INDR 450/550 HOMEWORK 1, Due Date: March 20, 2022

- Please work in groups of two or three or (individually) and submit one file for each group with all names.
- The data is on total monthly average price index of coffee in commodity markets as reported by the IMF commodity price data portal (data found in the blackboard page). The data starts from January 1990 and ends in January 2022 (included).
- Please perform all computations in python. You are expected to implement your own forecasts (don't use packages/functions from the web). Please submit (opload on blackboard) your commented (with explanations) python notebooks.
- In addition to the python notebook, **submit a short typed summary report** that includes the results (error tables, prediction intervals etc.) of all exercises. Also add a general assessment of the methods (which method is the best, which should be avoided etc.). **The report is part of the overall grade.**

Exercises

1. Forecasting Coffee Prices (monthly average price index of coffee in commodity markets as reported by the IMF commodity price data portal)
   (a) Plot the data and visually assess whether there is significant trend and seasonality.
   (b) To obtain a benchmark for errors, implement the naive one-month ahead forecast i) $\hat{y}_t = y_{t-1}$. Report the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) of the naive forecast for years 1991 through 2020. These error measures constitute a simple benchmark for all other approaches (i.e. hopefully you will obtain lower errors by more sophisticated methods). There is plenty of data in this case and we can start forecasting from 1990 but different methods require different initializations in the first year so it's best to compare the errors from 1991 to the end of 2020 and we use 2021 for some basic testing.

(c) Use a 5-period moving average to forecast the one month-ahead price. Report the MAE, MAPE and RMSE of the forecast for years 1991 through 2020. Report 95 percent prediction intervals (using the RMSE estimated in years 1991 to 2020) for the one-month ahead forecasts for year 2021. How do your prediction intervals perform for 2021?

(d) Use exponential smoothing to forecast the one-month ahead price. Perform an exhaustive search for the best smoothing constant α (that leads to the minimum MSE). Report the MAE, MAPE and RMSE of the forecast for years 1991 through 2020. Report 95 percent prediction intervals (using the RMSE estimated in years 1991 through 2020) for the one-month ahead forecasts for 2021. How do these compare with the benchmark and the MA-5 forecast?

(e) The data seems to have trend. Implement a naive forecast that includes trend: $\hat{y}_t = y_{t-1} + (y_{t-2} - y_{t-3})$ in order to forecast the one-month ahead price. Report the MAE, MAPE and RMSE of the forecast for years 1991 through 2020. Report 95 percent prediction intervals (using the RMSE estimated in years 1991 through 2020) for the one-month ahead forecasts for 2021. How do these compare with the previous forecasts?

(f) Implement an exponentially smoothed version of the previous forecast (you can take $\alpha = 0.7$ and $\beta = 0.2$): $\hat{y} = \alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1} + \beta z_{t-1} + (1-\beta)\hat{z}_{t-1}$ where $z_t = y_t - y_{t-1}$ and $\hat{z}_t = \hat{y}_t - \hat{y}_{t-1}$ for one-month ahead prices. Report the MAE, MAPE and RMSE of the forecast for years 1991 through 2020. Report 95 percent prediction intervals (using the RMSE estimated in years 1991 through 2020) for the one-month ahead forecasts for 2021. . How do these compare with the previous forecasts?

(g) Find the values of α and β that minimize the RMSE for **six-month ahead forecasts** for years 1991 through 2020. For the optimal values of the parameters, predict the mean coffee prices for the next six months from January 2021 to June 2022. Note that $\hat{y}_{t+k} = \alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1} + (k + 1)(\beta z_{t-1} + (1 - \beta)\hat{z}_{t-1})$. Report the MAE, MAPE and RMSE of the six-month ahead forecast for years 1991 through 2020.

(h) Your report for the exercise should include a table similar to the one below.

Method	Spec.	RMSE	MAPE
Benchmark	-		
MA-5	-		
ES	-		
Trend	-		
Smoothed Trend			
6-month ahead			

Note that the model specification for exponential smoothing is: $\alpha^* = ...$, $\beta^* = ...$.

2. Generate 500 realizations from an AR-1 process $Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$ where $c = 50$, $\phi_1 = 0.6$ and ϵ_t are normally distributed with mean zero and standard deviation 20.

   (a) Plot the generated observations.

   (b) Plot the Auto-Correlation Function (start from observation 100 to eliminate the effects of initialization).

   (c) Implement a naive forecast on the data $\hat{y}_t = y_{t-1}$. Compute the RMSE starting from period 100.

3. Investigating the Auto-correlation structure of the Coffee Price Index Data.

   (a) Plot the ACF and PACF of the coffee price index time series. State if there are interesting observations.

   (b) Difference the data: $u_t = y_t - y_{t-1}$. Plot the ACF and PACF of the differences u_t. State if there are interesting observations.

   (c) Based on the previous AC analysis, are you able to propose a model for the original data series?