

Assignment #2

In this assignment, you will use the provided dataset to practice implementing some of the concepts we have discussed in class. You must submit your completed results and some brief commentary about your results and decisions via a Dropbox submission folder on the course D2L website. You must also submit your code (as separate files, not attached in the report) so that your results can be entirely recreated.

1. Explore the attached dataset and find additional information from the resources listed below.

- a) The dataset comprises NY Stock Exchange with several additional predictors, as explained in the following paper (especially sections 5 and 6):
<https://www.sciencedirect.com/science/article/abs/pii/S0957417419301915>
- b) The original dataset was obtained from:
<https://archive.ics.uci.edu/ml/datasets/CNNpred%3A+CNN-based+stock+market+prediction+using+a+diverse+set+of+variables>
- c) Although the original dataset includes 5 different files, we will only use the NYSE.csv file, which includes values from 2010 to 2017.

2. Begin by using your knowledge of ARIMA (or SARIMA) modeling to conduct a univariate time series analysis and prediction of the NY Stock Exchange. Explain your process, present your chosen model, examine the residuals, and evaluate its performance as follows:

- a) Hold out the last 3 months of 2017 for out-of-sample prediction. Plot the predictions and confidence intervals and report the forecasting error using appropriate metrics.
- b) Use a rolling window approach, with a training window of 3 years and daily increments, predict the next day. Again, plot the predictions and confidence intervals and report the forecasting error using appropriate metrics.

3. Now, explore whether you can leverage additional information in the file as exogenous variables. Use appropriate tools to evaluate the suitability of using these variables and summarize your results. There is no fixed number of variables assigned – use your judgement and justify your decisions. Using an ARIMAX framework, show the impact of including additional variables as part of your prediction. Repeat the analysis in 2 (a) and (b) using the updated models.

4. Finally, repeat the analysis in 3 but for a VAR-based approach. Explore the relationships between the chosen variables using appropriate tools. Again, there is no fixed number of variables assigned – use your judgement and justify your decisions. Using a VAR framework, show the impact of including additional variables in your prediction. Repeat the analysis in 2 (a) and (b) using the updated models.