**Donald Trump’s Tweets Influence on Financial Market - Project Deliverable 2**

Sam Cutter, Xinrui Dong, Yixing Huang, Jia Yang

April 27, 2019

1. **Introduction**

Social media is becoming one of the main sources of unstructured communication data available today. Many financial analysts and regulators alike use Twitter to gather data and obtain meaningful insights. Many countries’ political leaders are also using social media as one of the key platforms to deliver messages to the public. The current (45th) President of the U.S.A., Donald J. Trump is no exception. His personal twitter handle (@realDonaldTrump) has posted more than 41,300 tweets, including over 6,530 tweets since his inauguration January 19th, 2017, and has nearly 60 million followers as of April 2019. High-profile politicians like President Trump have been using Twitter to communicate directly with the public to float his agenda and influence.

The President of the United States has political influences and executive powers; and it is often assumed that the U.S. President has access to information provided by different governmental institutions, financial agencies, and advisors. Therefore, in theory, the information shared through the President’s Tweets can be are considered to be informative signals that may influence consumers and affect investors’ decisions in the stock market.

1. **Project Goal**

The goal of this project is to examine whether President Trump’s tweets is a source of influence for the stocks market in the United States. The effect of tweets can be measured by coinciding changes in the stock market or changes in the share prices of certain companies targeted in the President’s Tweets [1]. Of the roughly 34,000 tweets used in this research between include were examined using keywords sentiment analysis to find which the overall tweet sentiment. Once determined, we selected the tweets representing the strongest positive and negative emphasis to use as our designed intervention point. We will also use the Causal Impact Timeseries Algorithm developed by Google to validate the pre-conceived relationship between Trump’s twitter attacks/praises directed at a company and its resulting market performance in the stocks.

1. **Analytical Techniques**
   1. **Simple Time Series Analysis**

At first, we are curious about whether financial market obeys time series, so we pick two parameters: S&P500 index and daily trading volume between January 2001 to December 2016. We set a training model using data from January 2001 to January 2016 and test on February 2016 to December 2016. Various time series prediction models are created, including the ets model, naive model, seasonal naive model, and Arima model. Then we use prediction models to predict the next 42 months financial market performance and compared with real market data.

* 1. **Sentiment Analysis**

In order to determine tweets that we should use for our causal impact study, we use sentiment analysis to find the most forcibleness tweets. Widely used lexicons such as affin, bing, and nrc, have been found to be very useful to a wide range of problems that are of interest to human-computer interaction practitioners and researchers. However, the inherent nature of microblog content - such as those observed on Twitter and Facebook - poses serious challenges to practical applications of sentiment analysis. Some of these challenges stem from the sheer rate and volume of user-generated social content, combined with the contextual sparseness resulting from shortness of the text and a tendency to use abbreviated language conventions to express sentiments.

We overcame these flaws by using an algorithm called Valence Aware Dictionary and eSentiment Reasoner (VADER). VADER is a reliable method for calculating the polarity of tweets as it is used frequently to analyze social media [4].

Afterwards, we decide to conduct the sentimental analysis on all the tweet and transform the sentimental result as a numerical value, then correlates the sentiment scores to the stock market. The results will show whether Trump’s tweets have significant impact on the entire U.S. stock market.

* 1. **Causal Impact**

Trump has tweeted a lot regarding specific companies, including but not limited to technology, media, retailers, aviation, and others. It is interesting to see whether his tweets impacted the stock prices of the companies he mentioned; and if they do, how much do they influence the stock. After research, it is determined that the Casual Impact R package is applicable for this kind of analysis.

Casual Impact is a package developed by Google [3], and it compares a response times series with a control times series, with a given event or intervention. In other words, it predicts the response time series if the event had never occurred, and compares with the actual time series. By using this characteristic, it is possible to predict if Trump’s tweet has influenced the stock price of the corporation.

There is an important assumption for this method: the control time series itself is not intervened by the event [2]. However, it is very difficult to find the control time series that is not intervened by any event. The stock price is extremely sensitive, and it will be affected by many different factors. Furthermore, industry indices are also affected by the event since the company’s stock prices contribute to the indices. It is also noticeable that the stock prices of the same company may act similarly in different markets. Therefore, the only viable choice is to treat a competitor's historical stock prices, in the same market, as the control time series, with the assumption that Trump’s tweet only influences the specific companies he targeted on.

If the casual impact is run successfully, the package will generate a summary, which includes the forecast if the event did not happen, and show the difference between the forecast and actual. The summary will also indicate whether it is statistically significant. Several plots will also be generated.

**4. Results and Discussion**

**4.1. Simple Time Series Analysis**

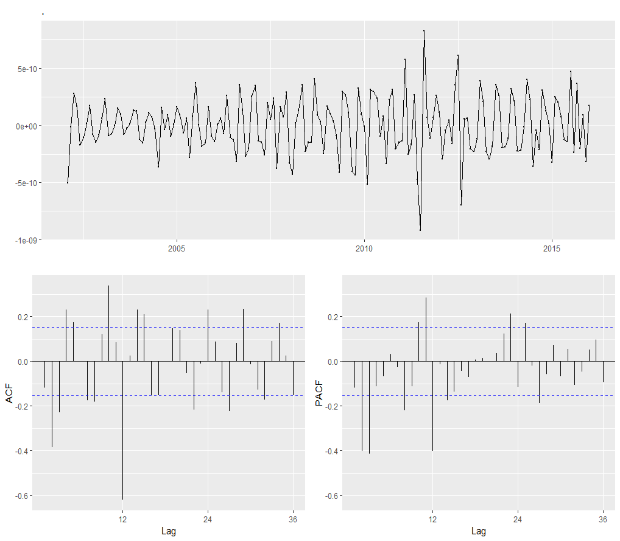
The raw data for the analysis is the S&P500 index and daily trading volume between January 2001 to December 2016, extracted from Yahoo Finance. After we compared the real market data and our prediction models, we found there is a very weak correlation and thus we would like to draw a conclusion that the financial market does not necessarily obey time series. In general, we found the seasonal naive model has the best accuracy but the white noise is still very small. Figure 1 and 2 show the accuracy of the seasonal naive prediction model and ets model, no obvious white noise is observed. It means that our models are not accurate in predicting the stock prices. After this, we decided to pursue another way for conducting analysis, and finally, we find sentiment analysis and causal impact would make a better interpretation.  


Figure 1: Summary of Seasonal Model

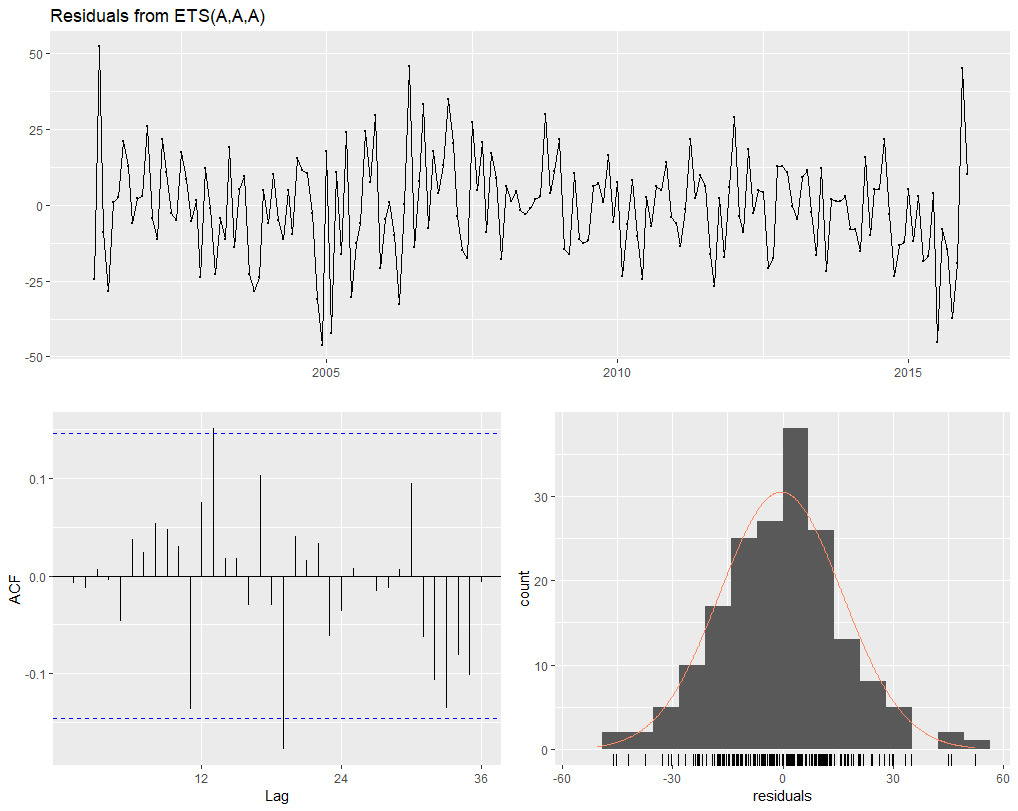


Figure 2: Summary of ETS Model

**4.2. Sentiment Analysis**

Before conducting sentiment analysis, additional data cleaning is done to separate and extract useful text data, characters and time, based on the dataset obtained in Deliverable 1. However, the Twitter records the time in UTC time zone, which does not correspond to the EST time zone for New York Stock Exchanges. In order to better examine the impact, we converted the time to fit EST time zone, and additional conversion is done to separate out time, date, day of the week, and stock market impact date if tweet occurred after trading has ended.

After examination, it is realized that VADER has a dictionary of social media vocabulary and matches words in tweets with this dictionary. It assigns a compound value to tweets and categorizes tweets as positive, negative and neutral. The sentiment score of each tweet is compounded and standardized between 1 and -1 with 1 being extremely positive, -1 being extremely negative and 0 being neutral. Here is an example of VADER analysis on one Donald Trump tweet:

@realDonaldTrump: “Just talked with Pfizer CEO and @SecAzar on our drug pricing blueprint. Pfizer is rolling back price hikes so American patients don’t pay more. We applaud Pfizer for this decision and hope other companies do the same. Great news for the American people!”

The VADER analysis assigns this tweet a compound score of 0.8799, which is high. As can be seen, VADER does well in accurately identifying the sentiment in the above tweet and classing it as a very positive tweet. Trump’s tweets are classified into positive and negative sentiments in this manner.

The next step we did was to calculate the correlation of the tweets between S&P500 to see if there is any correlation between them because we would like to see whether Trump’s tweets have significant impacts on the entire stock market.

First, we conducted the sentimental analysis on all the tweet and transform the sentimental result as a numerical value. Multiple indies, such as positive index, which means the percentage of positive words out of all sentimental words, are created. We pick positive index, negative index, fear index, affin index, VADER index as our main index. Second, we created a keyword list, which is used to search all the related tweets from all tweets. For example, the trade war keyword list contains words like China, tariff, and trade. Then we selected all the tweets that contain the keywords and make them in one data frame. Finally, the correlation between the sentimental index of these tweets and S&P500 can be calculated.

Finally, we found that most of the sentimental index has a weak correlation with the S&P500 index. However, if the time window is short, the correlation would be stronger than the longer time window correlation. Below is the result we get from trade war tweets analysis (Figure 3). There is a 0.49 correlation between the fear index and S&P500 when the keyword list is Trade War, with the 6-month time window.

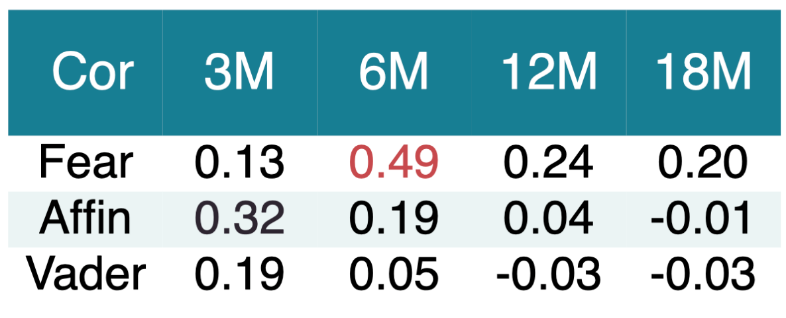


Figure 3: The list of correlation between fear, affin and VADER indices with S&P500

Moreover, we found that the distribution of sentimental index along the date is really random.

Below is the plot graph of trade war tweets’ VADER index in the past 2 years (Figure 4). No clear pattern is observed. It means that Trump does not reveal dominant sentiments when tweeting about the trade war.



Figure 4: The scatter plot of date of Trade War related tweets vs. VADER index

**4.3. Causal Impact:**

The stock prices of each individual company are obtained from Yahoo Finance by the get.hist.quote function of the tseries R package. The time range of the historical stock price is from 2015 to 2018. The closing prices of the trading days are selected as the points of the time series. Due to the and the number of tweets regarding the corporations, the Causal Impact can only be done for several companies. One big challenge for using causal impact is to create a semi-auto process of creating a data frame that contains the timestamp and corporation name for each tweet. It is important for quickly determining the actual day of the event. The R code is constructed to allow prompting such data frame. In addition, the Causal Impact can only be done on consecutive time series, meanwhile, there is no gap between each time point. However, since there are non-trading days for the stock markets, it is necessary to make the non-trading days have the same stock price as of the last trading days. It is also reflected in the R code.

Then the impact date can be assessed. Since there is no individual analysis on every single company in the industry, the companies to assess are chosen objectively. Trump has expressed his ideas on “America First” frequently on twitter, and one of the most targeted industry is the automobile. Based on this impression, several American based car industry companies are examined. The results show that Chrysler Fiat, Ford Motors and General Motors are mentioned at least 6 times each from 2017 to 2019, and the sentiments of these tweets are significant. Particularly, Trump has tweeted about how Chrysler Fiat moving the manufacture plants back to the United States several times. Then all of the tweets regarding the event are tested. Figure 5 and 6 show the causal impact, by comparing Chrysler Fiat, and General Motors for two tweets posted on January 11th, 2018. The examination window is set to 3 months to show the long-term impact on the stock price. correspondingly, the pre-period to assess is also set to 3 months.

The summary shows that for a 95% confidence interval the event has impacted the stock price from 23 to 27%. (Figure 5, 6). The P-value is only 0.001, meanwhile, it is statistically significant.

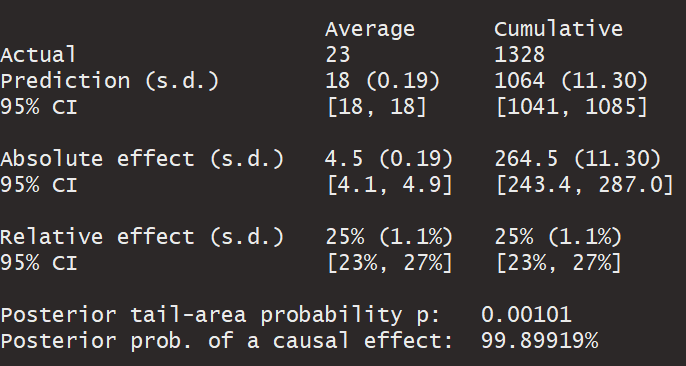


Figure 5: Summary for Causal Impact of Chrysler Fiat

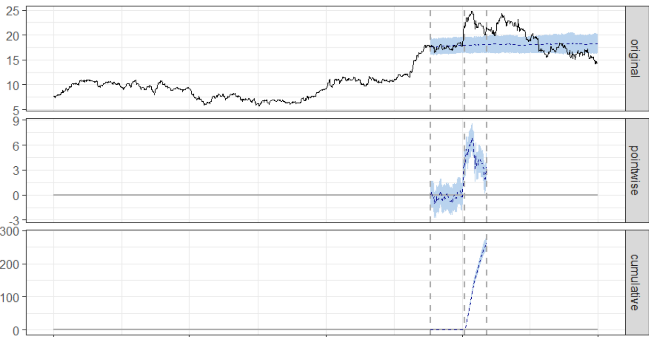


Figure 6: Graph for Causal Impact of Chrysler Fiat

On the contrary, negative sentiments might negatively impact the stock price. On January 5th, 2017, Trump tweeted: “Toyota Motor said will build a new plant in Baja, Mexico, to build Corolla cars for U.S. NO WAY! Build plant in U.S. or pay big border tax.” The Causal Impact analysis shows that Toyota’s stock fell from 5.7% to 7.7% in a 3-month window when comparing to Honda (Figure 7, 8).

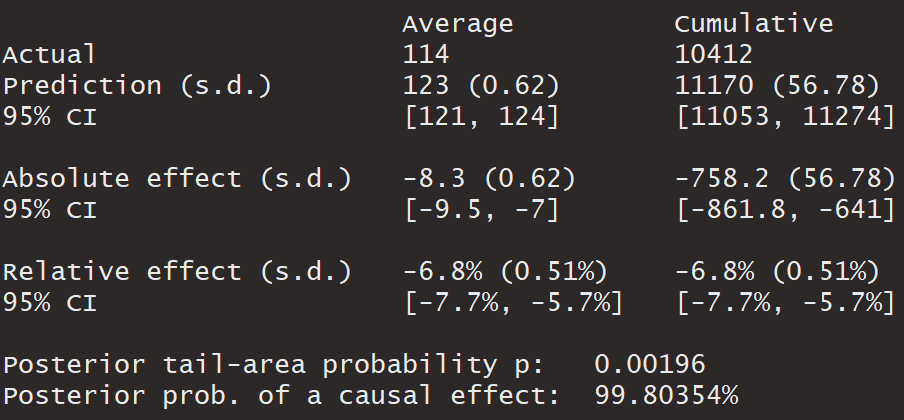


Figure 7: Summary for Causal Impact of Toyota

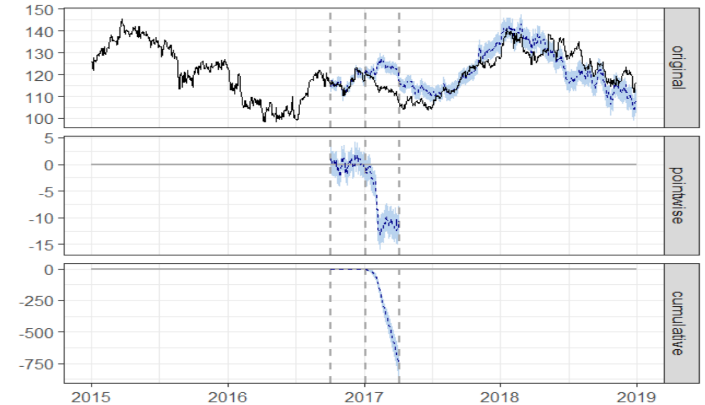


Figure 8: Graph for Causal Impact of Toyota

Another example can be found in the pharmaceutical industry. Pfizer and Merck are both American based drug companies and with comparable sizes. On July 9th, 2018, Trump has tweeted that “Pfizer & others should be ashamed that they have raised drug prices for no reason...”, while on the next day, Trump tweeted:” Pfizer is rolling back price hikes...We applaud Pfizer for this decision”.  The Causal Impact shows Pfizer experiences price growth from 5.4% to 6.9% for a 3-month window (Figure 9, 10).

The detailed assessment of the tweets for Chrysler, Toyota, and Pfizer, in different time windows, is summarized in table 1. The stock prices from all companies are obtained from the New York Stock Exchanges market.

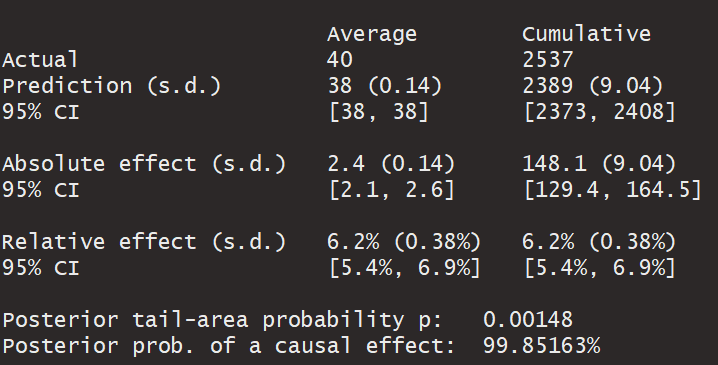


Figure 9: Summary for Causal Impact of Pfizer



Figure 10: Graph for Causal Impact of Pfizer

Table 1: List of Causal Impact for three companies

|  |  |  |  |
| --- | --- | --- | --- |
| Time Window to access impact | Chrysler vs. General Motors impact percentage (95% CI) | Toyota vs. Honda impact percentage (95% CI) | Pfizer vs. Merck  impact percentage (95% CI) |
| 1 week | [5.8%, 10%] | [-1.9%, -3.8%] | [-0.79%, 2.1%] |
| 1 month | [16%, 20%] | [-2.5%, -3.3%] | [0.19%, 2.5%] |
| 2 months | [22%, 28%] | [-5.9%, -7.3%] | [1.7%, 3.7%] |
| 3 months | [23%, 27%] | [-5.7%, -7.7%] | [5.4%, 6.9%] |

In general, Causal Impact does show Trump’s tweet has impacted on the automobile industry and pharmaceutical industry. It is also important to note that Trump tweeted and spread the information actually before the media and the company release. As a result, it adds to the credibility of the impact of the tweets. However, the Causal Impact method does not detect other significant findings for other companies. It is partly due to the fact that most of the tweets are regarding the politics, “fake news” and trade wars, but also partly that the more influential tweets are highly related to his own policies such as America First, and low drug prices. While it is may not as effective for others. In addition, Causal Impact might overestimate the impact of the tweets. Since we choose the date as the event, it will always be other events happening on the same particular date. Therefore, it is inevitable that the value relative effect of Causal Impact is influenced by these alongside events. It will be better to access the impact by smaller time unit, such as hours and minutes, but it will be much more difficult to do so.

1. **Conclusion**

As Twitter rises in popularity, the effect of celebrities and other politicians on financial markets can be analyzed. While @RealDonaldTrump has more twitter followers than the Twitter itself, he still does not break the top 10 list of most followed individuals (at the time of writing this paper @realDonaldTrump was the 12th ranked individual with 59,840,981 followers).

The sentiment analysis shows that sentiment of Trump’s tweets has a weak correlation with the S&P500 index, but our analysis also shows that the causal impact a single individual can have on organizations for the better or the worse. It is noticeable that Trump’s tweets impact the stock price of Chrysler, Toyota, and Pfizer.

Based on the analysis, it is recommended that regulators should scrutinize the effect of Twitter on financial markets more critically in the future.

1. **References**

[1] Ahmad H. Juma’h & Yazan Alnsour, 2018. "Using Social Media Analytics: The Effect of President Trump’s Tweets On Companies’ Performance," Journal of Accounting and Management Information Systems, Faculty of Accounting and Management Information Systems, The Bucharest University of Economic Studies, vol. 17(1), pages 100-121, March.

[2]

Google. (n.d.). CausalImpact. Retrieved from github: http://google.github.io/CausalImpact/CausalImpact.html#creating-an-example-dataset

Nishida, K. (2017, June 7).

[3] An Introduction to Causal Impact Analysis. Retrieved from Medium: https://blog.exploratory.io/an-introduction-to-causal-impact-analysis-a57bce54078e

[4] Rayarel, K. (2018). The Impact of Donald Trump’s Tweets on Financial Markets. Nottingham.