***Project Description***

**Tasks & Dataset files:**

**Task 1**

In a study of potential effects of both climate and pollution on disease specific mortality between the years 2010-2020 a team of researchers studied the disease specific averaged weekly mortality in Paris, France and the city’s local climate (temperature degrees Fahrenheit), size of pollutants and levels of noxious chemical emissions from cars and industry in the air - all measured at the same points between 2010-2020.

All 5 series i.e. mortality, temperature, pollutants particle size and two chemical emissions (chem1, chem2) between 2010-2020 (508 time points) are given in mort .csv Download mort .csv

You will use this data for the calculation of 4 weeks ahead forecasts for mortality.

Your task is to give best 4 weeks ahead forecasts in terms of R squared, AIC, BIC, MASE etc (as is appropriate) for the mortality series. Provide the point forecasts and confidence intervals and corresponding plot for the most optimal model for each method used (DLM, ARDL, polyck, koyck, dynamic, exponential smoothing and state-space model).

Multiple predictors are to be modelled i.e., use more than 1 predictor in regression type models, (i.e. multivariate). Point forecasts and confidence intervals should be reported along with appropriate graphs. Percentiles method for relevant covariates for the forecasts can be used.

Hint: Use MASE () function from the dLagM package to compute MASE for time series regression methods for model comparisons.

**Task 2**

In a study of 81 species of Australian plants Hudson & Keatley (2021) investigated whether the day of occurrence of a species first flowering (first flowering day, FFD, a number between 1 -365) is impacted by climate factors such as rainfall (rain), temperature (temp), radiation level (rad), and relative humidity (RH). The study by Hudson & Keatley essentially explores the influence of long-term climate on the FFD of 81 species of plants from 1984 to 2014.

Your data focuses on one species (of the 81) and contains 5 time series, the FFD time series of the given plant species and the contemporaneous yearly averaged climate variables measured from 1984 – 2014 (31 years). All series are available here in FFD .csv Download FFD .csv

Your task is to model FFD and forecast FFD. Single climate predictors (univariate models) are to be tested. Your task is to give best FFD 4 years ahead forecasts for the FFD series. Point forecasts and confidence intervals are required for the forecasts with appropriate graphs.

For forecasting, the future covariates are given here in Covariate x-values for Task 2 .csv Download Covariate x-values for Task 2 .csv

Provide the point forecasts and confidence intervals and corresponding plot for the most optimal model for each method used (DLM, ARDL, polyck, koyck, dynamic, exponential smoothing and state-space model).

Use MASE() function from the dLagM package to compute MASE for time series regression methods for model comparisons. Choice of optimal models within a specific method can be assessed from values of R squared, AIC, BIC, MASE etc (as is appropriate to the method).

Hints:

* Perform univariate analyse of predictors.
* Run models with and without intercept.
* Interpret the lag effects – note positive lag coefficient means that flowering is later and negative ones earlier flowering (lower FFD).
* Intercept denotes the intercept mean FFD. You are to analyse one given species (on Canvas).
* Plots are essential ACF PACF, forecasts etc.
* Percentiles method for relevant covariates for the forecasts can be used.

**Task 3**

Climate change has become a significant issue in Australia. Australia's extreme climate is vulnerable to climate shifts and has experienced many extreme weather events recently. Research has shown that climate change is influencing flowering timing and the relative flowering orders of plants. Changes in flowering order can have detrimental effects on plant growth and food availability. Little research has examined how flowering orders have changed over time. This data from Hudson & Keatley (2021) explores the influence of long-term climate shifts in Victoria on the relative flowering order similarity of 81 species of plants from 1983 to 2014. The species were ranked annually by the time taken to flower (FFD), and changes in flowering order were measured by computing the similarity between annual flowering order and the flowering order of 1983 using the Rank-based Order similarity metric (RBO). The earliest flowering species is ranked 1 and latest ranked 81 for the given year under study.

A decrease in similarity over time suggests that flowering orders are becoming more dissimilar from their original ordering in 1983. The aim of your analysis is to see if it is possible to determine how flowering orders respond to specific changes in climate by understanding the relationship between flowering order (RBO) and climate conditions. Climate variables are yearly averaged temperature, rainfall, radiation and relative humidity. These relationships may be used – in theory – to forecast future flowering orders. Note the term "flowering order similarity" refers to the similarity between the flowering order of the 81 plant species for a particular year and the flowering order of the same plants in the so-called baseline reference year, 1983 (the first year in the data collection period with enough observations) measured by a rank similarity metric, RBO. RBO values are therefore numbers between 0 and 1. Higher RBO values indicate higher similarity of the order of the first flowering occurrence (based on FFD) of the 81 species from 1983 compared to each of the subsequent years, 1984 to 2014, so the time series are of length 31.

Models will be built using RBO similarity values for the flowering orders of consecutive years as the dependent time series (Yt). Your data contains 5 time series, the RBO time series of the 81 plant species studied by Hudson & Keatley (2021) and the yearly averaged climate variables measured from 1984 – 2014, as 1983 is the reference year for flowering order FFD ranks. The time series are thereby of length 31, as the RBO metrics utilise 1983 rank orders of FFD as the baseline.

All series are available here in RBO .csv Download RBO .csv

For forecasting, the future covariates are given here in Covariate x-values for Task 3 .csv Download Covariate x-values for Task 3 .csv

Task 3 Part (a): Carry out your analysis based on univariate climate regressors (model one climate indicator at a time, i.e., univariate regressor).

Modelling methods to try (DLM, ARDL, polyck, koyck, dynlm).

Choice of optimal models within EACH a specific method can be assessed from values of R squared, AIC, BIC, MASE etc (as is appropriate to the method).

The goal is to forecast RBO three years ahead using each regressor one at a time and (use percentiles for the regressors) in forecasting for each of the best models within the methods utilised.

Point forecasts and confidence intervals are required to be obtained and reports for the forecasts. Percentiles method for relevant covariates for the forecasts can be used.

Task 3 Part (b): Flowering orders became more dissimilar over the most recent decades, particularly during the Millennium Drought (1997 – 2009), suggesting that flora in Australia is responding to changes in their environment. According to the BoM the drought period for Australia occurred from 1996 to 2009 How would you now accommodate for this in your analysis of the Rank RBO. Perform the appropriate analysis and obtain the 3 year ahead forecasts (suggest using the dynlm package) only for part (b)).