Using Data to Make Investment Decisions

Final Project: Preliminary Submission

ST362-B Spring 2022

George Mandl

200655330

mand5330@mylaurier.ca

Data Set

The data set that I will be using for analysis is one that I have created myself using Microsoft Excel and Python in the Juypter Notebook environment. The Python Pandas library was heavily used due to its data analysis and manipulation capabilities. The data set is called "SMP500MasterData.csv" and it will be referred to as the master data set or "masterdata" in this report and sequential reports. To create the data set, I first created an excel sheet that tracks various attributes of SMP500 stocks using the Stocks Data Type in Excel. The spreadsheet is exported everyday after market close and is added to the Juypter notebook repository. In the repository there is also a file that contains a function that reads the spreadsheet and the master data set. It then appends the spreadsheet to the master data set and then reads the new master data set back onto its location in the repository. The only user interaction needed is to once daily export the excel spreadsheet, upload it to the repository, and run the function to append.

Note: The dataset that is submitted with this report is set to change since I am appending results to the master data set daily

Motivation

The data set is created for the purpose of exploring the quantitative side to investing. Quantitative trading, also knows as quant, is an area of interest of mine because it is an intersection of statistics, artificial intelligence, and finance. Academically, I am passionate in all three of these areas and I created the data set to explore their combination. Further interest within the field of quant that I have become interested in is the use of data to drive investment decisions through quantitative trading strategies.

Preliminary Outline

Through this project, I am hoping to explore various variables that can used as predictors to evaluate investment decisions. There are two potential valuations for making an investment decision. The first is the price of a stock and whether there is value in the price it is listed at. If the actual price is higher than the price that the model predicts then the stock is considered overpriced and a bad buy. And vice versa, if the actual price is lower than the price that the model predicts then the stock is considered underpriced and a good buy. The other valuation is finding predictors that result in positive or negative changes in price and the magnitudes of these changes. An optimal predictor, or combination of predictors, would result in high magnitude positive changes and low magnitude negative changes. We can find this by looking at the expected change. In short, the two variables we will look at as our dependent variables are "Price" and "% Change". I will attempt to create functions that will automate the creation and evaluation of various regression models to find the best quantitative trading strategy. I don't know exactly what questions will be asked yet, but they will be something along the lines of:

- Why of the variables is most significant in predicting "Price"?
- Is it statistically significant that added ____ variable to the model will improve the accuracy of the model?