# Course Project for ECM607C
# Macroeconometrics

## INSTRUCTION:

This course project aims to build the ARMA model for the time series of the foreign exchange between US dollar and British pound and evaluate the forecasting performance of the model. The econometric analysis should be implemented by the statistical software R and its integrated development environment (IDE), R studio.

The project report should be written in the style of academic article, including abstract, introduction, econometric analysis, conclusion, and references (in <u>Harvard style</u>[1]). For the econometric tasks, you need to report the econometric results from R, and, more importantly, provide the technical description on the methods and the interpretation/discussion on the econometric results.

For example, you should include the following content for the task of ADF test:
- technical description of ADF test;
- how to choose the setting of ADF test (whether you want to allow drift and/or trend; how to select lags);
- R code to conduct the analysis;
- R output (**Important**: the results should be presented in nicely formatted tables, rather than the raw screenshot from R studio);
- interpretation and discussions on the results.

The deadline for the course project submission is *April 20<sup>th</sup>, 2020 (12.00 noon)*. Only electronic submission (PDF file) is required via Blackboard.

## DATA:

The file "*EXUSUK.xls*" contains the monthly exchange data from *Jan 1971 to Feb 2020*, in total 590 observations. The data was obtained from the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/series/EXUSUK). The data webpage also provides more information on the time series. Denote the original exchange rate as $Y_t$, and the variables that we are interested are the logarithm of the exchange rate $\log(Y_t)$ and its growth rate $\Delta \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$. **Important**: note that you will lose the observation in Jan 1971 when you calculate $\Delta \log(Y_t)$, and you need to delete the observation of $\log(Y_t)$ in Jan 1971 accordingly.

## ECONOMETRIC TASKS:

Before carrying out the following tasks, firstly split the dataset into in-sample period and out-of-sample period. The in-sample period is from *Feb 1971 to Dec 2012*, and the out-of-sample period is from *Jan 2013 to Feb 2020*. **Important**: you should use the data in the in-sample period[2] to perform Tasks a–d and then make forecasting for the out-of-sample period in Tasks e–f.

---

[1] https://libguides.reading.ac.uk/citing-references/referencingstyles
[2] To be clear, use data between Feb 1971 and Dec 2012 to conduct Tasks a-d.

Task a.   **[Pre-analysis]** Provide the data visualisations (time series plot, box plot, histogram, and ACF plot) and the descriptive statistics of $\log(Y_t)$ and $\Delta\log(Y_t)$. What can you spot from various plots? What is the meaning of the descriptive statistics?

Task b.   **[Unit root/stationarity tests]** Perform ADF test, PP test, and KPSS test on $\log(Y_t)$ and $\Delta\log(Y_t)$. You can use AIC to select the best lag length for the ADF test and use "short" lag length for the PP and KPSS test.

Task c.   **[Model selection]** Select the best $ARMA(p, q)$ model for $\Delta\log(Y_t)$ by AIC. Please set the maximum orders as 5, i.e. $p, q \leq 5$.

Task d.   **[Model diagnostics]** Perform model diagnostics for the residuals from the best selected model. To be specific, carry out the time series plot, ACF plot, and Ljung-Box test on the residuals. For the Ljung-Box test, please set the lag lengths as 20, 30, and 50. Comment on the results of the model diagnostics.

Task e.   **[Forecasting]** Choose the best model based on the data in the in-sample period and make forecasting for the out-of-sample period. Plot the predicted values from the best model versus the true values. Calculate the forecasting performance (MSE, MAE, MAPE, and %correct sign) of the best model.

Task f.   **[Rolling-window Forecasting]** The idea of rolling-window forecasting is to iteratively make one-step-ahead prediction using the most recent $w$ data observations, which is referred to as the window of data. In practice, we have one new monthly observation coming in each month. Thus, it is potentially helpful for forecasting if can we include the newest observation (whenever it becomes available) and exclude the oldest observation (since it become less relevant). For this rolling-window forecasting task, you need to update the best model (and its parameters) in each window and then only make one-step-ahead prediction. For each time-point $t$ in the out-of-sample period, you need to use the most recent $w$ observations (i.e. from $t - w$ to $t - 1$) to make prediction for the value at time-point $t$. In this task, please set the window length $w = 503$. To be fully clear:

- use data from Feb 1971 to Dec 2012 to select the best model, estimated the model, and make the prediction for Jan 2013 only.
- use data from Mar 1971 to Jan 2013 to select the best model, estimated the model, and make the prediction for Feb 2013 only.
- ⋮
- use data from Mar 1978 to Jan 2020 to select the best model, estimated the model, and make the prediction for Feb 2020 only.

Collect the predicted values from each rolling-window. Plot the predicted values versus the true values. Calculate the forecasting performance (MSE, MAE, MAPE, and %correct sign) of the rolling-window forecasting method. Compare the forecasting performance of rolling-window forecasting method in Task f and the forecasting method in Task e.

[*Hint: 1. write a function to produce one-step-ahead prediction based on a given dataset; 2. call that function in a for loop; 3. store the predicted values in a container.*]

*Tips: you need to provide the technical description of the econometric tools, how to choose the setting, R code, econometric result, and interpretation/discussion for each task above. Only providing the result is not adequate.*

**Academic misconduct**:
- ***You should do this coursework independently and you should not share your answers with others***. All submitted coursework will be checked by the Turnitin. Any case of plagiarism will be reported.
- To respect the fairness to every student, I will make response to queries only related to the clarity issues, and **'answer-snooping' queries should be discouraged**.

## RUBRICS (Marking Criteria):

| | Weights in the row | | | | Weight of each section |
|---|---|---|---|---|---|
| | Technical Description (and Setting) | R Code | Results | Interpretation/ Discussion | |
| Pre-analysis | 25% | 25% | 25% | 25% | 10% |
| Unit root/stationarity tests | 25% | 25% | 25% | 25% | 20% |
| Model selection | 25% | 25% | 25% | 25% | 10% |
| Model diagnostics | 25% | 25% | 25% | 25% | 10% |
| Forecasting | 25% | 25% | 25% | 25% | 20% |
| Rolling-window Forecasting | 25% | 25% | 25% | 25% | 20% |
| Abstract, Introduction, and Conclusion | | | | | 5% |
| Overall formatting and reference | | | | | 5% |
| Total Weight | | | | | 100% |