BRYAN SHENG CONG TAM (bst35)
ORIE 5270 – Big Data Technologies
March 04, 2018

Project Proposal

Introduction

This project aims to determine the profitability of day-trading strategies using a portfolio of the most volatile stocks from the previous trading day. Such strategies try to take advantage of market overreaction to various events by suggesting trend reversal trades. A paper[1] by Bondt and Thaler demonstrates evidence of market overreaction. As such, these strategies have strong empirical underpinnings, and this project explores whether traders can consistently take advantage of patterns in the evidence.

Background

The NYSE Daily Trades File details the price and volume for every trade on every security in the US. This data allows us to determine the most volatile securities for every trading day and their intraday price trends, which we believe could give us indicators on trends for the next trading day. Such indicators require factors, such as the ratios of low to close prices and high to open prices, to be constructed from the underlying data.

Methodology

This project is expected to involve several steps, starting with processing the data. The raw data for every trading day is about 2.5GB, and significant effort will be required to parse the data and convert it into a readable format.

The necessary information, such as the most volatile stock for a trading day, then needs to be extracted from the data. Factors, such as the ratio of low to close price for each security, and the dependent variable, which is the return on a security for the next trading day, also have to be constructed from the data.

[1] Werner F. M. De Bondt, and Richard Thaler. "Does the Stock Market Overreact?" The Journal of Finance 40, no. 3 (1985): 793-805. doi:10.2307/2327804.

The information, factors and dependent variable are next fed into various machine learning algorithms for analysis. The types of machine learning algorithms used may vary depending on the dependent variable and predictive power of each algorithm. This project intends to start with simpler algorithms, such as linear regression if daily returns in its original continuous form is used, and logistic regression if threshold returns in its processed binary form is used. If the simpler algorithms are unsuccessful, the project intends to proceed with more complex algorithms such as decision and regression trees. The use of cutting-edge but computationally expensive algorithms involving deep learning and neural networks are also possible.

The results of each algorithm will then be analyzed to determine the predictive power of each algorithm. If the algorithms are at least somewhat successful, the project will then move on to construct optimized portfolios based on the algorithms' predicted returns for the target securities. The returns of these portfolios will then be analyzed. If the returns are satisfactory, the project will then propose a trading strategy utilizing the aforementioned algorithms.

Conclusion

We expect this project to contribute new insight on the profitability of trading off market overreactions, as well as demonstrate the use of machine learning in finance. While we cannot guarantee a successful strategy, we look forward to gaining a better understanding of what may or may not work in financial markets.