For this assignment please compile a word document using Rmarkdown and submit. You could also submit a pdf file.

Problem 1. Simple Regression and normality test (40 points)

For this problem you will need to download daily data for equities on the US markets. To do so you may use any data source. For example, you can use a Bloomberg terminal or even Yahoo or Google Finance. Please download daily equity trade data for Coca Cola Company (KO) and PepsiCo (PEP). Please download two different periods: from January 2017 to Dec 2019 as DATA 1 and from January 2020 to Dec 2021 as DATA 2. You need to download minimum the day of the record, Opening day value, Closing day value, and the volume of shares traded.

1. Import the data in your program. Please calculate the daily simple returns for each of the two stocks using:
$$R_i = \frac{C_i - O_i}{O_i},$$
   where C_i is the close value of day i and O_i is the opening value for day i. Please pay attention to the order of days in your imported dataset. For this part submit the 5 number summary and a histogram for daily return of each equity. Perform separate analysis for DATA 1 and DATA 2. Please comment on the differences/similarities you observe.

2. Perform a normality test using Coca Cola daily simple returns for each of the two periods. What is the conclusion at the 95% confidence level?

3. Perform a regression using return for Coca Cola as response and Pepsi as predictor. What is the R^2? Comment on the differences you observe between the two periods.

4. Test if the coefficients are equal to 0. Give the p-value of your tests. State the conclusions in terms of the problem. Write down the regression relationships.

5. Predict the return of Coca Cola if the daily Pepsi return is 0.01 and if it is -0.007. Give a 95% confidence interval for each prediction. Compare your prediction results from the two periods.

6. Consider the daily returns of the two companies for the two periods. Test whether or not the returns for these companies are the same or if they are significantly different statistically for DATA 1 and DATA 2. State the null and alternative hypotheses. What is the p-value of the test. Write your conclusion in terms of the variables and parameters under study. Give a 95% confidence interval for the difference in daily returns.

7. Repeat the previous part but this time compare returns between the two companies. For this part please perform a matched pairs test where the returns are matched every day. Comment on the differences you observe between the two companies.

Problem 2. Autocorrelation (30 points) Using the returns of Coca Cola you calculated in Problem 1, do the following for each of the two periods (DATA 1 and DATA 2).

(a) Calculate the daily log returns.

(b) Use R's `acf` function to plot the first 30 lags of the autocorrelation function.

(c) Provide the values of the first 5 lags of the autocorrelation function.

(d) Discuss the results from the two periods.

Problem 3. Fitting an AR model Using the returns of Pepsi Cola you calculated in Problem 1 from DATA 1 and DATA 2 above please fit an AR model for each of the periods.

(a) Use the PACF function to plot the partial autocorrelations for each dataset. Based on the plot what is the recommended order of AR models.

(b) Using the ARIMA function and the order p you determined in previous problem fit an AR(p) model to the returns.

(c) Now use the built in function `ar` in R. Using criteria of your choice obtain two models as outputted by the `ar` function.

(d) Compare the coefficients for the models obtained between the two periods. Comment on the differences you observe.

Problem 4. Evaluating a model performance Please download stock data for PEP from Jan 1, 2022 to Jan 31, 2022. Please note that during holidays there is no trading data.

(a) Using the models you evaluated in problem 3 please forecast the return for: 3 days, 5 days and 10 days.

(b) Please calculate a 95% confidence interval for the forecasted values.

(c) Compare the forecasted values with the real values in the dataset.

(d) Please comment on your observations.