Financial Data and Stock Price Correlation

Research Question: How do various quantitative pieces of financial data affect and correlate with their company's current share price?

We are interested in the intersection of data science and finance and hope to analyze any possible correlations between financial data and closing stock price. We became interested in this topic after watching an interview with Jim Simons. He tracked the correlation between an uncountable amount of variables and stock prices. It is commonly said that they tracked this correlation between stock price and almost every possible variable that could be tracked over time.

Our Data: Our data consists of Compustat and Yahoo Finance exported csv files for our two companies, Apple and Google. We collected the quarterly data from Jan 2000 - Jan 2021 for all major collected financial data on Compustat, and the monthly close price data on Yahoo Finance.

We have over 600 variables for the 84 quarters in our time range, with an entry representing an equivalent value on the 10-Q or 10-K. An empty space represents 0 or N/A, as provided by Compustat. Relevant variables will be taken when needed using the following key: https://www.crsp.org/products/documentation/quarterly-data-industrial (https://www.crsp.org/products/documentation/quarterly-data-industrial) Stock price data was taken from one csv and imported into the df.

Data Analysis: We are working with dataframes using pandas to simplify how we worked with the data. Our functions either calculate a ratio and add it to our dataframe, or graph the ratio against the stock price. For each function, we calculated the correlation betweeen stock price and the associated ratio and reported this.

Data Interpretation: After analyzing the 8 apple and 7 google graphs, it is clear that certain metrics and relationships are hard to predict, even across companies in the same industry. For example, in the case of apple and google's current ratios, it is extremely unexpected for apple's to be very close to 1 in comparison to google's 3. Our graphs are extremely easy to follow by being labeled and titled. We also graphed the linear regression for the information. Our next steps include analyzing more companies in the tech industry and possibly branching off into other industries. We also hope to analyze the most 5 or 10 years instead of just the last 20 years, since the tech industry is constantly innovating.