

Assignment 3: Predicting the amount of legislation

Is the Danish society becoming more civilised? Judging by the amount of legislation you would think so, with the size of the legislative body more than doubling between 2002 and 2021. What about the future? As enlightened citizens that know the rules of their country how long time can we expect to study new legislation per year? What dynamics does law-production experience? And what political party is the most productive? These questions will be studied in this assignment, where we will first use the ARIMA formulation and afterwards one of the transfer function formulations to model and forecast the amount of words in Danish legislation.

The data for this assignment can be found in "DataAssignment3.csv", and consists of the yearly net change in the amount of words in the Danish legislation, along with the name of the party from where the sitting prime minister came from during that year (in years with more than one prime minister, the one who had the office for the longest period is reported). The first year of the data is 1919 while the last is 2020. More detailed data can be provided for the interested, in which all changes to the amount of words can be found, with time stamps, thus yielding a continuous time series. The data on legislation has been provided by Jonas Herby who is a Special Advisor at CEPOS, while sitting prime ministers have been extracted from Wikipedia.

This assignment will not be graded using peergrade, and so you do not have to worry about obtaining the same models as I did. Instead, you should focus on presenting your findings in a clear and concise way. You should try and limit the size of your assignment while still conveying what you find to be interesting. There might be things not explicitly asked for in the assignment that would still be favourable for you to include. You decide how to choose between model candidates, e.g. using information criteria, cross validation or visual inspection of residuals. For this assignment you are required to hand in all code used to solve it as well, either as readable code in an appendix or in a separate file.

Question 2.1: Visualising the data Plot the total amount of words in the Danish legislation, and comment on the overall pattern. Is the data stationary? What about the variance?

What things can this visual inspection tell you about the needed structure of the ARIMA model? If you think transformations will be necessary, plot and comment on these transformations of the data, including differencing.

Question 2.2: Identifying a suitable ARIMA model Use the ACF and PACF function to identify a suitable ARIMA model. Guide the reader through the steps. When a suitable model has been found, do a residuals diagnostic to show whether it fits the data well. Write the mathematical formulation of your final model and report the estimated parameters.

Note: When differencing using "arima" from R, it will automatically not include an intercept even if you put "include.mean=TRUE". Consider fixing this using either the "xreg" argument or by manually differencing the data.

Question 2.3: Forecasting future legislation Forecast the total amount of words in the Danish legislation for the years 2021 to and including 2030. Do include prediction intervals and

comment on the results.

Question 2.4: Interpreting the ARIMA model Inspect the parameters of your final model and interpret them. Argue why they make sense or explain why you are surprised by the results.

Question 2.5: Including the political party Make a transfer function model, with the political party as input. You are allowed and indeed encouraged to not use the transfer function model in its full generality but rather one of the simpler versions found in the slides of week 7. You are free to search for libraries that can fit these models, but make sure that you understand exactly the model being estimated, for example by doing something similar to the first exercise of week 7.

You do not have to show the steps you take to make this model, but for the final model:

- If you used a library, state the name of the library and the specific code/model call that you used to estimate the model.
- Write out the mathematical formulation.
- Present the parameter estimates including e.g. their standard deviation or confidence intervals.
- Interpret the parameter estimates and argue why they make sense or why you are surprised.
- What political party produces legislation at the highest rate according to your model, and how fast is this? (You have to figure out how this is quantified for your particular model)
- Forecast the amount of words in the Danish legislation for the years 2021 to and including 2030 given that the prime minister is a Social Democrat from now and until 2030. Plot it including prediction intervals.
- For each of the 5 given political parties, forecast the amount of words in the danish legislation for the years 2021 to and including 2030 given that the prime minister is from that political party. Plot all in the same plot *without* including prediction intervals.

Question 2.6: Conclusions If you have found other interesting conclusions or have comments based on the work, then report them. Note that it is not necessary to report anything here for full points.