IE 360 Statistical Forecasting and Time Series
Homework 2, due May 7th, 2021

Instructions: Please solve the following exercises using R (http://www.r-project.org/) or Python (https://www.python.org/). You are expected to use GitHub Classroom and present your work as an html file (i.e. web page) on your progress journals. There are alternative ways to generate an html page for you work:

- A Jupyter Notebook including your codes and comments. This works for R and Python, to enable using R scripts in notebooks, please check:
  - https://docs.anaconda.com/anaconda/navigator/tutorials/r-lang/
  - https://medium.com/@kyleake/how-to-install-r-in-jupyter-with-irkernel-in-3-steps-917519326e41

  Things are little easier if you install Anaconda (https://www.anaconda.com/). Please export your work to an html file. Please provide your *. ipynb file in your repository and a link to this file in your html report will help us a lot.

- A Markdown html document. This can be created using RMarkdown for R and Python-Markdown for Python

Note that html pages are just to describe how you approach to the exercises in the homework. They should include your codes. You are also required to provide your R/Python codes separately in the repository so that anybody can run it with minimal change in the code. This can be presented as the script file itself or your notebook file (the one with *.ipynb file extension).

The last and the most important thing to mention is that academic integrity is expected! Do not share your code (except the one in your progress journals). You are always free to discuss about tasks but your work must be implemented by yourself. As a fundamental principle for any educational institution, academic integrity is highly valued and seriously regarded at Boğaziçi University.

Time Series Regression for Predicting Macroeconomic Indicators

This homework extends the descriptive analyses you have performed with the data from Data Delivery System: https://evds2.tcmb.gov.tr/ of Central Bank of the Republic of Turkey Electronic. If you click "All Series" link on the upper menu, you will see the data categories in an alphabetical order on the left. Once you click one of the categories, you will be able to see the available data under the corresponding category.

The aim of this homework is to forecast a selected indicator at a monthly level. You will use the following target variables based on your student id. Based on the modulo 10 of your student id (i.e. last digit of your student id), use the respective target variable below:

Following four variables are from "Price Indices" menu:

0- Consumer Price Index (2003=100)(TURKSTAT) -> "FOOD AND NON-ALCOHOLIC BEVERAGES"
1- Consumer Price Index (2003=100)(TURKSTAT) -> "PURCHASE OF VEHICLES"
2- Consumer Price Index (2003=100)(TURKSTAT) -> "ALCOHOLIC BEVERAGES AND TOBACCO"
3- Consumer Price Index (2003=100)(TURKSTAT) -> "CLOTHING AND FOOTWEAR"

Following four variables are from "Housing and Construction Statistics" menu:

4- House Sales Statistics - Turkey - Second hand sale (TURKSTAT)
5- House Sales Statistics - Turkey - First sale (TURKSTAT)
6- House Sales Statistics - Turkey - Mortgaged sales (TURKSTAT)
7- House Sales Statistics - Turkey - Total sales (TURKSTAT)

Following two variables are from "Production Statistics" menu:

8- Firms Statistics-Newly Established and Liquidated (TOBB) -> Newly established Total (Number)
9- Firms Statistics-Newly Established and Liquidated (TOBB) -> Liquidated Total (Number)

As independent variables, you can use any possible relevant predictor you came up with. However the use of data from the one or few of the following is strongly suggested:

- Survey question answer statistics from "Seasonally unadjusted Consumer Confidence Index and Indices of Consumer Tendency Survey Questions" under "SURVEYS" menu. You do not need to use all of them, you can select questions that is relevant to your target.
- Relevant statistics under "INTEREST RATES" menu.
- Relevant statistics under "EXCHANGE RATES" menu.

You are expected to make your series selections so that your analyses make sense. To be more clear, the selected data should not be irrelevant. You are expected to build a time series regression model for forecasting next month's value of the target variable. The target series are monthly observations and they are available up to 2021-03. Our aim is to provide a forecast of 2021-04 (so that we can check our prediction error later when 2021-04 statistics are announced. Note that we need to make use of lagged varıables ın order to forecast future. In other words, regressors are also not available for 2021-04. This may require you to use the information you have in the previous periods to forecast the future.

Please follow the time series regression steps (like the one we did in class and problem solving sessions) and describe your analysis steps using the following report format.

You will be evaluated based on the following criteria:

- Format (10 points)
- Soundness of your research question (20 points) – An introduction section
- Data manipulation and visualization (60 points) – An analysis section where you read and merge the data, create time series objects, visualize data, find correlations and comment based on your research question and findings.
- Concluding remarks (10 points) – A conclusion section