Assignment 1

Exercise 1

Download the current version of the FRED-MD database and load it into *R* (or another statistical software of your choice). Note that the second line in the CSV-file denotes the suggested transformation, you have to remove it.

- Create a function that takes a vector containing observations of a time series as input and returns a dataframe with the following transformed series in its columns as output:
  - the original time series in its raw form.
  - the log-transformed time series.
  - month-on-month growth rates in percent.
  - year-on-year growth rates in percent.
  - the first lag of the year-on-year growth rates of the time series.
- Use the created function to create a dataframe with the various transformation for US industrial production (mnemonic *INDPRO*), plot the logged time series and the yearly changes produced by the function. Briefly describe the properties of the time series.
- Using suitable functions from the *stats* and *urca* package, assess the properties of both logged industrial production and its yearly growth rate. Plot the autocorrelation function and perform Dickey-Fuller tests to test for a unit root (note the different specifications, i.e. including a drift or a trend), interpret the results.
- Estimate a suitable AR model (e.g. using the *ar.ols()* function) for the stationary time series (as determined in the previous point).[1] How is the lag order determined by default? Use the estimated model to produce forecasts for the next year and plot them. Interpret their behaviour (i.e. are they converging towards a certain value? What could that be?). Use the produced forecasts to also forecast the change in the original time series.
- Bonus: Create a function that computes the RMSE of a given AR model based on a lag order and a holdout period that you can specify. That is, the function should take as inputs the time series, the number of lags to include in the AR model and the number of time periods that will be used for computing the RMSE (the holdout period). Note that prior to estimation you should remove the holdout period from the end of the sample, estimate the model, produce forecasts for the holdout period and then compute the RMSE based on the predicted values and the realized values of the time series. Using this function, compare a number of AR models (i.e. with different lag orders) to assess the predictive performance of them. Which one would you choose and why? Might your answer differ with different holdout periods (e.g. 6 versus 12 months forecasting horizon)?

Exercise 2

Read Kilian & Park (2009), who discuss the effects of oil price shocks on the US stock market, focus on Sections 2 and 3.1-3.3. Load the provided data by Kilian & Park (2009), which contains a measure of change in oil production, a measure of real economic activity, the real price of oil, and changes in real US dividend growth from 1973M1 to 2016M12.

- Using the the packages *vars* in R (or an equivalent one in another language), estimate the VAR described in section 2.2 using the variables in the same order as specified by Kilian & Park (2009).[2]
- Using the estimated VAR, compute impulse response functions (take a look at the *irf()* function in *vars*, it uses the same identification scheme as Kilian & Park (2009) propose (recursive ordering based on a Cholesky decomposition of the vcov-matrix of the errors) by default. Replicate Figure 1 and the lower panel in figure 3 of Kilian & Park (2009).[3] Interpret the results.
- Calculate forecast error variance decompositions for the included variables (take a look at the *fevd()* function in *vars*). Replicate Table 2 of Kilian & Park (2009). Interpret the results.

[1] Ignore more general ARMA models for now.

[2] The order of variables in this setting is important for the identification of the VAR, i.e. when recovering the structural form of it. We did not cover identification in the lectures yet, however take a look at the argumentation of the authors on how they approach it.

[3] Note that the bottom panel of figure 3 shows reponses of real dividends, not the change in real dividends. Hence, you will have to calculate *cumulative* impulse responses. Note also that the oil supply shock is a negative shock (i.e. lowering oil production), while the other two are positive ones. With a linear VAR, you can retrieve a negative shock by inverting the signs of the responses to a positive shock.

- Note that the dataset provided misses US stock market returns (due to the licensing of the underlying time series). Look for alternative data on the US stock market,[4] create a variable similar to the one used by Kilian & Park (2009). Re-estimate the model and replicate Figure 1 again as well as the top panel of Figure 3 and Table 1.[5] Interpret the results.

[4]Take a look at the FRED-MD database from exercise 1 for that. The S&P-200 might be a suitable alternative. Note that in their paper, the authors use *real* stock returns.

[5]Note that there might be deviations from the results of Kilian & Park (2009) due to using a different underlying variable for stock market returns.