An Exploration of the Effects of the United States Executive Branch under the Trump Administration on the Stock Market through Sentiment Analysis and Linear Regression

Authored by: Bryan Ma, Neil Keltcher, Teddy Pomianek

Problem Statement and Background

Our group sought to find a measurable effect of the many orders committed by the Trump administration, and settled on analyzing change in United States stock market (the market) evaluations. We chose this topic as all members were interested in seeing if the government actions under the Trump administration were considered positive or negative by the overall population, represented by opening, closing, and intermediate prices in the market. For every reported executive action we plan to measure and analyze the effect of the announcement on the market values, and draw conclusions from our gathered data.

This topic has been in heavy contention across the country, and we seek to supply a quantitative measurement of the impact of the Trump administration so far through fluctuations in market prices. It should be noted that while this is a concrete quantitative measurement of the administration's performance, it does not take into account other effects of the administration that may not impact the market, so the findings cannot be used to conclusively make claims on administration policies.

Introduction to the Data

Our project necessitated two main categories for our data: stock data and executive branch report data.

Our stock data came from the yfinance library for Python, which while not directly affiliated with Yahoo Finance, takes data from publicly available Yahoo APIs to display a variety of stock data for every ticker on the market. For each ticker, yfinance returns (within a specified date / date range): opening price, closing price, daily high, daily low, and volume traded. Given all stock data is publicly available and accessed through the Yahoo Finance library, we expect no biases to appear in our stock data. That being said, due to time constraints, our group opted to focus on VOO as our stock, given it contains the top 500 best performing stocks on the market, believing it would be a good indicator for market activity as a whole. However, because VOO is such a large and consolidated stock, any changes in the price through the days and weeks is quite small, leading to potential issues later on.

Our group considered a number of options for fetching reports on executive actions. We considered the White House website, CNN, MSNBC, Fox News, NewsData, Google News, and more. Ultimately, technical roadblocks with our web scraping left us with the White House website as our final choice (https://www.whitehouse.gov/news/). The White House's news reports include articles, briefings and statements, presidential actions, fact sheets, and remarks. We scraped all the categories to compile our sentiment analyses of presidential actions. One key issue that should be noted is that the White House news is managed by officials appointed by the executive branch, so there are implicit and explicit biases in the reports. Explicit biases would be the wording of the articles and how they address positive vs. negative topics for the executive branch. A quick skim of the site shows several titles that could be analogous to propaganda. An example of implicit biases could be whether the site reports or ignores negative media.

Data Science Approaches

Our problem statement was to examine the correlation between the stock and the sentiment of all the written statements produced by the White House government website. We were able to address our problem statement through the use of web scraping, Pandas dataframes, sentiment analysis, normalization, correlation of directional movement of price and volatility changes, calculation of volatility, and visualizations.

First we had to get our text data from the white house website. All the text was in a format of a main page with all the article links on that page. Then there were many more other main pages of past articles. To access these main pages we only needed to add to the end of the url, "/page/2", to find main page 2. To scrape all written content from every article we first had to scrape every main page for the article links. Then we took every link to every article and scraped the content and the data from each.

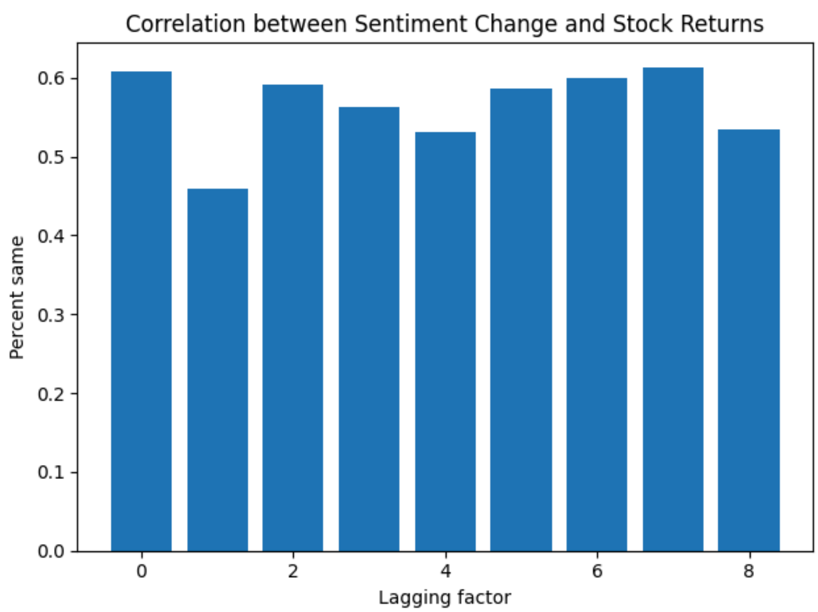
Then we used Pandas dataframes instead of lists or dictionaries because of its organized tabular format with rows and columns. By storing the stock data and sentiment scores in the dataframe, we were able to easily align values by date, handle missing data, and perform calculations much easier than if we used a dictionary or a list. For example the stock data did not have all the same dates as the white house data. This is because there was no stock data on the weekends or holidays because the markets were closed, but the White house still wrote articles on these days. But, with a simple inbuilt function I was able to erase the columns that had the missing data.

The next data science approach we used was to use sentiment analysis. We started out with just raw text from the white house website. This was just qualitative data and there is no way to perform calculations with the quantitative stock data. But using sentiment analysis it turns the text data into a polarity score, a number from -1 to 1. So, for each day in the dataframe, we calculated the sentiment score. If a day had multiple text write on that day we would just take the average sentiment score from that day.

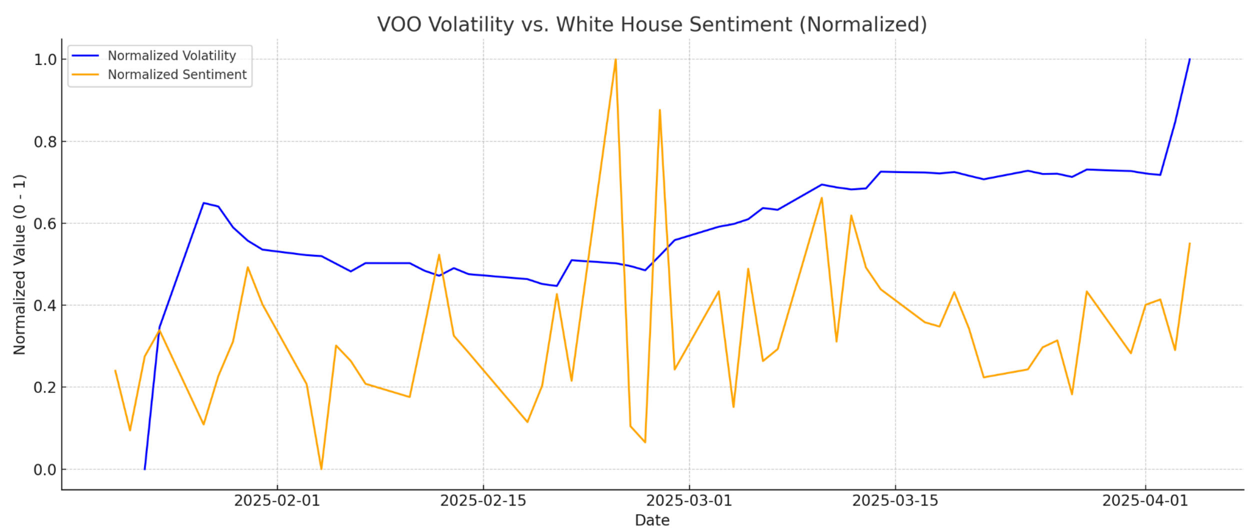
Now that we have two sets of quantitative data to perform calculations and plots we need to normalize it. The reason is that the sentiment data ranges from -1 to 1, while the price of the stock ranges from 500 - 550. If we tried to make a line graph with this we would get a nice line for the stock price, but the sentiment would look like a flat line because its range is only 2. So we have to normalize the two sets of data from a range of 0 to 1 to be able to easily see the correlation in a line plot (Figure 1, in results and conclusions). When we normalized the sentiment data we added 1 then divided by 2, so 0.5 would be like the zero when not normalized.

The next part was finding if there was a correlation between the sentiment and the stock price. We would not be looking for a linear correlation, but we would be looking for if the stock and the sentiment increased or decreased at the same time. The first thing we needed was the daily returns. To find this, we took the current day's closing price and subtracted that by the previous day's closing price. After this we normalized daily returns of the sentiment and the stocks(Figure 2, in results and conclusions). If it was a negative number we assigned the value a 0, if it was a positive number we assigned the value of 1. After this we formed a list of ones and zeros for both the daily return of sentiment and stock prices. Then, to find the correlation of directional movement, we counted the number of times the two lists had matching indexes, then dividing the count by the length of the list to get a number between 0 and 1.

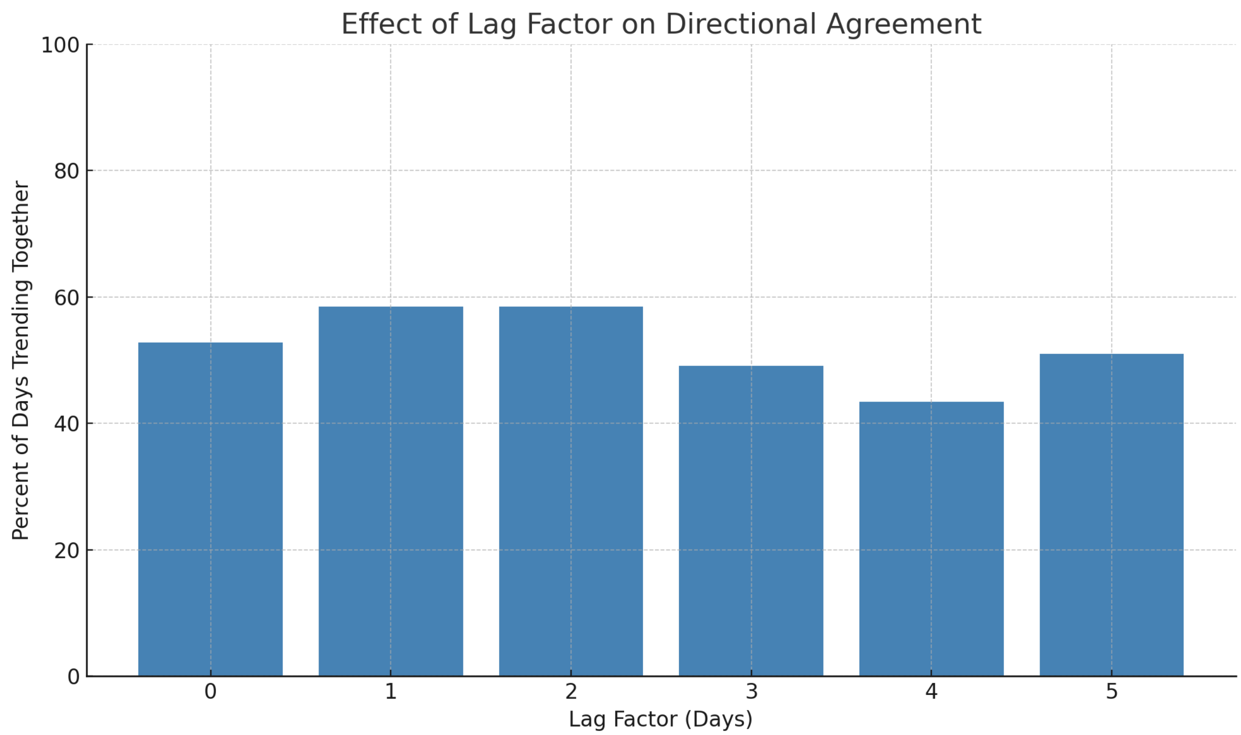
But, if we only found the correlation on the same day we would not be getting the full picture. We needed to see if the correlation was lagging. It would make more sense because it would take time to write an article. Most articles would be talking about events that happened yesterday or days prior, or to take into account the delay of the market reacting. We decided to add a variable lagging factor to our daily change in sentiment time series. This lagging factor shifted sentiment change forward in time. For example, with a lagging factor of 1, the sentiment change from April 11th would be compared to the volatility change or price change from April 10th. A lagging factor of 2 would move our sentiment time series by 2 days and so on. We decided to find the correlation of directional movement with a lagging factor of 8 days and showed the correlation of each day through a bar graph.



Next we wanted to look at volatility graphs to draw conclusions of whether the sentiment of White House reporting can be used to predict market conditions. In order to calculate the volatility of VOO, we created a time series of the percent change in the daily closing prices over the time period January 21 2025 to April 6th 2025 (Trumps 2025 presidential term up to April 6th). To calculate volatility from the closing price data, we needed to find the daily returns. To find this, we took the current day's closing price and subtracted that by the previous day's closing price. Then we divided that difference by the previous day's closing price to find a percent change. After creating a time series of percent changes of stock price, we used that time series to calculate the standard deviation using the sample standard deviation formula. Finally, we needed to annualize our volatility and to do that we multiplied our standard deviation by square root of the total stock trading days in a year: $\sqrt{252}$.

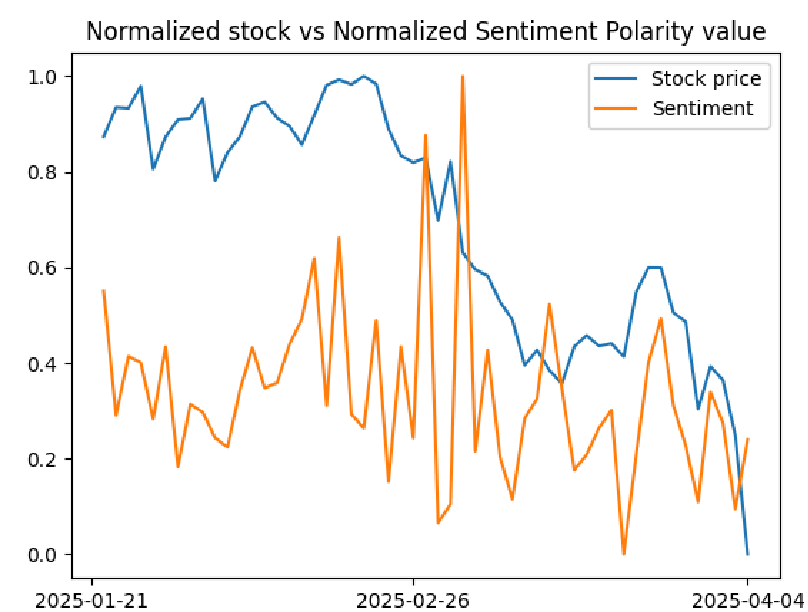


In addition to our volatility of VOO vs sentiment graph, we also created a daily-change of volatility and sentiment graph with a lagging factor. The bar chart below represents the distribution of the percent of days in which volatility and sentiment trend together across various lagging factors.

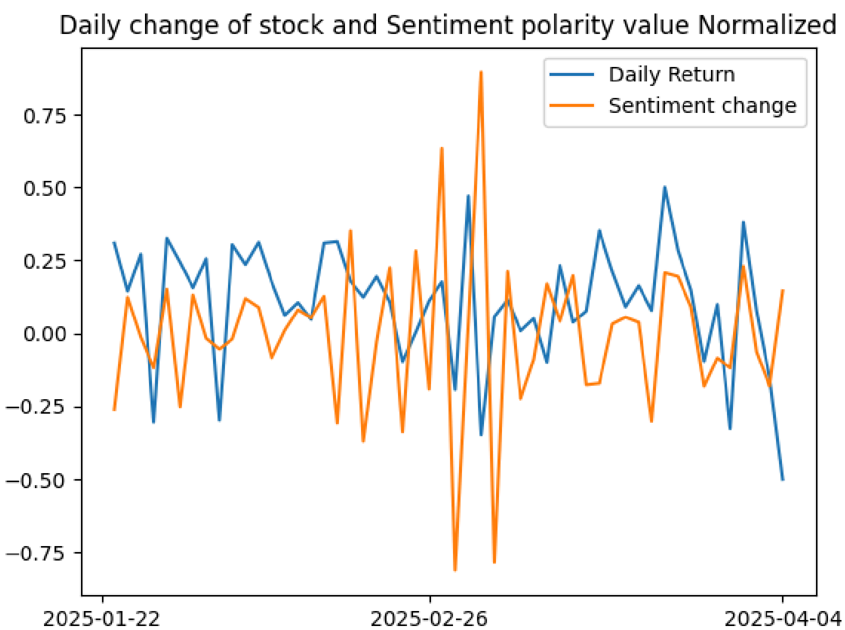


Results and Conclusions

We began this project searching for any correlation between stock market prices and executive branch actions and statements by comparing the sentiment analysis value of said executive statements with fluctuations within the market. We are happy to conclude that we have found seemingly conclusive evidence that there is indeed a positive correlation between the sentiment of the statement and the movement of the market.



The line graph to the right (Figure 1) contains the normalized stock price of VOO plotted against the normalized sentiment of executive reports. As you can see, there are very visible peaks and valleys at the same time across both lines, meaning there must be at least some correlation between the two values. Although it is acknowledged that there can be actions taken by a third party that would similarly affect both the stock market and the sentiment of executive statements,



this is at the very least a small indication that our hypothesis was correct. Another indication of our conclusion is shown in the graph to the left (Figure 2). This line graph shows the normalized sentiment polarity value for the days the Trump administration has been in office, and compares it to the daily change in VOO, which is also normalized. In this graph it is once again clearly visible that the change in polarity of the executive branch's statements mirrors the daily change in VOO almost to the dot. Most peaks and valleys in the polarity are echoed in the market, showing a positive correlation between the two elements and cementing our belief in our hypothesis. One last thing that should be noted when looking at the graphs is that the market typically lags a few days behind presidential reports, so the peaks and valleys seen in the sentiment line will be a tiny bit before the ones in the price line, but they are still evidence of a direct correlation.

Future Work

There are a variety of ways this project can be expanded beyond the scope of our work over these last few weeks. The majority of these we considered when planning our project, yet did not implement due to time constraints or some other issue we encountered. A few of these ideas include: expanding the stock selection to multiple index funds, changing the stock portfolio to include individual stocks from every sector of the market, scraping articles from multiple news sources, and possibly comparing the sentiment of Democratic/Republican-backed news outlets. These are just a few of the other ideas we considered throughout the duration of the project, and all places we can take the project in the future.