ORIE4741 Project Proposal

Rui Dai (rd576), Marshall Guan (jg2262), Bojun Li (bl755)

October 2020

General Description

One of the toughest challenges in investment management is to identify which aspects of a public company's fundamentals affect its stock price. Even for a seasoned stock analyst, distilling information from large swaths of operating performance data and news remains difficult. Furthermore, financial data embody the concepts of "big messy data" because of its notoriously noisy nature and high-dimensionality. We believe that by using computer science, statistics, and big data analytics, we could gain profound insights into the complex financial data environment, and identify correlations between operating and stock performances.

Question

1. How fundamentals of a publicly traded company affect its stock price?

In practice, modeling techniques such as Discounted Cash Flow or Ratio Analysis were widely applied to forecast and evaluate stock price based on companies' fundamental data. In this project, we would explore predictive power of machine learning techniques and compare their effectiveness against DCF and Ratio analysis.

2. Do fundamentals of public traded companies reflect the health of the overall economy?

Public companies are important drivers of many economic activities, including innovations, lending, trades, etc. Moreover stock prices also function as significant indicators of market sentiments. As our dataset includes both stock data and operating data, we would have the opportunity to study those relationships.

Dataset

Link: https://www.kaggle.com/dgawlik/nyse
Our dataset contains 77 explanatory variables which represent the fundamentals of public listed stocks in NYSE according to their SEC 10K annual fillings as well as the response variable which is the daily historical prices of the stocks. The time frame of this dataset is from 2012 to 2016. The dataset is downloaded from Kaggle and the original sources are: Yahoo Finance (stock prices), Nasdaq Financials (fundamentals) and extended by some fields from EDGAR SEC databases.

Feasibility Analysis

Empowered by a descent dataset, we skip the trouble of compiling and matching company operating performances. In the future, we could add alternative dataset, such as textual data of company filings, and twitter data commenting on company's stock performances to expand our research.