INDR 450/550 HOMEWORK 2, Due Date: April 4, 2022

- Please work in groups of two or three or (individually) and submit one file for each group with all names.
- The first data set is the same as in HW 1: 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).
- The second data set is the quarterly beer production in Australia (in megaliters) starting from the first quarter of 1956 to the second quarter of 1994.
- Please perform all computations in python. You are expected to implement your own forecasts (don't use packages/functions from the web). Please submit (upload 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. (40 points) Forecasting Coffee Prices (monthly average price index of coffee in commodity markets as reported by the IMF commodity price data portal). This continues from the first HW.
   (a) Plot the data and visually assess whether there is significant trend and seasonality.
   (b) Check the ACF and PACF plots after detrending and deasonalizing (if necessary). Check whether there is significant AC visible.
   (c) Fit an ARIMA model to the whole data set based on the autocorrelation structure and the patterns (trend, seasonality etc.). Explore the significance of the fitted coefficients, the residual diagnostics and assess its performance by calculating the MAE, MAPE and RMSE.

(d) Experiment with a different ARIMA model for comparison.Explore the significance of the fitted coefficients, the residual diagnostics and assess its performance by calculating the MAE, MAPE and RMSE. Compare with the previous model.

(e) Separate the data into a training set and a test set (first 70 to 80 % of the data should be the training set and the remaining data test set). Fit the two models above on the training set and compute the forecast errors on the test set.

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

2. (40 points) We had looked at the monthly Australian Beer Production data. This time we investigate the quarterly production data starting from the first quarter of 1956 to the second quarter of 1994.

(a) Plot the generated observations.

(b) Fit a SARIMA(0,1,0)(0,1,0,4) model to the whole data (i.e. seasonal differencing of 4 quarters and first order differencing). Explore the significance of the fitted coefficients, the residual diagnostics and the error performance. This could be a benchmark model for other comparisons.

(c) Detrend and deseasonalize the data. Plot the Auto-Correlation Function and the Partial Auto-Correlation Function of the detrended and deseasonalized data.

(d) Based on the ACF and PACF, fit a SARIMA model to the whole data that has some AR and MA coefficients. Explore the significance of the fitted coefficients, the residual diagnostics and assess its performance by calculating the MAE, MAPE and RMSE. Compare with the previous benchmark model.

(e) For the best model you found, find a one-quarter ahead forecast for quarter 155 (using all data available until the end of quarter 154) and report the 95 % prediction interval for your forecast.

(f) Separate the data into a training set and a test set (first 70 to 80 % of the data should be the training set and the remaining data the test set). Fit the two models above on the training set and compute the forecast errors on the test set.

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

Table 1: Summary of Results for Exercises 1 and 2

Method	**Spec.**	**RMSE (Train)**	**RMSE (Test)**
Benchmark (from HW1)	-		
Model 1	$\phi_1 =$, $\theta_1 =$, etc.		
Model 2			
Model 3 (if any)			

The model specification includes the ARIMA specification (i.e. ARIMA (1,0,2) or SARIMA(1,0,2)(0,1,0,4)) and the fitted coefficients.

3. (20 points) Fit ordinary linear regression models to the quarterly Australian Beer Production data.

   (a) Fit a linear regression model that uses seasonal dummies as predictors (note that there are four quarters and three corresponding seasonal dummies are needed). Check the R^2 and the significance of the coefficients and compute the MSE.

   (b) Fit a linear regression model that uses seasonal dummies and a linear trend term as predictors. Check the R^2, adjusted R^2 and the significance of the coefficients and compute the MSE.

   (c) For both of the models in part a and part b, find a one-quarter ahead forecast for quarter 155 (using all data available until the end of quarter 154) and report the 95 % prediction interval for your forecasts.

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

Table 2: Summary of Results for Exercise 3

Method	**Spec.**	R^2	**RMSE**
Benchmark (from HW1)	-		
Model 1	$\beta_0 =$, $\beta_1 =$, etc.		
Model 2			
Model 3 (if any)			