original (not from R packages), recent, suffciently long and must include a seasonal component. The data should be analyzed using at least two techniques, and results should be compared.

Good Luck, and Happy Forecasting!