Time Series Analysis 02417
Spring 2024

Assignment 1

Instructions: The assignment is to be handed in via DTU Learn "FeedbackFruits" latest at February 23rd at 23:59. You are allowed to hand in in groups of up to max 4 persons. You must hand in a single pdf file presenting the results using text, math, tables and plots (do not include code in the main report - if you want to include your code as an appendix, you can add this in a separate file). Arrange the report in sections and subsections according to the questions in this document. Please indicate your student numbers on the report.

1 Plot data

In this assignment we will be working with data from Statistics Denmark, describing the number of motor driven vehicles in Denmark. The data is provided to you, but if you are interested you can find it via `www.statistikbanken.dk` (search for the table: "BIL54"). Together with this document are two files with data - one with "training data" including a timeseries of 59 observations of monthly data starting in January 2018. The second file with data includes "test data" with 12 extra months (beginning December 2022). To begin with we will ONLY work with the training data. The variable of interest is *number vehicles in total* (in Danish "Drivmidler i alt"). We will ignore the other variables in the dataset.

1.1. Make a time variable, x, such that January 2018 has $x_0 = 2018$, February 2018 has $x_1 = 2018 + 1/12$, March 2018 has $x_2 = 2018 + 2/12$ etc. and plot the training data versus x.

1.2. Describe the time series in your own words.

2 OLS

We will now make a global linear model of the form:

$$Y_t = \beta_0 + \beta_1 \cdot x_t + \epsilon_t \tag{1}$$

2.1. Estimate the parameters β_0 and β_1 using the training data (OLS model). Describe how you estimate the parameters.

2.2. Present the values of the parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ and their estimated standard errors $\hat{\sigma}_{\hat{\beta}_0}$ and $\hat{\sigma}_{\hat{\beta}_1}$.

2.3. Make a forecast for the next 12 months - i.e., compute predicted values with corresponding prediction intervals. Present these values in a table.

2.4. Plot the fitted model together with the training data and the forecasted values (also plot the prediction intervals of the forecasted values).

2.5. Comment on your forecast - is it good? What could be better?

2.6. Investigate the residuals of the model. Are the model assumptions fulfilled?

3 WLS - local linear model

We will now use WLS to make the linear model in Eq. (1) as a local trend model, i.e., the observation at the latest timepoint ($x_N = x_{59} = 2022.833$) has weight $\lambda^0 = 1$, the observation at the second latest timepoint ($x_{N-1} = x_{58} = 2022.750$) has weight λ^1, the third latest observation ($x_{N-2} = x_{57} = 2022.667$) has weight λ^2 etc.
We start by setting $\lambda = 0.9$.

3.1. Describe the variance-covariance matrix (the $N \times N$ matrix Σ) for the local model and compare it to the variance-covariance matrix of the corresponding global model.

3.2. Plot the "λ-weights" vs. time in order to visualise how the training data is weighted. Which time-point has the highest weight?

3.3. Also calculate the sum of all the λ-weights. What would be the corresponding sum of weights in an OLS model?

3.4. Estimate and present $\hat{\beta}_0$ and $\hat{\beta}_1$ corresponding to the WLS model with $\lambda = 0.9$.

3.5. Make a forecast for the next 12 months - i.e., compute predicted values corresponding to the WLS model with $\lambda = 0.9$.
Repeat (estimate parameters and make forecast for the next 12 months) for $\lambda = 0.8$, $\lambda = 0.7$ and $\lambda = 0.6$.
Plot the training data (Y_t), the predicted values ($\hat{Y}_t$ - one series for each λ), and the forecasted values (also $\hat{Y}_t$, but for the next 12 months and also one series for each λ).

3.6. Comment on the forecasts - do the slopes of each model correspond to what you would (roughly) expect for the different λ's?

3.7. Which model would you use for decision making - or how would you decide which one to use knowing only the results generated until here?

4 Iterative update and optimal λ

In the section above we used four different values of forgetting, but didn't quantify how they affect the prediction performance. In order to choose an optimal λ we will now *iterate* the model through the training data - i.e. for each time point, x_t, in the training data, we will compute a local trend model, using only the training data up until x_t, and then make predictions for 1, 6 and 12 months ahead. These predictions can be compared to the training data. You may use the updating formulas for F_N and h_N, see the book page 55.

4.1. Provide L and $f(0)$ for the model.

4.2. Provide F_1 and h_1.

4.3. Using $\lambda = 0.9$ update F_N and h_N recursively and provide F_{10} and h_{10}. We will not calculate predictions for these first 10 steps.

4.4. Now update the model recursively up to F_{59} and h_{59}, while also calculating predictions at each step. You should calculate predictions for 1 month ahead, 6 months ahead and 12 months ahead.

4.5. Plot the resulting 1-month, 6-month and 12-month prediction together with the training data.

4.6. Repeat the iterative predictions for $\lambda = 0.55, 0.56, 0.57, ..., 0.95$, and calculate the root-mean-square of the prediction errors for each forecast-horizon (1 month, 6 months and 12 months) and for each value of λ.
Plot the root-mean-square of the prediction errors versus λ for both 1 month, 6 months and 12 months predictions.

4.7. Which value of λ is optimal for predicting 1 month ahead?

4.8. Which value of λ is optimal for predicting 6 months ahead?

4.9. Which value of λ is optimal for predicting 12 months ahead?

4.10. It would be problematic to make λ as small as 0.5. Why is that? (hint: compare the sum of weights to the number of parameters).

4.11. Compare the 1-month-ahead prediction with $\lambda = 0.55$ to the *naive* persistence model (i.e. using only the last observation to predict 1 month ahead). Which one is better?

4.12. Now choose the best forecasts at time $t = 59$ of Y_t for time $t = 60$, $t = 65$ and $t = 71$, i.e. $\hat{Y}_{59+1|59}$, $\hat{Y}_{59+6|59}$, and $\hat{Y}_{59+12|59}$ and plot them together with the training data AND the test data (i.e. 2nd data file).

4.13. Comment on the models prediction performance for both 1 month, 6 months and 12 months ahead.