Empirical project 2023

- Form a group of 2 students (at most)
- Use a software of your choice
- Do the analyses as detailed below
- I will provide a sketch of the solutions (i.e., values of the Root Mean Squared Error of models)
- You should be able to explain discrepancies if any
- Hand in via ARIEL a single PDF file with name, surname, and student ID
- Use the template that I provide to answer.
- Do not attach data or code
- Have fun!

1 Dataset

You will be working with one of the most-used datasets in empirical macroeconomics and econometrics, the FRED-QD dataset. It is a large database of US macroeconomic data that you will use to forecast real GDP.

I've already pre-processed the data for you. The database is in the file "DatiProject.xls"

Details and a description are in the file.

2 Forecast Experiment

You will construct 1-quarter-ahead forecasts of Real GDP (log-level): $y_t = \log(GDPC1_t)$ where $GDPC1$ is the code of the series in the dataset (sheet: Data).

All models are estimated using a rolling window of 100 observations. The first forecast origin is $R = 1985 : Q2$, and the last forecast origin is $T = 2018 : Q3$, for a total of $P = 134$ forecasts. At each forecast origin, you will estimate models and save the forecasts from the models, as detailed below.

Forecast evaluation is carried out comparing forecasts for the level with Real GDP $GDPC1_t$ and computing the root mean squared error, RMSE (i.e., exponentiate the forecasts before computing RMSE).

Useful links:

- Website: https://research.stlouisfed.org/econ/mccracken/fred-databases/
- R: https://github.com/cykbennie/fbi
- Python: https://pypi.org/project/FredMD/
- Julia: https://github.com/markushhh/FredApi.jl

NOTE: R code is provided by FRED, Python, and Julia are not and seem to be working with FRED-MD (i.e., monthly data).

3 Models

- Random Walk (without drift)
- AR(4) for $\Delta y_t = (y_t - y_{t-1})$
- VAR(4) model with Δy_t, a measure of inflation $\pi_t = \log(PCECTPI_t) - \log PCECTPI_{y-1}$ and the term spread defined as $Tspread_t = GS10_t - TB3MS_t$ (sheet: Data)
- Same model as before, but select a new lag order "p" with Akaike Information Criterion each time you estimate the model to produce a new forecast.
- An AR(4)-X model for Δy_t that includes the lag of the first principal component ($\hat{F}_{1t-1}$) as exogenous variable. $\hat{F}_{1t-1}$ is computed using all the variables in the FRED-QD database, except $GDPC1_t$. (sheet: DataForFactors)

$$\Delta y_t = a + \sum_{j=1}^{4} b_j \Delta y_{t-j} + c\hat{F}_{1t-1}$$

  You can apply PC analysis directly to the series in the sheet "DataForFactors". These have already been transformed to stationarity, but you have to standardize them (zero mean, unit variance).

- (BONUS) Propose your own model.

4 Details

1. Section 1: Plot all the series that you are using in the empirical analyses (not all the series in the database): y_t, Δy_t, $\log PCECTPI_t$, π_t, $Tspread_t$.
2. Section 2: use the first sample of 100 observations that ends in $R = 1985 : Q2$. Plot the sample Autocorrelation (ACF) function of the series entering the VAR(4) model. Use AIC to select lag order. Estimate the model. Plot the residuals of the three series and their sample ACF and Partial ACF.
3. Section 3: use the first sample of 100 observations that ends in $R = 1985 : Q2$. Plot the PC factor and try to interpret it.
4. Section 4: write the equations for all models, estimate them and produce forecasts. Provide a separate plot for each forecast against real GDP. Provide a table with RMSE computed for real GDP (i.e. $GDPC1_t$).

NOTE: All sections need to be shortly commented with the aim of interpreting results and not explaining what the code does.