**Filelist**

**R Files**

* **P1-Main-Code.R:** This is the main code which produces the results for the loaded target variable.
* **P2-functions.R:** Necessary file with custom-made functions which is loaded in P1-Main-Code.R

**Required R Libraries**

Before running the codes, the user is required to have installed "forecast", "lubridate", "randomForest", "xgboost", "gbm", "kknn", and "glmnet" libraries.

**Data Files (.csv)**

* **namq\_10\_gdp\_working.csv:** The data for quarterly GDP to be used in the econometric analysis.
* **namq\_10\_gdp\_comparison.csv:** The data for quarterly GDP to be used in the evaluation of results.
* **weights.csv:** The data which includes the EU27 weights for Production/Output, Retail Trade and Producer Prices variables.

**Output from the above R files using the above data files.**

* **namq\_10\_gdp.RData:** all the environment is exported in this .RData file.
* **namq\_10\_gdp\_CR1.csv:** the coverage rates for the confidence intervals.
* **namq\_10\_gdp\_CR3.csv:** the coverage rates for the confidence intervals.
* **namq\_10\_gdp\_CR5.csv:** the coverage rates for the confidence intervals.

Replacing the data files above with any file/variable of your choice, you can re-run the same code and obtain the new output. The GDP is used as a generic example here.

**Instructions**

The R codes are written with a large number of detailed comments and instructions that explain the purpose of each variable and part of the code. Below, we provide some guidance to facilitate the applied research and promote the applicability of our methodological report.

**Line 11:** All environment is empty, and the code starts with clean memory.

**Line 15:** The user is required to type where the final output will be stored.

**Line 18:** The user is required to type the exact location of the **P2-functions.R** file which is necessary for the calculation of results.

**Line 21:** The necessary libraries are loaded; the user is required to have installed "forecast", "lubridate", "randomForest", "xgboost", "gbm", "kknn", and "glmnet" libraries.

**Line 29:** The user is required to type where the exact location of the datafile which contains the data for the “working” time series; i.e. the time series which will be used in the econometric analysis. For the hereby discussed GDP example, this series is the GDP measured in millions.[[1]](#footnote-1)

**Line 32:** The user is required to type where the exact location of the datafile which contains the data for the “comparison” time series; i.e. the time series which will be used in the evaluation of the estimates. For the hereby discussed GDP example, this series is the GDP growth.[[2]](#footnote-2)

**Line 35:** The user is required to type the prefix to be used in all the output files. For example, using fnam3 <- "namq\_10\_gdp" will return the following output: namq\_10\_gdp.RData, namq\_10\_gdp\_CR1.csv, namq\_10\_gdp\_CR3.csv and namq\_10\_gdp\_CR5.csv. But using, say, fnam3 <- "naidq\_10\_gdp" we will obtain the following output: naidq\_10\_gdp.RData, naidq\_10\_gdp\_CR1.csv, naidq\_10\_gdp\_CR3.csv and naidq\_10\_gdp\_CR5.csv.

**Line 47:** The user is required to type where the exact location of the datafile which contains the weights to be used in the EU27 aggregate.

**Line 48:** The user is required to choose one of the following options: 1, 2, 3 or 4. The options are as follows: 1 means that the EU27 aggregate does not use weights (i.e. all countries have a weight of 1), 2 uses the Eurostat weights for the Industrial Production, 3 uses the Eurostat weights for the Retail Trade and 4 uses the Eurostat weights for the Producer Prices.

**Line 53:** This allows the user to choose between the Eurostat official EU27 aggregate or a calculated EU27 aggregate (default option is 1 and it is advised not to be changed).

**Line 63:** The user is required to choose the stationarity transformation to be used. Two options are allowed: “fdiff” which uses the first difference and “ldiff” which uses the log growth.

**Line 66:** Additional option of stationarity transformation which is used in the comparison series; usually, for consistency, this matches the previous option in Line 63 and this is why it is set to be identical.

**Line 71:** The user is required to set the out-of-sample periods. This is chosen from the bottom part of the dataset. For example, using usr\_CV <- 21 will use the bottom (i.e. most recent) 20 observations as the out-of-sample.[[3]](#footnote-3)

**Line 87:** Some machine learning methodologies, e.g. Random Forests, have a natural element of "randomness" in their calculation. This option controls the number of iterations. A relatively large number is required. It is set to 200 (as the default value) which already slows down the calculations.

**Line 90:** The user is required to type, in the form of a list() object, the missing value cases, i.e. which countries we suppose that they have missing values and, hence, they will be estimated to produce the EU27 aggregate estimate.

--- The remaining lines do not require any input from the user. ---

**Lines 111 – 118:** the “working” data file is loaded and manipulated to a matrix() object.

**Lines 121 – 124:** the “comparison” data file is loaded and manipulated to a matrix() object.

**Lines 127 – 129:** internal checks and errors to the user in case of mismatching of the above matrices.

**Lines 132 – 135:** the “weights” data file is loaded and manipulated to a matrix() object.

**Lines 138 – 144:** weights are set to a global variable (to be used later in the EU27 aggregation) and data is organised in different variables.

**Lines 147 – 154:** variables are transformed to stationarity.

**Lines 157 – 172:** sparse objects are introduced which will hold the results/calculations.

**Line 175:** loop across missing value cases opens.

**Lines 188 – 200:** sparse objects are introduced which will hold the results/calculations.

**Line 202:** loop across the out-of-sample dates opens.

**Lines 207 – 218:** available data is selected and missing values are added.

**Lines 223 – 290:** all forecasts from the univariate models in levels are obtained.

**Lines 302 – 383:** all forecasts from the univariate models in stationary transf. are obtained.

**Lines 386 – 443:** all forecasts from the machine learning models in stationary transf. are obtained.

**Line 446:** loop across the out-of-sample dates closes.

**Lines 449 – 538:** all output for the specific missing value case is calculated.

**Line 539:** loop across missing value cases closes.

**Lines 541 – 560:** flatten of missing value cases to be used as column names.

**Lines 562 – 565:** output is saved in .RData and .csv files.

1. In particular, we use the GDP which is the output of the following options in the "namq\_10\_gdp" Eurostat table: "B1GQ", "SCA", and "CLV15\_MEUR". [↑](#footnote-ref-1)
2. In particular, we use the GDP which is the output of the following options in the "namq\_10\_gdp" Eurostat table: "B1GQ", "SCA", and " CLV\_PCH\_PRE". [↑](#footnote-ref-2)
3. One observation is added for the stationarity transformation. For example, if we require 20 periods in the out-of-sample we select usr\_CV <- 21 where the first is used to calculate the stationary version of the series (i.e. 21-1=20). [↑](#footnote-ref-3)