Project 2 Proposal – Predicting Investment Returns by Tracking "Insider" Trading and Twitter Activity

Team Members: Jack Galvin, Wilford Bradford, and Josh Fram

Repository Link: https://github.com/UC-Berkeley-I-School/Project2_Fram_Galvin

Project Summary: The SEC requires all executives at US listed public companies to disclose any trading activity conducted in the shares of companies for whom they work. This "Insider" buying and selling activity is updated frequently and publicly available for investors to download from the SEC's website. Some investors have developed strategies that track this data, driven by the notion that if a company's CEO is buying stock in their own company, it is an indication that they believe it is a good investment. This is ostensibly an important indicator given that the CEO is the person who should be most familiar with the company and its prospects. The reverse holds as well – if an executive is reporting sales of their company, it may be a bad signal about the prospects of that company.

We propose an exploratory data analysis to uncover whether there is a relationship between "Insider" activity and a company's stock price. As we analyze the data, we will almost certainly come up with additional questions based on our findings - these will be explicitly discussed and included as well.

Primary Dataset: The SEC's EDGAR Database maintains a running list of all "insider" buys and sells reported via Form 4 submissions. This will be the primary data set. Specifically, our sample set will be isolated to companies that saw Insider activity in the 4th quarter of 2018.

- ***https://www.gurufocus.com/insider/summary***

Secondary Dataset: We will source daily stock returns for a 1-year period following any insider activity that occurred in our sample period (4Q 2018). This data will be sourced from Bloomberg (a sample is provided from Yahoo Finance below for Ticker ***SLS***.

- ***https://finance.yahoo.com/quote/SLS/history?period1=1546300800&period2=1577750400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true***

Combining Datasets: Joining insider activity data with stock price data from Bloomberg will rely largely on stock tickers and dates. The relevant columns are as follows (we may find there are more, but at least these):

- SEC Edgar Database: The Direction (Buy / Sell), Amount of each trade, Ticker Symbol, and Date
- Stock Market Data (Bloomberg): Ticker Symbol, Date, Price (note that this will be daily data, starting from one day prior to the first instance of Insider activity and extending 1 year after the date of the first instance of Insider activity)

Report: We intend to cover whether there is any meaningful correlation between Insider trading and a stock's performance, both on absolute terms and relative to the S&P 500. Within our report, we intend to include several visualizations, including, but not limited to:

- Line plot (time vs stock price)
- Scatter plot (days since first insider activity vs price), including regression line
- Heatmap (to uncover any correlations among several columns of data)

- Pairplot (again, to uncover any correlations among several columns of data)
- Others (as we uncover relationships, engineer additional features, etc.)

Our report will take the form of a pdf word document and make use of the figures above. It will include:

- Abstract
- Introduction
- Data Collection / Processing
- Discussion, including assumptions, strengths, and limitations of our analysis