Please submit your answers by midnight (11:59pm) on Friday, 11th March 2022 **via Canvas only**. Late penalty: -2% for each hour past deadline. Your discussion of results (50% of the marks) **and** working code (the other 50%) should be contained in a **single** Jupyter Notebook using markup functionality or commentary. Clearly identify the parts of the project by sectioning (e.g., using markup section ## Question 1.1, Question 1.2, etc...). You don't need to provide data. In marking your empirical project, I will be executing each of your Python notebooks, so please make sure that the code is working to avoid a loss of 50% marks for code that does not compile or throws an error on execution.[1]

When submitting your files please:

1. Use the file-naming convention: LASTNAME, FIRSTNAME - PROJECT 1.
2. At the beginning of your Jupyter Notebook file, indicate whether you:
   (a) agree to share your work on Dropbox with the rest of the class with your identification intact ("I-am-proud-of-my-work!" option) [default option if you forget to mention it in your submission];
   (b) agree to share your work on Dropbox with the rest of the class but with your identification removed ("I-am-happy-to-share-but-I-feel-shy" option);
   (c) do not like to share your work with anyone.

Question 1

1. [1 mark] Obtain adjusted closing prices from 01-Jan-2015 to 1-Mar-2022 for
   - the DJIA index (Yahoo ticker: ^DJI),
   - Tesla (Yahoo ticker: TSLA), and
   - Freeport-McMoRan Inc (Yahoo ticker: FCX).
2. [3 marks] Before you can proceed with time series modeling, you have to make sure that your data are stationary (does not contain unit root). Perform the following:
   (a) Check your price series for stationarity using ADF and KPSS tests.
   (b) Convert your closing prices to *log returns* and check your return series for stationarity using ADF and KPSS tests.[2]
   (c) What do you conclude?[3] In (a) and (b) above, did you use *constant only* or *constant and a trend* model as as your benchmark and why? Discuss.
3. [1 mark] Plot *cumulative returns* for all three assets on the same graph originating at \$100 (the progression of the \$100 invested on 1-Jan-2015 to 1-Mar-2022). Make sure your x-axis represents dates and the legend with the names of the three assets is visible. Axis labels for x and y should indicate "Time" and "Cumulative Return", respectively.
4. [1 mark] On a 3-by-3 subplot, plot the *returns* in the top row as well as ACF (2nd row) and PACFs (3rd row). Based on your visual inspection of *returns*, ACF, and PACF plots, would you consider an ARMA model?
5. [4 marks] Retain the last 10 observations for checking forecasting ability (your test data), and use the rest of your *returns* sample (your train data) to select the optimal $ARMA(p,q)$ model based on BIC for each of the three assets. Set maximum model complexity to 5 (that is, $p = 0...5, q = 0...5$) and assume Gaussian residuals (this is commonly the default setting in any software).
   (a) Construct a 3D plot with p and q values on x and y axes and BIC on z axis.
   (b) What values of p, q are optimal based on BIC?
   (c) What values of p, q are optimal if you are interested in accuracy of 10-day forecasts from these models based on RMSE? Did you select rolling or recursive forecasting scheme? Why? Discuss.

[1] You only need to include "import" statements for Python libraries you are utilizing. No need to include "pip" or "conda" install commands.

[2] Simply referred to as *returns* hereafter.

[3] In econometrics, "conclusions" are based on hypothesis tests with analyses of p-values and chosen level of significance.

(d) Discuss your findings and propose the final $ARMA(p, q)$ model that you favour the most.

6. [1 mark] Perform Step 5 again, but this time use AIC to select the optimal $ARMA(p, q)$ model. Did your conclusion change?

7. [1 mark] On a 3-by-3 subplot, plot the *squared returns* in the top row as well as ACF (2nd row) and PACFs (3rd row). Based on your visual inspection of *squared returns*, ACF, and PACF plots, would you consider a GARCH type model?

8. [1 mark] Perform Engle's ARCH test for each of the 3 assets to reconfirm your conclusion from the step above.

9. [4 marks] Retain the last 10 observations for checking forecasting ability, and use the rest of your sample to select the optimal $GARCH(p, q)$ model based on BIC for each of the three assets. Set maximum model complexity to 5 (that is, $p = 0...5, q = 0...5$) and assume Gaussian residuals (this is commonly the default setting in any software).

   (a) Construct a 3D plot with p and q values on x and y axes and BIC on z axis.

   (b) What values of p, q are optimal based on BIC?

   (c) What values of p, q are optimal if you are interested in accuracy of 10-day forecasts from these models based on RMSE?

   (d) Discuss your findings and propose the final model that you favour the most.

10. [1 mark] Perform Step 9 again, but this time assume Student t residuals when fitting $GARCH(p, q)$ models. Did your conclusion change?

Note: Many of the steps in this empirical project are repetitive with only few parameters varied. You can substantially simplify/reduce your code if you use/define functions via "def". Your code efficiency will not go unnoticed.