Data Science Winter School
Time Series Analysis and Forecasting

Tutorials

I: We analyse the credit score data (`credit.csv` or `credit` data set from `fpp2` package in R. The objective is to forecast the credit score of bank clients relying on several explanatory variables.

1. Import the data and make appropriate visualisation (`pairs, ggpairs`)

2. Split the data randomly into training and test data set (`sample`).

3. Run the regression using training data set. Assess the fit.

4. Compute the confidence intervals for the parameters.

5. Compute the forecasts (point, prediction and confidence intervals) for the test data set (`predict, forecast`).

6. Compute forecast errors and analyse them.

7. Make appropriate data transformation and repeat the above steps for the new model.

8. Compute the marginal effects for the original and the transformed models (`margins`).

9. Compare the forecasting ability of the two models using loss functions and using appropriate tests.

10. The score variable is slightly skewed. Takings logs shall overcome this problem. Check the forecasting ability of the log-score model.

II: Next we consider the quarterly UK passenger car production (`ukcars.csv` or `ukcars`) data set. The objective is to model the production data using regression techniques.

1. Plot the data and choose an appropriate order of time polynomial for modelling. Keep the last 10 observations as test data set.

2. Run a simple regression on the time polynomial and check the residuals (using Durbini-Watson).

3. Compute the forecast for the test data set and the corresponding losses.

4. Create dummy variables for the quarters and repeat the last two steps.

5. Create a sin and/or cos explanatory variables to capture variation within calender years and repeat the above steps.

6. Compare the performance of the models (no tests, but in- and out-of-sample fit), visualize the results. Compute the average forecast and compare it with the individual forecasts.

III: The data set `elecequip` or `elecequip.csv` contain monthly data on the sales of electronic devices in EU (17 countries, 2005=100). The objective is to decompose the data using simple smoothing techniques.

1. Plot the data trying to stress the attention on the components.

2. Try several moving average techniques to extract the trend (`ma`). Which orders/form of moving averages do provide the best result?

3. Apply the loess regression for trend extraction.

4. For one of the above computed trends compute the seasonal and the residual components.

5. Check the ACF of the residual component. Is there anything left in the time series?