ITC6002 B1: Exploring and Analyzing Data

Final Project

Due on Dec. 12, 2023

A. Choose a data set of a quantity whose daily values over two years are publicly available. [For example, this could be Covid19 cases, stock market data, oil / gas prices, meteo data, etc.]. Go over this data set and make sure to clean it up of any inconsistencies, outliers, missing data, etc.

B. Using the 1st year data of your chosen (and cleaned up) data set, try to forecast the data of months 2-12 based on the first month's data. For this:

- **a. Use 3 different forecasting techniques to perform the forecasting of level (without trend or seasonality adjustments). If a technique is parametric, try to choose the optimal value for this parameter.**
- **b. Compare these 3 techniques in 2 different ways.**
- **c. Try to see if there is a trend in the 1st year's data set, estimate it and then do a trend-adjusted forecasting.**
- **d. Then try to see if there is seasonality in the 1st year's data set, estimate the seasonality factor and then adjust your forecasting for seasonality as well.**

C. Do the same (as in B.) for the 2nd year's (cleaned up) data set. Keep in mind that any parameters may have to be re-optimized for the 2nd year data set.

D. Compare the 2 data sets in terms of the forecasts you made and the available actual data and try to describe quantitatively how they differ.

E. Finally, choose a monthly data set (i.e. only 12 data points) from either the 1st or the 2nd year of your data (at regular intervals, e.g. the 1st day of each month). Then find at least another 10 similar data sets from other sources. For example, if your data set is the peak temperature of a municipality of Athens on the 1st of each month, find another 10 Athens municipalities that you expect to have similar weather, and collect the corresponding data. Run your best forecasting technique for the 12 point data set. Then adjust it taking advantage of the other 10 similar data sets.

Note: Provide a full report (about 3000 words) including methodology, code, plots and presentation.