Linear Time Series Assignment :

ARIMA modelling of a time series

To be submitted:

- **before Thursday, May 22 at midnight** by dropping it on this Dropbox link
- into a .zip file :
  - called "*fname1_fname2*" where "fname1" is the first letter of the first member's first name and their name.
  - containing **the PDF report, the .R script and the time series used as a .CSV file**.

Guidelines : The assignment must be done with R. Scoring will take into account the rigor and exactitude of the use of econometric tools, of the conciseness and clarity of the results' presentation. The report (a maximum of 6 pages, appendix not included, in English) will show the commented programs in the appendix. This assignment **must be done in pairs**.
Scoring is indicative : Part 1 : 30%; Part 2 : 30%; Part 3 : 40%.

We are interested in the modelling and predicting of an observed time series of the French economy. It is strongly advised to pick a French Industrial Production Index (IPI) series. (link to the IPI series). You will only work on observed data. Using the INSEE's time series databank, you must choose an aggregate series corrected from seasonal variations and working days (CVS-CJO), in monthly or quarterly frequency, from any sector of the industry and containing at least 100 observations. You can use another series as long as it checks the previously mentioned characteristics (avoid current-euro series, price data or financial data).

- **Part I : the data**
  1. What does the chosen series represent ? (sector, potential data processing, logarithmic transformation, etc.)
  2. Transform the series to make it stationary if necessary (differentiate it, correct the deterministic trend, etc.). Thoroughly justify your choices.
  3. Graphically represent the chosen series before and after transforming it.
- **Part II : ARMA models**
  4. Pick (and justify your choice) an ARMA(p,q) model for your corrected time series X_t. Estimate the model parameters and check its validity.
  5. Write the ARIMA(p,d,q) model for the chosen series.
- **Part III : Prediction**
  Denote T the length of the series. Assume the series' residuals are Gaussian.
  6. Write the equation verified by the confidence region of level α on the future values (X_{T+1}, X_{T+2}).
  7. Give the hypotheses used to get this region.
  8. Graphically represent this region for $\alpha = 95\%$. Comment on it.
  9. Open question : let Y_t a stationary time series available from $t = 1$ to T. We assume that Y_{T+1} is available faster than X_{T+1}. Under which condition(s) does this information allow you to improve the prediction of X_{T+1} ? How would you test it/them ?