ECON-UB 251

Assignment 4, Fall 2022

The learning goals of this assignments are:

1. familiarize with time series analysis
2. developing an understanding of the concept of non-stationarity and its implications
3. use time series models in forecasting

I *strongly* prefer that you complete the assignment in Rmarkdown and a sample template is provided in Brightspace. You can knit the document to Word or PDF (or export to PDF the Word document). Set the `echo` option to `TRUE` so that I can see the code you are using to conduct the analysis. You can discuss the assignment with other students, but each student should submit his/her original work.

Submit in `Brightspace` by 2pm on **Tuesday December 13th, 2022**.

The goal of this assignment is to analyze the properties of an economic or financial time series and to construct a forecasting model that can produce accurate forecasts of the release of the variable in the following period.

1. [**20%**] **Pick an economic or financial time series** (for the US or any other country) that *you* are interested in modeling from FRED, Yahoo Finance or any other source (please reference the data source in your assignment). For data downloadable from FRED and Yahoo Finance you can use the two `R` packages used in the slides:

```
# examples of how to use the alfred and tidyquant packages
library(alfred)
payems <- get_fred_series("PAYEMS", observation_start = "1950-01-01")

library(tidyquant)
stocks  <- tq_get("SPY", get = "stock.prices", from = "1990-01-01")
payems  <- tq_get("PAYEMS", get = "economic.data", from = "1950-01-01")
```

- Download the time series for your chosen variable from the earlierst available date to the latest available (it's fine to use a shorter sample as long as you provide a justification)
- Show the code where you are reading or downloading the data
- Print the first and last three rows of the data frame using `head()` and `tail()`

2. [**10%**] Plot the time series and provide a qualitative discussion of the main features of the time series

- For several variables (e.g., CPI, GDP, stock prices) you might want to consider the *logarithm* of the variable rather the level of the variable (see examples in the slides)
- Features to discuss (among others): range of variation, persistence, trend, outliers, behavior during recessions/expansions, impact of the COVID-19 pandemic

3.[**10%**] Plot the ACF of the time series up to 20 lags and discuss the results.

- For this question (and the following questions that involve estimation) check the sensitivity of your results to excluding the pandemic (i.e., observations after `"2019-12-31"`)

- If the results are similar use the full sample but if they are significantly different exclude the pandemic. Justify your choice and provide in your report only the results for the sample you decided to use.

4. [**20%**] Test the non-stationarity of the series using the Augmented Dickey-Fuller (ADF) test using the `adf.test` function from package `aTSA`. Perform the test at 5% significance level and discuss whether a drift or trend should be considered in the ADF test. Draw the appropriate conclusion whether you should model the series in level (or log-level) or take the first difference.

5. [**10%**] Based on the conclusion for the previous question, estimate an AR(1) and AR(4) model for the variable that you decided to use.

- Calculate BIC for both models and select the best model (make sure to adjust the sample to account for the lags)
  - AR(1) and AR(4) are general guidelines and you might want to use other lags based on the frequency of your time series
- Provide the estimation results for the selected model and discuss the goodness-of-fit

6. [**10%**] Consider an ADL(p,q) model by adding one X variable that you believe could be relevant to predict your time series

- First, consider if the additional variable needs transformation to first difference (or log difference) because of non-stationarity
- Fix the p in the ADL specification to the AR order selected in the previous question; for q use 1 and 4 and choose the optimal value based on BIC
- Report the estimation results of the ADL model and discuss its goodness-of-fit in comparison to the AR model in the previous question

7. [**10%**] Use the estimated AR(p) and ADL(p,q) models to forecast the chosen variable in the period after the last one that you downloaded

8. [**10%**] Construct a 90% forecast interval for the one-period ahead forecast (use the in-sample SER for the RMSFE)