Using Social Media Data as Early Warning Signals in Risk Management

Yingyue Luan
MA Capstone Project Group A
University of California, Berkeley

May 9 2018

1 Problem Description

The topic of this project is to use social media data as early warning signals in risk management. We are motivated by an incident in February 2018 that Snapchat lost $1.3 billion after celebrity Kylie Jenner's tweet[1]. We are interested to find out that whether social media data sentiment analysis can be used to predict stock trends and if mentions of specific companies will affect stock values. As the percentage of U.S. population who currently use any social media has increased from 24% in 2008 to 81% in 2017[2], social media is inevitably a major source affecting our judegement and decision-making. This project will contribute to the investigation of the significance and utilization of social media data in business and finance industries.

Previous work by Bollen, Mao and Zeng[3] in 2011 analyzed tweets to measure public moods in six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) and they found a 87.6% accuracy in predicting daily up or down changes in Dow Jones Industrial Average (DJIA) closing values. Nguyen, Shirai and Velcin[4] further integrated the sentiments in social media into the prediction of stock price movements and the result shows a 2.07% better accuracy than prediction only using historical prices.

In this project, instead of analyzing public mood in general, we want to focus on two specific recent events and see if unexpected change in sentiments solely intervenes with stock prices. The first event is about United Airlines where puppy died on one of their aircrafts after being locked in the overhead bin for the whole flight[5]. The second incident we are following is about President Donald Trump's Twitter rant on Amazon started on March 29th, 2018[6]. The data we extracted are tweets directed to the company's Twitter page (i.e. @united for United Airlines and @amazon for Amazon) and stock prices.

We intend to consider the unexpected change of sentiment score as intervention point and measure its effect by time series analysis such as AutoRegressive Integrated Moving Average (ARIMA) model, Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model and especially, intervention analysis.

2 Data Description

The Twitter data are collected through the twitteR package in R after creating an API (Application Programming Interface) using one group member (Luna)'s Twitter account[7]. This API allows us to pull tweets from the past six to nine days. For example, there are over 200,000 tweets directed at United Airlines (@united) from March 7, 2018 to March 16, 2018 and over 300,000 directed at Amazon (@amazon) from March 24, 2018 to April 12, 2018. For every tweet, the "searchTwitter" command returns its text, username, ID, created time and other status variables. We intend to conduct sentiment analysis on the text of every tweet.

Daily stock closing prices from Yahoo Finance are gathered via "quantmod" package[8]. Besides stock prices for United Airlines and Amazon, market indices from New York Stock Exchange (NYSE), S&P 500 and NASDAQ are also collected for volatility comparison.

The intra-day trading prices data are accessed via Wharton Database Research Services provided by the Wharton School at the University of Pennsylvania. We are granted the access from Haas School of Business at the University of California, Berkeley. Only transactions during stock market hours (9:30 am to 4:00 pm) are collected. For United Airlines (ticker symbol: UAL), 1,444,514 transactions are obtained for a two-month period from January 16, 2018 to March 16, 2018. 8,528,921 transactions from January 1 to April 18 are collected for Amazon (ticker symbol: AMZN). The data include variables of time down to millisecond, trading price and trading volume. Figure 1 shows the differences of trade prices for United Airlines and Amazon.

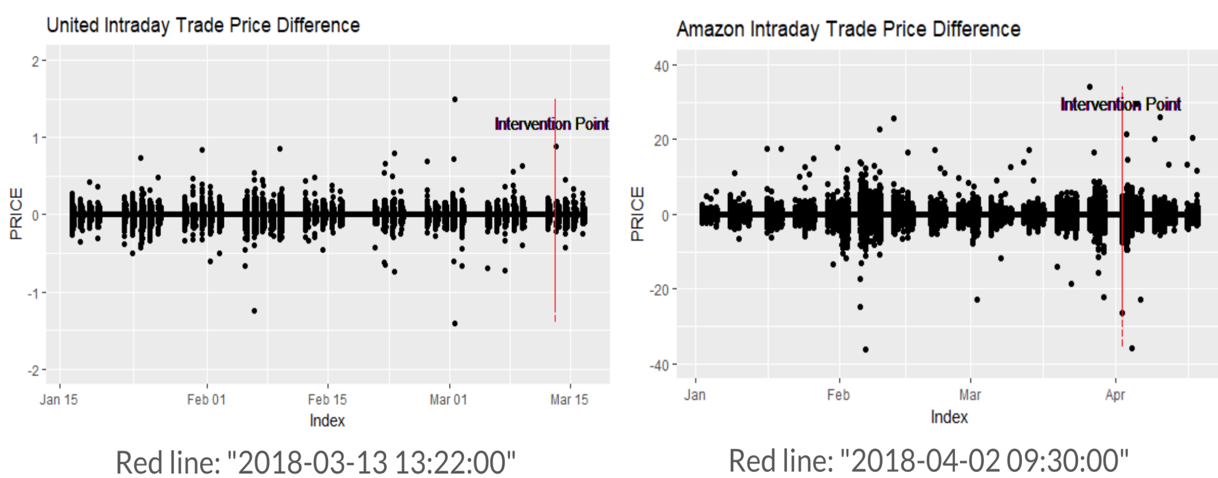

Figure 1: Trade Price Differences for United Airlines and Amazon

3 Methods

3.1 Sentiment Analysis

3.1.1 Rudimentary Sentiment Analysis

To process Twitter data, we conduct sentiment analysis which is a computational task that automatically assigns a score to the feeling users expressed through texts. It is an important topic in natural language processing (NLP), computational linguistics and text mining. We first implement a rudimentary approach using the "Opinion lexicon English" text file consisting a list of approximately 6,000 positive and negative words. We then count the number of negative, neutral and positive words of each text. This method allows us to add more colloquial words to the lists of positive and negative words. This feature is very helpful to our study since we aim to analyze social media texts and we would like to include more casual uses of language like "lit" and abbreviation like "nvm (never mind)". But this approach has a major drawback that it does not take account of the nuances and contexts of words such as sarcasm, negations, jokes, and exaggerations.

3.1.2 sentimentr

We further explore the "sentimentr" package which takes into consideration valence shifters (negator, intensifier, downtoner, and conjunctions)[9]. Besides tagging the words with polarity score, this augmented dictionary method also treats the four words before and two words after the target word as valence shifters and each shifter assign a weight to the score of the target word.

3.1.3 SentimentAnalysis with Loughran-McDonald Financial Dictionary

Another package we implement is called "SentimentAnalysis"[10]. This package utilizes various existing dictionaries, such as Harvard IV, and finance-specific dictionaries, such as Loughran-McDonald Financial Dictionary (approximately 3000 words). According to Prollochs, Feuerriegel and Neumann[11], manually-selected word lists such as the Harvard-IV psychological dictionary does not fully understand the connotations and contexts of financial news. They employed different Bayesian approaches to generate domain-specific dictionaries for sentiment analysis of financial news. Comparing to existing dictionaries, ridge regression achieves an improvement in correlation between sentiment value and abnormal stock market return of 93.25%. Therefore, it is more suitable to use a domain-specific dictionary (Loughran-McDonald Financial Dictionary) for our analysis since it relates the most to the content of financial study and it generates more accurate results than "sentimentr".

3.2 Market Comparison

To make sure certain event or intervention is unique to the interested company itself instead of a market-wise influence, we compare the return, log difference and three-day rolling standard deviation of target company with those of the market. The calculation of return can be written as

$$\text{return}_{t,0} = \frac{\text{price}_t}{\text{price}_0}$$

Log difference can be interpreted as the percentage change in a stock but does not depend on the denominator of a fraction.

$$\text{change}_t = \log(\text{price}_t) - \log(\text{price}_{t-1})$$

The three-day rolling standard deviation is calculated using "rollapply" command in "zoo" package[12]. The formula of rolling standard deviation is

$$\sigma = \frac{\sqrt{N \sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2}}{N}$$

where $N = 3$ in our case and x_i refers to stock price.

3.3 GARCH

Volatility is an universally important attribute of stocks and a good way to check if stock changes are consistent with sentiment movements. To measure how volatility spikes upwards and then decays away, we use the estimation of a generalized autoregressive conditional heteroskedasticity (GARCH) model. We will also use the GARCH volatilities of stock to check its consistency with sentiment variances[13].

A standard GARCH (sGARCH) Model is

$$\sigma_t^2 = \omega + \sum_{i=1}^{m} \zeta_i v_{it} + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

where $a_t = \sigma_t \epsilon_t$, ω is a constant and v_{it} denotes exogenous variables.

To fit a GARCH model, we implement "rugarch" package by specifying a variance model and a mean model. We use a standard GARCH with order (1, 1) for variance model and a ARMA model with order (0, 0) for mean model. Both orders refer to (p, q) parameters where p is the order of the autoregressive model and q is the order of the moving-average model. The choices of p and q are both commonly used in financial studies. One way to check if a GARCH model is appropriate is to look at its persistence. Persistence of a GARCH model represents how fast large volatilities decay after a shock. We check persistence by examining the two main parameters, α and β. If the sum of two parameters are greater than 1, the volatilities are too explosive to be trusted and thus the GARCH model is not a good one[14].

3.4 Intervention Analysis

We explore the trade data using time series analysis. Commonly, time series is a sequence of data points at equal-difference times. Because our data is only from open market hours, we use the "highfrequency" R package to aggregate trading prices into 30-second intervals by the median and fill in empty intervals with the value of the closest interval[15]. Thus, the trading prices are transformed into a continuous sequence. The time series is then checked for stationarity by Augmented Dickey-Fuller Test[16]. If the original time series is not stationary, a common approach to reach stationarity is by differencing: computing the differences between consecutive observations.

We use Autoregressive Integrated Moving Average (ARIMA) models to analyze time series. ARIMA model involves three parameters: p is the order of the autoregressive model, d is the degree of differencing, and q is the order of the moving-average model. Parameters p and q are determined by inspecting the autocorrelation function (ACF) and partial autocorrelation function (PACF). The parameters can be also determined by "eacf" function in "TSA" R package[17].

The intervention analysis determines if an intervention point affects the value of the series and estimates how much it has changed the series. In time series, intervention analysis focuses on the mean level change after an intervention. It uses the data before the intervention point to generate the ARIMA model and then forecasts the values after the intervention. The difference between the actual values and the forecasted values is then applied to determine the intervention effect and overall model[18].

The usual ARIMA model is

$$x_t - \mu = \frac{\Theta(B)}{\Phi(B)}\omega_t$$

where x_t is the observed series, μ is the mean, w_t is the error series, $\Theta(B)$ is the usual MA polynomial and $\Phi(B)$ is the usual AR polynomial. We then introduce z_t represents the amount of change at time t due to the intervention ($z_t = 0$ before intervention and may or may not be zero after). The overall model becomes[18]

$$x_t - \mu = z_t + \frac{\Theta(B)}{\Phi(B)}\omega_t$$

There are four common patterns of z_t. The first is a constant permanent change that z_t equals to a constant after the intervention.

$$z_t = \delta_0 \text{ for } t \geq T$$
$$z_t = 0 \text{ for } t < T$$

The second pattern is a temporary constant change that z_t equals to a constant only for a period after the intervention.

$$z_t = \delta_0 \text{ for } T \leq t \leq T + d$$
$$z_t = 0 \text{ for all other } t$$

Third pattern is a gradually increasing intervention that eventually levels off, which can be written as

$$\text{when } t < T, z_t = 0$$
$$\text{when } t \geq T, z_t = \omega_1 z_{t-1} + \delta_0 = \frac{\delta_0(1 - \omega_1^{t-T+1})}{1 - \omega_1}$$

The forth pattern is an immediate change following a trend back to zero.

$$\text{when } t < T, z_t = 0$$

$$\text{when } t = T, z_t = \delta_0$$

$$\text{when } t > T, z_t = \omega_1 z_{t-1}$$

In all equations, δ_0 is a constant change, $|\omega_1| < 1$ and both δ_0 and ω_1 can be estimated by the data.

4 Results

4.1 United Airlines

Our sentiment analysis of Twitter data directed at United Airlines from March 7, 2018 to March 16, 2018 found a point of sudden increase in the number of tweets and sudden decrease in the sum of sentiment scores. We denoted this point (2018-03-13 13:22:00) which represents the start of sharp changes in both tweet volume and sentiment score as intervention point. At this point, the total sentiment score dropped from -0.88 to -6 and the one-day rolling variance doubled.

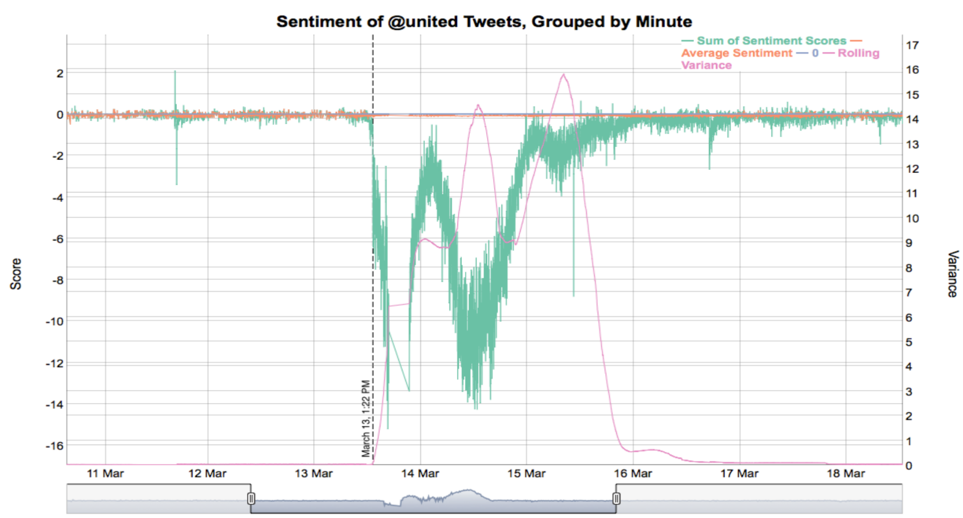

Figure 2: Sentiment of United Airlines Tweets

According to the comparison of United Airlines' daily stock with NYSE (where United Airlines is listed) and S&P 500 market indices, United's return and stock change are more extreme compared to markets at the intervention point (2018-03-13). The unusual change makes it reasonable to continue analyzing this point and checking if its volatility is also high.

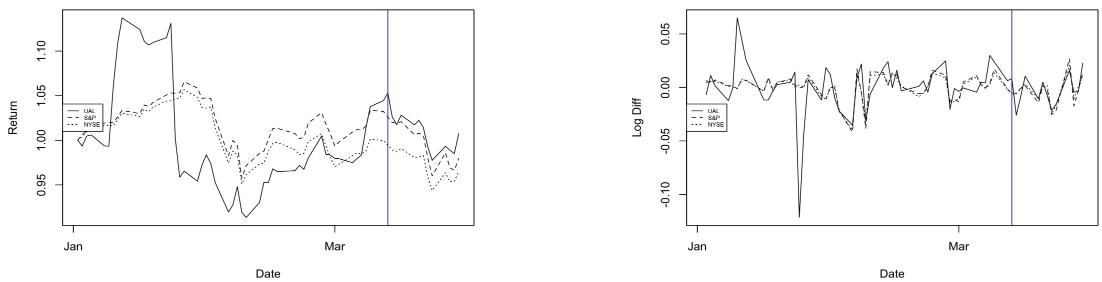

Figure 3: Return and Log Difference of United Daily Stock and Market Indices (Blue line: 2018-03-13)

After fitting a GARCH model with order (1, 1) “sGARCH” variance model and order (0, 0) mean model, we recognized that the intervention point (2018-03-13 13:22:00) has the second highest volatility. Besides, the sum of parameter $\alpha = 0.7579$ and $\beta = 0.2321$ is less than 1. Thus, this GARCH model is persistent and reliable. The high volatility also conforms to the doubled variance of sentiment scores.

μ	6.75e+01
ω	8.38e-13
α_1	7.59e-01
β_1	2.32e-01
shape	5.56e+01

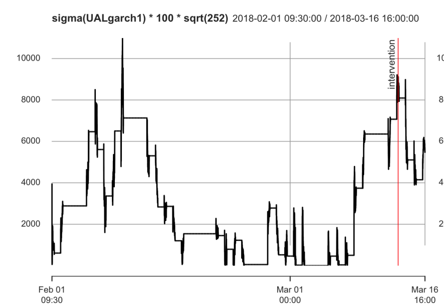

Figure 4: United Airlines GARCH Sigma

As the comparison with market and volatility both show that the intervention is affecting United Airlines' stock price, we then conduct the intervention analysis and aim to quantify the effect. We first fitted an ARIMA model on intra-day stock. To better describe the effect, we focus on a single day high frequency stock (2018-03-13) since social news and events are generally popular and important for a couple of hours. From the plot, we recognized a similar pattern to the fourth one mentioned in the Methods section that an increasing intervention eventually leveling off. We further modified the formula for the effect to be

$$z_t = \omega_0 I_t + \omega_0(1 - |\omega_1|)^{t-T+1} P_t$$

where $I_t = 1$ for $t = T$ and 0 otherwise, $P_t = 1$ for $t \geq T$ and 0 otherwise. T represents the intervention point and ω_0 and ω_1 are coefficients from MA and AR model, respectively. We replicated the same procedure on United Airlines daily stock and the result reinforced that the intervention caused short-term increase to stock prices with a trend back to mean value.

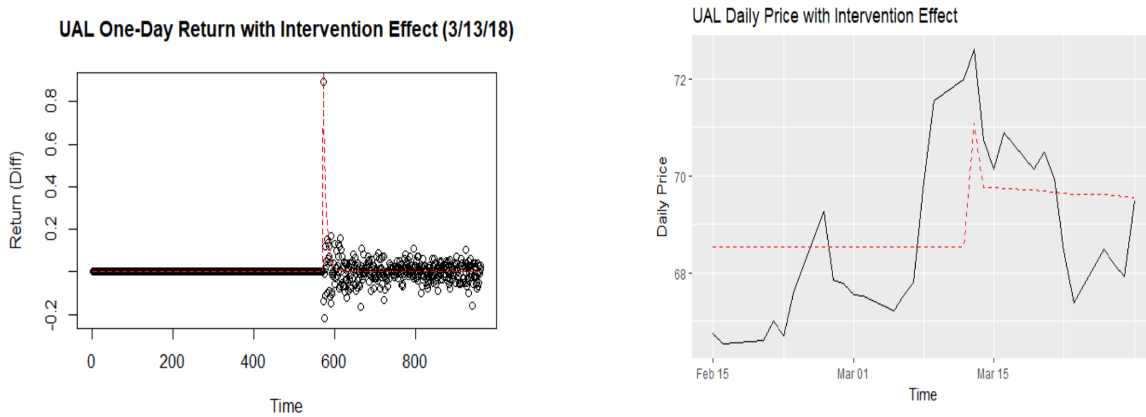

Figure 5: United Airlines Intervention Effect

Surprisingly, the stock demonstrated an increase while the sentiment scores dropped from -0.88 to -6. The reason this contradiction happened could be that we picked the point where sentiment score began to drop instead of the point with the lowest score. As United's stock was on an upward path right before the incident, it is reasonable to hypothesize that the drop after is due entirely to the spread of discussion of the tragedy on Twitter.

4.2 Amazon

From the Amazon sentiment time series, it is not hard to notice two abnormally low sum of scores. The first is on March 30 and the second is on April 2. We went through President Trump's tweets and found his first tweet expressing his concerns about Amazon was on March 29, followed by two tweets on March 31, one on April 2 and one on April 3. Because the first abnormal point (March 30) was on holiday, the second point (2018-04-02 08:19:00) is selected. It is further slightly pushed back to the next market opening hour for accommodation of the stock data. Therefore, the final intervention point for Amazon is "2018-04-02 09:30:00".

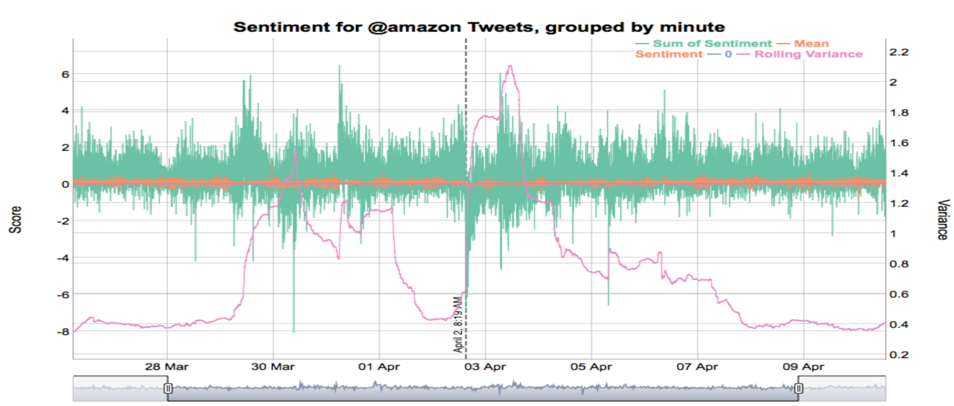

Figure 6: Sentiment of Amazon Tweets

Comparison of Amazon's daily stock with NASDAQ (where Amazon is listed) and S&P 500 market index shows that Amazon's stock movements are inconsistent with those of markets around the intervention point (2018-04-02). Its percentage of change is less stable, and thus it is practical to continue on measuring the intervention effect.

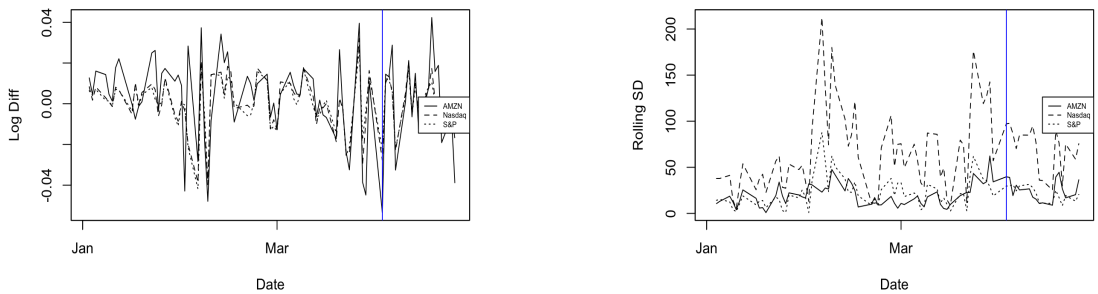

Figure 7: Log Difference and Rolling SD of Amazon Daily Stock and Market Indices (Blue line: 2018-04-02)

GARCH model also exhibits high volatility and the highest volatility point (2018-04-02 15:44:00) is actually pretty close to the intervention point (2018-04-02 09:30:00). The sum of parameters α and β is again less than 1, showing that "2018-04-02 09:30:00" is a dependable point of intervention analysis and quantification.

μ	1.50e+03
ω	3.17e-06
α_1	8.77e-01
β_1	1.18e-01
shape	8.28e+01

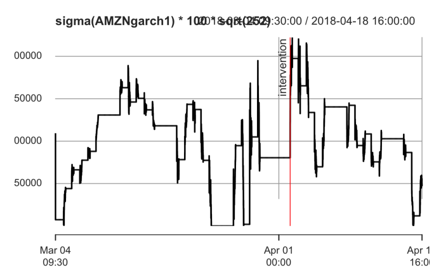

Figure 8: Amazon GARCH Sigma

We observe a similar pattern of the intervention effect on Amazon to that of United Airlines. However, this time the stock shows an immediate decrease followed by a return to mean level (Figure 9). Thus, the same quantification formula is also applicable to Amazon case with a negative MA model coefficient ($\omega_0 < 0$).

5 Conclusion and Next Steps

From the analysis of United Airlines and Amazon above, we conclude that mentions on social media platforms and sentiments of a company do affect the stock prices. However, slight movement of Twitter sentiment may not affect the stock. When we tried to repeat same analysis on other companies like Disney and Alibaba, the intervention point selected from

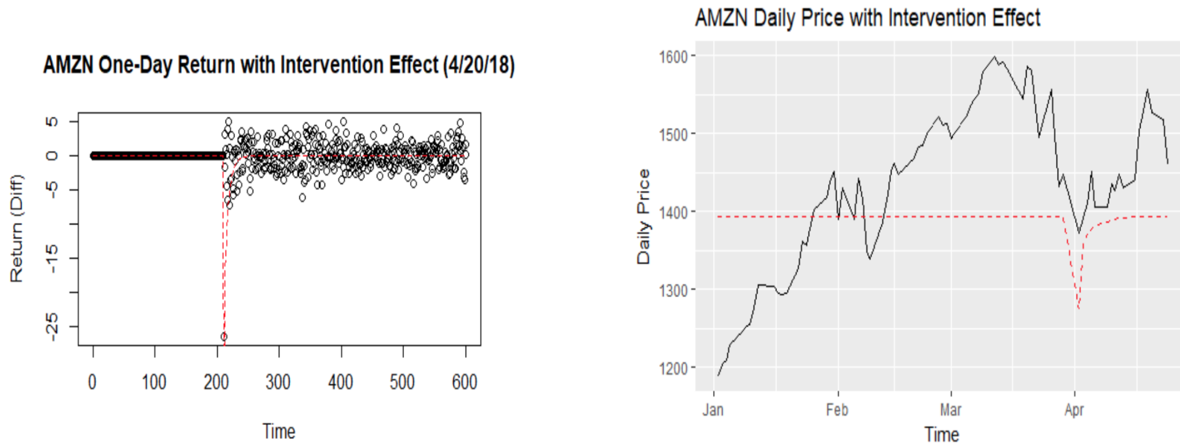

Figure 9: Amazon Intervention Effect

the sentiment analysis does not have corresponding volatile movement in stock prices. We reckon the reason might be that although the sum of sentiment scores change dramatically, the rolling variance does not shift a lot. It is also due to our limitation on the access of tweets. As mentioned in the Data Description section, we are only granted tweets access of the past six to nine days. Even though we have been constantly downloading tweets, the period is still not long enough to cover significant news and interventions.

In this project, we only considered positive and negative numbers of sentiment scores. Expansion to non-polarity sentiment analysis could be an interesting next step. For example, the six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) mentioned in Bollen et al.[3] can be explored.

Another aspect can be improved is tweets selection. In this study, we emphasized on all tweets directed to target company. The sentiment analysis could be more accurate if some cut-off criteria are assigned, such as only include tweets from users with more than 500 tweets posted and only use tweets that have been liked or re-tweeted.

References

[1] Yurieff K. 23 February 2018. *Snapchat stock loses $1.3 billion after Kylie Jenner tweet*

[2] Edison Research, Triton Digital, Salesforce.com. *Percentage of U.S. population who currently use any social media from 2008 to 2017.* In Statista - The Statistics Portal. Retrieved May 9, 2018.

[3] Bollen J., Mao H., and Zeng X. February 2011. *Twitter mood predicts the stock market* Journal of Computational Science, Volume 2, Issue 1, 2011, Pages 1-8.

[4] Nguyen T. H., Shirai K., and Velcin J. July 2015. *Sentiment analysis on social media for stock movement prediction*. Expert Systems with Applications, Volume 42, Issue 24, 2015, Pages 9603-9611, ISSN 0957-4174,

[5] Powers M. 13 March 2018. The Washington Post. *Puppy dies after being put in overhead bin on United flight from Houston to NYC*

[6] Chmielewski, D.C. 2 April 2018. *Amazon Stock Pummeled Monday Following Trump Twitter Rant*. Deadline.

[7] Gentry J. twitteR: R Based Twitter Client. R package version 1.1.9.

[8] Ryan J. A., Ulrich J. M. quantmod: Quantitative Financial Modelling Framewor. R package version 0.4-13

[9] Rinker T. sentimentr: Calculate Text Polarity Sentiment. R package version 2.2.3.

[10] Feuerriegel S., Proellochs N. SentimentAnalysis: Dictionary-Based Sentiment Analysis. R package version 1.3-2

[11] Prollochs N., Feuerriegel S. and Neumann D. *Generating Domain-Specific Dictionaries Using Bayesian Learning*. 23rd European Conference on Information Systems (ECIS 2015). Münster, Germany. May 26-29, 2015.

[12] Zeileis A., Grothendieck G. and Ryan J. A. zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations). R package version: 1.8-1.

[13] Business 41202. University of Chicago Booth School of Business. Spring Quarter 2015. Mr. Ruey S. Tsay.

[14] Pat. 6 July 2012. *practical introduction to garch modeling*. R-bloggers

[15] Boudt K., Cornelissen J., & Payseur S. highfrequency: Tools for Highfrequency Data Analysis. R package version 0.5.3.

[16] Hornik K. 9 February 2018. tseries: Time Series Analysis and Computational Finance. R package version: 0.10-43.

[17] Chan K. 13 November 2012. TSA: Time Series Analysis. R package version: 1.01.

[18] STAT 510. PennState Eberly College of Science. Applied Time Series Analysis.