

# Goldsmiths College, University of London

## Advanced Econometrics

Lecturer: **Tomás Rotta**

### Project

The purpose of the project is for you to demonstrate that you can **code everything correctly in R** and that you can **correctly interpret the results** from the estimations of the different approaches and models. It is not enough to run the code correctly. You also need to give your interpretation of the results. For example: if you find that two series are cointegrated, what does that mean? Should we expect these series to be cointegrated in the first place? Why? Should we expect A to cause B, or for B to cause A? Why? When you run your ARDL model, what interpretation do you offer for the coefficient estimates? Why would you prefer a VEC model over any other model? When you run your panel data model with fixed effects, how do you interpret the regression estimates?

### Instructions

Complete **part 1** of this project. **Part 2 is optional** but I encourage you to complete it to train your panel data and coding skills, which are valuable skills in the job market and they look good on your CV. There is no grade penalty if you only complete part 1.

Write the **R code** for all steps. **Submit your project together with your R code as a single PDF file on the Moodle VLE**. The **R code** embedded within the project file does **not** count toward the **2,000-word limit**. A template in Word format is available on the Moodle page.

The easiest way to format your project is to write it in MS Word, combining the R code and the R output step-by-step for each task, pasting both the output text and the plots. Then save the Word file as a PDF and submit it on the Moodle VLE. The submission is anonymous but please include your student ID# on the cover page.

If you are the type of student who loves a new challenge, you can write your entire project in R using **Rmarkdown**, which is a free open source package for R. **Rmarkdown** is fully integrated into Rstudio, even though you can run Rmarkdown without Rstudio. More information on Rmarkdown is available on our Moodle page. **Rmarkdown** allows you to automate many features in your own project. On our Moodle page you will also find a **template Rmd file** which contains the R code to render your project in PDF, HTML, and Word formats using Rmarkdown. I suggest you render your project in PDF format as in my template.

## Part 1: Time Series

Use the following two time series with **monthly frequency**:

Capacity utilisation rate: <https://fred.stlouisfed.org/series/TCU>

CPI inflation rate: <https://fred.stlouisfed.org/series/CPIAUCSL> (\*)

(\*) You need to use the monthly percentage change (i.e. growth rate) of CPI from one year ago, which means you are actually using the **monthly inflation rate** but correctly calculated **year-on-year** (not month-on-month). Do **not** use the monthly inflation rate calculated month-on-month!

Explain what the data are about and how each variable is measured.

**Plot** both series in levels and in first differences.

Run the **ADF**, **PP**, and **KPSS** unit root tests. Are the series stationary  $I(0)$  or nonstationary  $I(1)$ ?

Run the **Engle-Granger cointegration test**. Are the two variables cointegrated?

Run the **Johansen cointegration test**. Are the two variables cointegrated?

Run the **ARDL** model using the “dynamac” package in R, assuming that one of the variables is exogenous and the other is endogenous. Recall that the ARDL model combines variables in differences (for the short-run effects) and variables in levels (for the long-run effects). You can test for the significance of the ARDL long-run relationship in levels if you set “ec=TRUE” when running the ARDL model and then running the command for the **PSS bounds test**. If the F statistic is above the  $I(1)$  critical value you can conclude that the long-run variables in levels are jointly significant. Remember to also run the **diagnostic checks** (Shapiro-Wilk, Breusch-Godfrey, Jarque-Bera etc.). Finally, set “simulate=TRUE” and then plot the **impulse-response functions**.

Run the **VAR model**, assuming that both variables are endogenous to each other. If the Johansen cointegration test or the Engle-Granger cointegration test tell you that there is cointegration then you need to include the **error correction mechanism (ECM)** in the VAR model. This in fact means that you need to estimate the **Vector Error Correction (VEC) model**. If there is no cointegration than use the VAR model, not the VEC model.

Run the **Granger-causality tests** in both directions.

Use a **Cholesky decomposition** to make the structural errors orthogonal to (independent of) each other. Plot the **impulse-response functions**.

Run the **ARIMA model** for one of the two times series. Which ARIMA model has the best fit? You can use the “auto.arima()” function in R to automatically detect the best model. Use the estimated ARIMA model to **forecast** the series in the next 10 months. Plot your forecast.

## Part 2: Panel Data (Optional)

Open the R code file for panel data models, then scroll down to the last section titled "Panel Data Example using World Bank Data (WDI 1960-2015)".

Explain what the data set is about and how each variable is measured.

**Include one more variable in your panel data model** that is available in the WDI dataset but not already included in this example in the R code. Explain why you have chosen to include the additional variable.

Run the **Random Effects (RE)** model.

Run the **Fixed Effects (FE or 'within' model)**.

Run the **Pooled OLS** and then **test for poolability**. What do you conclude?

Run the **Hausman test** and decide which model is preferred.

Run the **diagnostic tests on the residuals** of each model. Are the residuals **white noise** (free from auto-correlation and heteroskedasticity)?

\*\*\*