The Forecasting Power of Twitter Sentiment on Future Stock Performance

A micro analysis on customer-facing & non-customer-facing companies

Yuhui Dai

Capstone Advisor: Professor Christian Haefke
New York University Abu Dhabi

April 24th, 2017

Thesis submitted to the Division of Social Sciences at New York University Abu Dhabi for the Economics Capstone Project in 2017

I submit this Thesis for the Capstone project in Economics to the Division of Social Sciences at New York University Abu Dhabi. The Thesis provides sufficient information on data sources, data periods, and data manipulation and transformation to allow for a full replication of all results.

Academic Integrity Declaration

I hereby declare that this Thesis is solely my work and contains no material I have previously submitted for assessment at New York University or elsewhere, without the express approval of all instructors, including my Capstone advisor(s). To the best of my knowledge and belief, the Thesis contains no material previously published or written by another person except where due reference is made in the Thesis.

I declare that I have read, and in undertaking this research, have complied with the Policy for Academic Integrity of Students at New York University. I also declare that I understand what is meant by plagiarism and that this is unacceptable. Except where I have expressly indicated otherwise, this Thesis is my own work and does not contain any plagiarized material in the form of unacknowledged quotations or any other material. Any data sources and computer code used to derive the results presented in this Thesis—including digital and electronic media source—have been properly acknowledged.

Name: Yuhui Dai

Signature:

Date: 2017/04/24

Table of Contents

List of Figures

The Forecasting Power of Twitter Sentiment on Future Stock Performance[1]

A micro analysis on customer-facing & non-customer-facing companies

Yuhui Dai
Class of 2017

Abstract

Social media are increasingly reflecting and influencing behaviors of other complex systems. In this paper, I investigate whether Twitter's tweet sentiment (aka: tweet sentiment) can forecast firm specific stock performances. In particular, I compare, in a period of 62 days, tweet sentiment's predictive power on stock performance in both customer-facing companies and non-customer-facing companies. I want to explore whether tweet sentiment correlates with individual stock prices and returns differently depending on whether the company is customer-facing or non-customer-facing. Therefore, I analyze 5 customer-facing companies and 5 non-customer-facing companies. My hypothesis is the following: Negative tweet sentiment can more accurately forecast individual firm's future stock performances than positive tweet sentiment. Non-customer-facing companies' stock performances correlate with tweet sentiment more significantly than customer-facing companies' do. However, my research finds that in the short term, there is little predictive power of tweet sentiment on stock performance at an individual firm level, and the Efficient Market Hypothesis holds.

Keywords: Sentiment, Twitter, Stock Performance

[1] Special thanks to Professor Christian Haefke for generous mentorship and support. Many thanks to Lucas Siga for hosting capstone seminar.

1. Introduction

Economic researchers have been interested in finding ways to predict what will happen in the future. The recent technological revolution with widespread presence of computers and Internet has created an unprecedented situation of data deluge, changing dramatically the way in which we look at social and economic sciences. In business, quantitative analysts started to recognize the value of 'Big Data' in the recent two decades[2]. In particular, financial investors employ all kinds of means to predict asset values. In academia, diverse research communities investigate whether statistical tools can create accurate predictions. There are many data modeling techniques developed in the fields of financial engineering, operation science, and management science and engineering.

Recently, both investor and customer sentiment data forecasting has received significant attention. Investor sentiment is defined as a belief about future cash flows and investment risks that are not justified by commercially available data (Baker and Wurgler, 2006). Customer sentiment is a statistical measurement and economic indicator of the overall health of the economy as determined by customer opinions[3]. Using sentiment data to predict asset value movement has become an important topic in both academia and industry. Researchers have utilized numerous methods and independent sentiment variables to predict future asset prices, such as stock price, commodity and precious metal futures. The topic is relatively new, with a significant amount of research on behavioral finance conducted within the past decade. Among the existing research, there are mainly three categories of sentiment data: the media-based sentiment value, the pre-constructed sentiment index, and the

[2] Lohr, Steve. "The Age of Big Data." The New York Times. February 11, 2012. Accessed December 07, 2016. http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

[3] Root. "Customer Sentiment." Investopedia. October 25, 2010. Accessed November 18, 2016. http://www.investopedia.com/terms/c/customer-sentiment.asp.

self-constructed sentiment proxy. The majority of research focuses on using sentiment data to forecast stock market movement with pre-constructed sentiment index.

Among the three categories of sentiment data, the media-based sentiment data produce by far the most consistent and promising results. The majority of the past research focuses on media-based sentiment's predictive power of aggregate stock movements, such as the movements of S&P 500, Dow Jones, and DAX, with a minority focusing on industry specific movements, such as in financial sectors and technology sectors.

Sentiment reflects the level of optimism investors have towards the market. The sentiment, no matter driven by rational analysis or irrational intuition, is an important factor for us to study. On the one hand, if the sentiments are driven by rational analysis of data that are available to the investors, they can be used to extrapolate the future performance of the market. On the other hand, even if the sentiments are irrational, herding behavior can potentially aggregate these sentiments and influence stock volatility. As Keynes suggests, investors can make money by predicting how the crowd is likely to behave in the future and beat the crowd by doing that first. Therefore, I intend to analyze the predictive power of sentiment on stock performance.

In contrast to the existing literature, my research investigates the sentiment data's predictive power on an individual firm level. The objective of this paper is to analyze whether or not the sentiment of Twitter users' tweet content (aka: tweet sentiment) towards specific companies can be an indicator of the company's future stock performances. In particular, I am comparing the predictability of tweet sentiment in both customer-facing and non-customer-facing companies' stock performances.

This paper is structured as the following: section 2 summarizes existing literature on relevant topics and states my research hypothesis; section 3 deals with the data collection, variables creation, and methodology used, while section 4 highlights sample statistics and results. Lastly, section 5

discusses results and potential future improvements.

2. Literature Review

Different types of sentiments, both investor sentiment and customer sentiment, have been studied worldwide. Although most of the research done in this area focuses on the US market, similar studies investigating sentiment index' predictive power on stock prices have been carried out in South Korea (Kim and Park, 2015), China (Jiang and Wang, 2009), the UK (Hudson and Green, 2015), Germany (Lux, 2010), India (Kumari and Mahakud, 2015) and 16 other countries (Schmeling, 2009). Overall, it has been found that there is a weak relationship between sentiment index and future stock returns. One significant discovery is that sentiment tends to be a more important determinant of returns in an economic crisis period than at other times. However, the direction of the relationship varies from study to study. The discrepancy in the results may be explained by different sources of sentiment measurements. Across the past literature, researchers usually employ one of the following three types of sentiment measurements: the pre-constructed sentiment index, the media-based sentiment value, and the self-constructed sentiment proxy.

2.1 Pre-constructed Sentiment Index

Pre-constructed sentiment index includes the sentiment measurements offered by third-party organizations, either from academic institutions or the research department in business entities. Baker and Wurgler (2006) construct six indexes that have been widely used to predict stock values. These indexes include market turnover, closed-end fund discount, new equity issuances, number of IPOs, first day return on IPOs, and the difference in book-to-market ratios between dividend payers and dividend non-payers. Using these six indexes, Canbas and Kandir (2009) and Spyrou (2012) observe that stock portfolio returns seem to affect all investor sentiment proxies, not the other

way around. Mian and Sankaraguruswamy (2012) find that the predictive power of sentiment is more pronounced for specific stocks: small stocks, young stocks, high volatility stocks, non-dividend-paying stocks, and stocks with extremely high and low market-to-book ratios. Furthermore, according to Baker and Wurgler (2006), the above-mentioned stocks yield low subsequent returns if previous sentiment is high, and yield high subsequent returns if the previous sentiment is low. These stocks are characterized as more speculative, and thus less persistent, making the valuation of the NPV (net present value) more difficult and subjective. Therefore, there is a more significant effect of sentiment on these stocks' performances.

Commercial research institutes are another source of popular investor sentiment data, which researchers can purchase. The sentiment surveys from the American Association of Individual Investor (AAII), and Investor Intelligence (II) are the most popular sources that researchers frequently utilize. These sentiment indexes are obtained from investors' opinions. II tracks the number of market newsletters that are bullish, bearish, or neutral. AAII's investor sentiment survey shows the percentage of investors who are market bullish, bearish, or neutral on stocks. Using these measurements, which are updated weekly, Wang, Keswani and Taylor (2006) discover a lagged result that is similar to Spyrou's (2012). In addition, Brown and Cliff (2004) find that sentiment has little predictive power for near-term future stock returns.

A third source of sentiment data is the Customer Confidence Index (CCI) and the Economic Sentiment Index (ESI), which instead of measuring sentiment of investors, measure customers' sentiment towards the market. In 2006, the University of Michigan's Customer Confidence Index was validated as an effective sentiment measurement tool (Lemmon and Portniaguina, 2006). Researchers find contradictory results with these customer-based sentiments. Fisher and Statman (2003), by analyzing the relationship between customer confidence and stock values from S&P 500 and NASDAQ, find that the CCI

has no significant predictive value of stock returns, and tends to follow investors' sentiment. Kadilli (2015) detects an insignificant effect of sentiment on future returns during non-crisis times, yet a positive and strongly significant effect during crisis times. However, Chung et al (2012) discovers the exact opposite effect: the CCI is significantly more correlated to future stock returns during economy expansion rather than crisis times.

2.2 Media-based Sentiment Value

Both traditional media, including news and message boards, and social media such as Twitter and Facebook provide sentiment data that can potentially forecast future stock movements. Hautsch and Groß-Klußmann (2011), and Lee (2015) utilize Thomson Reuters News Analytics, a popular news analytics tool of the Reuters Company, and discover a significant positive relationship between news sentiment and future stock returns.

Sentiment data can be also obtained from online stock message boards. An online stock message board, such as the Yahoo Finance Message Board, can be used as a herding device to temporarily forecast stock prices (Sabherwal, Sarkar and Zhang, 2011). It is discovered that the online traders' credit-weighted sentiment index is positively associated with contemporaneous returns but negatively predicts the returns next day and two days later. The correlation between contemporaneous returns and sentiment index could be explained by the herding behavior among online traders. The subsequent negative correlation could be possible reactions to the contemporaneous returns. For instance, a significant drop could be due to the dumping by influential posters who have originally pumped up the stock. After profit has been fully exploited by them and their followers on the message board, the stocks return to their fundamental values. Furthermore, these messages can help predict market volatility (Antweiler and Frank, 2004). During recession, news sentiments yield significant predictive results (Garcia 2013). However, Kim (2014) disagrees with these results: he finds that at both

aggregate and individual level, news sentiment forecasts of future stock values are inaccurate.

Besides acquiring sentiment data from traditional media, social media, which reveal customers' and investors' thoughts, behaviors and feelings, can capture a vast range of events and topics in the market (Sprenger et al, 2014). Therefore, they become popular tools for researchers to predict stock movements. According to Yu et al (2003), social media has a stronger forecasting power of stock performances than traditional media (e.g., news). Among them, Twitter stands out as the most frequently analyzed social medium to forecast stock values both on an industry level and an aggregate market level.

Bollen et al (2011) uses financial tweets (tweets that use stock tickers to specifically discuss financial issues) and their associated investor's mood in order to predict the Dow Jones Industrial Average Index. His team finds that the accuracy of the DJIA forecasts could be notably improved by including two of the 6 mood dimensions, 'Calm,' and 'Happy,' rather than 'Alert,' 'Sure,' 'Vital,' and 'Kind.' 'Hope' and 'Fear' derived from twitter feeds are also correlated with Dow Jones, S&P 500 and NASDAQ (Zhang 2011). An indicator of Twitter Investor Sentiment and the frequency of occurrence of financial terms on Twitter in the previous 1-2 days are found to be statistically significant predictors of daily market returns (Mao 2011).

At an industry specific level, both significant and insignificant results are observed. Wu et al (2016) finds that compared with positive emotions, firm specific negative twitter sentiment in the financial sector can predict future and stock movements. However, in the technology industry, Corea (2016) finds that it is not the tweet sentiment that can predict future stock returns, but the volume of the tweets subjected to a specific company. Similarly, Ranco et al (2015) finds that during peaks of Twitter volume, the dependence of tweet sentiment and stock abnormal returns is highly significant.

2.3 Self-constructed Sentiment Proxy

The third category of sentiment data is the self-constructed sentiment proxy that has been frequently employed by researchers. These measurements include VIX, Put-Call Ratio, TRIN (Simon and Wiggins, 2001), Trading Volume, Market Liquidity (Baker and Stein, 2004), among others.

3. Data and Methodology

Section 2 sums up the existing literature for sentiment analysis forecasting. Among the three categories of sentiment data (the pre-constructed sentiment, the media-based sentiment and the self-constructed proxy), the media-based sentiment generates the most promising results. Among the available sub-components of media-based sentiment, Twitter, which offers open API (Application Programming Interface)[4], become the most feasible data source for my research. Using data from Twitter and ten S&P 500 companies over a two-month period, I hope to explore whether tweet sentiment can forecast firm specific stock performances. In particular, I classify the ten companies into 5 customer-facing and 5 non-customer-facing companies. Customer facing companies interact directly with consumers and offer business to consumer service, whereas non-customer facing companies offer business to business service. I want to explore whether tweet sentiment correlates with individual stock prices and returns differently, depending on whether the company is customer-facing or non-customer-facing. My hypothesis is the following: Negative tweet sentiment can more accurately forecast individual firm's stock future performances than positive tweet sentiment. Non-customer-facing companies' stock returns correlate with tweet sentiment more significantly than customer-facing companies do.

[4] "API Overview." Twitter Developer Documentation. Accessed September 14, 2016. https://dev.twitter.com/overview/api.

3.1 Twitter API Streaming

Twitter, a rapidly growing microblogging platform, allows users to post and read text-based messages (aka: tweets) of up to 140 characters in length. As of the 3rd quarter of 2016, Twitter has over 313 million monthly active users worldwide.[5] Over three quarters of the Fortune Global 100 companies own one or more Twitter accounts at the corporate level and for specific brands (Malhotra and Malhotra, 2012). Most importantly, Twitter offers open public API, where one or more filtering parameters can be specified. When an API request with a specific filtering parameter is sent to twitter, a limited, randomly sampled stream of tweets that satisfy the parameters, will be returned in json[6] format that would be processed by my python programs. In this study, both the company's full name and its ticker are used as filtering parameters. For instance, Michael Kors' search terms are 'MichaelKors,' and '$KORS.' My capstone analyzes five customer-facing companies: Michael Kors, Coca Cola, Starbucks, Nike and Hershey, and five non-customer-facing companies: Goldman Sachs, Morgan Stanley, Accenture, Exelon, and Ever Source. Table 1 shows specific search terms associated with each company. Red represents customer-facing companies, and blue represents non-customer-facing companies. I aggregate Twitter data for 12 hours from 8am EST to 8pm EST daily for each of the ten companies for 62 days from November 14, 2016 to Jan 13th, 2017.

[5] "Twitter Usage/Company Facts." Twitter Company. Accessed November 14, 2016. https://about.twitter.com/company.
[6] "Introducing JSON." JSON. Accessed November 16, 2016. http://www.json.org/.

Company	Ticker	Search Name
Coca Cola	$KO	CocaCola
Hershey	$HSY	Hershey
Michael Kors	$KORS	MichaelKors
Nike	$NKE	Nike
Starbucks	$SBUX	Starbucks
Accenture	$ACN	Accenture
Ever Source	$ES	EverSource
Exelon	$EXC	Exelon
Goldman Sachs	$GS	GoldmanSachs
Morgan Stanley	$MS	MorganStanley

Table 1. Ten Companies and their search terms

3.2 Sentiment

Twitter data become a useful source of information for opinion mining and sentiment analysis. There are mainly two ways of measuring tweet sentiment in the existing literature. One way is to employ Natural Language Processing (NLP) tools. For instance, He et al (2016) uses Lexalytics to detect tweet sentiment. Bollen et al (2011) employs OpinionFinder[7] and GPOMS to measure twitter sentiment. The other way to measure sentiment is to manually code the tweet sentiment. For example, Ranco et al (2015) hires 10 financial experts to label over 100,000 tweets with 'positive,' 'negative' and 'neutral' sentiments, which are used to build a Support Vector Machine (SVM). Mao et al (2011) has constructed the "Twitter Investor Sentiment" as a function of the number of occurrences of "bullish" and "bearish" in the tweets.

[7] "OpinionFinder | MPQA." OpinionFinder | MPQA. Accessed November 14, 2016. http://mpqa.cs.pitt.edu/opinionfinder/.

In this research, I use the sentiment analysis functionality of Microsoft Cognitive Services[8] to identify the positive, negative or neutral sentiment within a given text for a specific target, towards which Microsoft will analyze the text sentiment. The targets are the same as the filtering parameters I used to collect Twitter streaming data. For instance, I used "MichaelKors" and "$KORS" to filter tweets related to Michael Kors. When I try to analyze the sentiment towards Michael Kors via Microsoft's API, I not only input the exact tweet I obtain as the text, but also specify "MichaelKors" and "$KORS" as my targets. Once my sentiment API receives and analyzes the text, it returns target sentiment scores between -1 to 1 for the targeted object, in this case, "MichaelKors" or "$KORS." A sentiment score between -1 and 0 denotes negative sentiment, whereas a score between 0 and 1 represents positive sentiment. 0 means the text is neutral. I specify a target to obtain sentiment score because its sentiment score can be different from the overall sentiment score of the text. For instance, if the text is "I hate Hershey but M&M is fine," the overall sentiment of the text is 0.34 but it does not tell much useful information. The target sentiment for M&M is 0.33, which is positive, whereas the target sentiment for Hershey is negative, -0.56. When processing tweets and determining their sentiment scores, I remove URLs because they normally do not represent relevant content but rather point to it. I specify my target as either the company's name (e.g., "MichaelKors") or its tickers (e.g., "$KORS").

Finally, after collecting the sentiment data, I sort the sentiment data by percentile, weighted by the number of followers associated with each account. Specifically, I retrieve the scores of 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 99th percentiles in the daily sentiment distribution. In addition, I also calculate the mean and standard deviation of the daily sentiment. These sentiment

[8] Onewth. "Quick Start Guide: Machine Learning Text Analytics APIs." *Quick Start Guide: Machine Learning Text Analytics APIs | Microsoft Docs*. N.p., n.d. Web. 31 Jan. 2017.

variables are respectively regressed on the stock prices and returns of each company with different regression model weights.

3.3 Company Specific Stock Performance

I collect stock prices and returns from the Bloomberg database. Bloomberg data cover the ten companies' daily stock price from November 14th, 2016 to January 13th, 2017. I use the companies' closing stock price at 4:00pm each day as the daily stock price. For the regression model, I regress the sentiment variables on the current-day, next-day and third-day prices and returns to analyze whether sentiments correlate with the stock prices and returns over different time periods.

Stock price reflects the long-term effect of investors' expectations and evaluations of the companies. Return, on the other hand, reflects investors' short-term belief of the companies' performance. Therefore, I am analyzing whether sentiment correlates with any of those parameters.

4. Statistics & Results

According to the data, on average on a 12-hour streaming daily basis, non-customer-facing companies accumulate significantly fewer tweets than customer-facing companies. The specific statistics can be found in Table 2. Among the five non-customer-facing companies, Accenture and Goldman Sachs are mentioned most frequently on Twitter. Among the five customer-facing companies, Nike and Starbucks attract the most attention on Twitter. Trading day and non-trading day volume fluctuation is not significant for customer-facing companies. However, for non-customer-facing companies, there are fewer tweets gathered from a non-trading day than a trading day. This may be because non-customer-facing companies will be discussed more intensively on a trading day than a non-trading day.

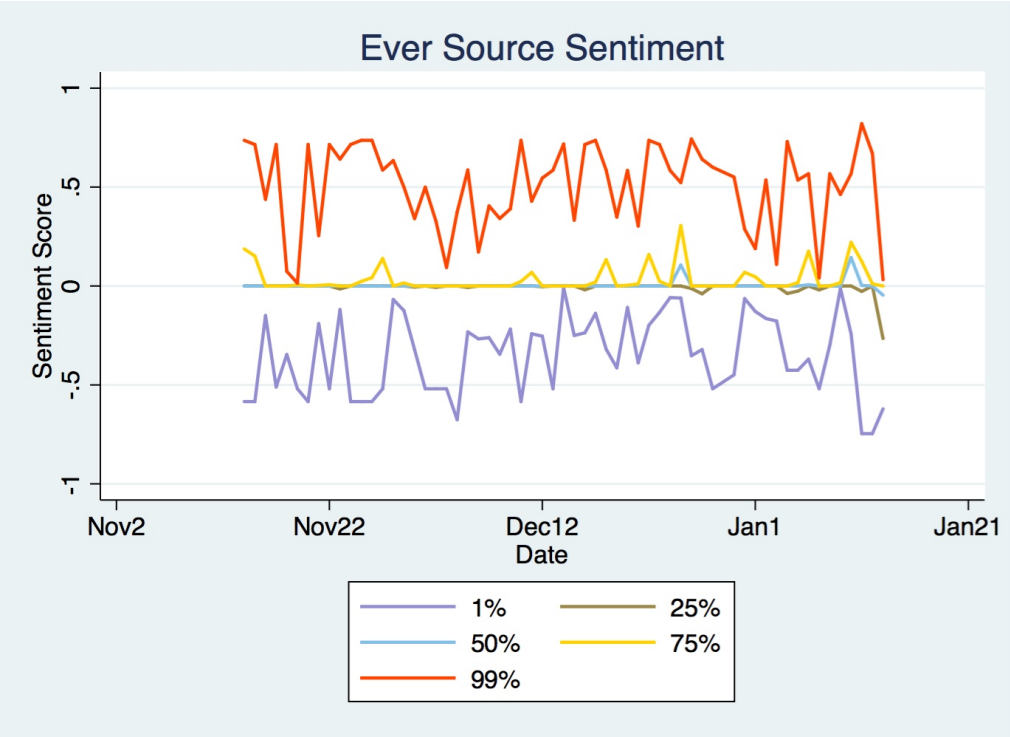
Company (non-consumer-facing)	Accenture	Ever Source	Exelon	Goldman Sachs	Morgan Stanley
Number of Tweets (Standard Error)	441 (218)	100 (63)	136 (141)	743 (678)	143 (104)
Company (consumer-facing)	Coca Cola	Hershey	Michael Kors	Nike	Starbucks
Number of Tweets (Standard Error)	1103 (634)	2210 (1256)	264 (203)	27622 (8914)	15374 (6547)

Table 2. Summary of average number of daily tweets for each company

I target sentiment for each company to ensure that the sentiment score is specifically aiming at the company that I am analyzing and to prevent a strong sentiment towards a third-party from distorting the sentiment data. Table 3 compiles a list of sample tweets and sentiment scores for each of the ten companies that I analyze.

Company	**Sentiment Score**	**Tweet Content**
Accenture	0.818465	@AccentureStrat: Accenture Strategy's Kevin Quiring takes customer experience to a whole new level.
Coca Cola	0.863061	@CocaCola: Pork sliders become otherworldly delicious when a refreshing Coke joins the party.
Hershey	0	PowerBuilder Business Analyst - Hershey, PA
Exelon	-0.200472	Question: Is Exelon introducing the bill to the legislature? Or a representative? Seems like a strange framing

Table 3. Sample tweets and sentiment scores



Graph 1. Ever Source Weighted Sentiment distribution

Graph 1 displays the time series sentiment distribution of Ever Source, an electric service company, over the analyzed period. The graph displays 1st percentile, 25th percentile, 50th percentile, 75th percentile and 99th percentile of the daily sentiment distribution. The distribution graphs for the rest of the companies can be found in the Appendix A.

My original hypothesis is that negative tweet sentiment can more accurately forecast an individual firm's stock performances, and the effect is more prominent in non-customer-facing companies. Therefore, I analyze the predictive power of sentiment at different percentiles because lower percentiles denote more negative sentiment than higher percentiles. In my regression model, I regress each of the sentiment percentile over the stock return and price of current day, second day, and third day respectively. In addition, I added the weight to my regression in Stata. The three weights are: $\frac{1}{variance}$, $\frac{\text{number of tweets on that day}}{\text{max number of daily tweets I ever collected during the 62-day period}}$, and

$$\frac{\text{number of tweets on that day}}{(\text{max number of daily tweets I ever collected during the 62-day period}) * (\text{variance})}$$, a combination of both. The regression formula are the following:

$$P = C_1 + \alpha S_{t0} + \beta S_{t-1} + \gamma S_{t-2} + \varepsilon_1 \quad (1)$$

$$R = C_2 + \alpha S_{t0} + \beta S_{t-1} + \gamma S_{t-2} + \varepsilon_2 \quad (2)$$

P stands for daily stock price and R stands for daily returns. S represent sentiment variables: different percentiles and standard deviation. The subscript denotes whether it is the sentiment is the current day, or it is from yesterday, or two days ago. Table 4 and 5 are examples of the regression table I create, with highlighted t-values significant at a p smaller or equal to 5%. Table 4 is Coca Cola's regression statistics, and table 5 is Ever Source's. Different weights for the regression model are labeled on top of each t-stats table. For instance, the upper left chunk of Coca Cola t-stats table is "Price weighted by volatility": namely, $\frac{1}{variance}$

Coca Cola		**(mean: 1103 tweets per day)**			**(sd: 634 tweets)**							
		Price weighted by volatility				Price weighted by # of tweets				Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	
1	1.0414699	3.7460298	0.92788362	1	2.1160002	5.2618099	4.2982521	1	0.5093831	4.1629368	0.6203275	
5	0.8651465	4.3270018	0.5982953	5	1.7455889	4.7545472	4.2177477	5	0.3329977	4.7946939	0.1478355	
10	0.7775865	4.7142341	0.34583657	10	1.6694539	4.2107537	3.4473216	10	0.3766824	5.1860525	-0.0934587	
25	-0.1854471	0.2003792	0.19277271	25	1.1426789	1.2793855	1.1105975	25	0.0477422	-0.0629101	-0.063908	
50	-3.4852414	-4.8402897	-0.16838824	50	-1.2120023	-2.1595489	-2.0150245	50	-3.5865571	-5.4380771	-0.092823	
75	-0.9473833	-4.8393053	-0.27529061	75	-1.5959419	-4.1952585	-3.1398393	75	-0.5713815	-5.3261626	-0.0451628	
90	-0.9662028	-4.1834692	-0.38551198	90	-1.712693	-5.1643587	-4.7537683	90	-0.5946743	-4.6730328	0.0261105	
95	-1.114866	-3.9497661	-0.52067247	95	-1.7003411	-5.07106	-4.9512406	95	-0.6798102	-4.4201607	0.0243724	
99	-1.100119	-3.6281592	-1.0333482	99	-2.0627881	-5.5599881	-4.6851807	99	-0.5891514	-4.1019251	-0.1514146	
sd	-0.93	-4.39	-1.24	sd	-1.85	-5.19	-4.51	sd	-0.44	-4.91	-0.65	
		Return weighted by volatility				Return weighted by # of tweets				Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	
1	-4.0251753	-0.1428792	0.42416643	1	-2.1346309	1.0340863	0.5013731	1	-3.7665337	0.8484559	0.4628578	
5	-3.9431297	0.2998346	0.1890167	5	-2.3145233	0.7078807	0.1763737	5	-3.7374136	1.2602616	0.1163838	
10	-2.8087616	0.6786096	0.08994078	10	-1.9271253	0.3790406	0.2108253	10	-2.8457291	1.574038	-0.01004	
25	2.0416521	0.0507371	-0.06000586	25	0.181036	-0.2723663	0.1884238	25	1.6968264	0.0863495	-0.1568411	
50	-0.1724078	-1.829686	-0.04109772	50	1.0106426	-0.6581694	-0.5638249	50	-0.579136	-2.3793254	-0.0967511	
75	2.6829579	-0.8777831	-0.04697584	75	2.2985164	-1.1822586	-0.0685088	75	2.6533099	-1.7239299	-0.0287295	
90	2.6177532	-0.2109678	-0.06029189	90	2.2371383	-0.6374172	-0.1982137	90	2.5960895	-1.1781261	0.0204675	
95	3.1503402	-0.0238761	-0.06263272	95	2.3580844	-0.4324759	-0.277515	95	2.9551529	-1.0103345	0.081556	
99	3.7594531	0.1533754	0.22867897	99	2.2351998	-0.8979714	-0.6701004	99	3.47138	-0.8464855	0.6293227	
sd	3.65	-0.35	-0.07	sd	2.37	-0.83	-0.43	sd	3.46	-1.34	-0.19	

Table 4: Coca Cola T-stats table

Accenture	(mean: 441 tweets per day)				(sd: 218 tweets)						
	Price weighted by volatility				Price wieghted by # of tweets				Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-1.6554776	-1.501662	0.10367976	1	-0.4816377	-0.0489468	0.56541674	1	-2.1582025	-1.9983661	0.47858593
5	-1.2776591	-1.2664121	-0.4507212	5	-0.2389525	0.31895419	0.61224908	5	-1.6216961	-1.5907218	0.23445161
10	-0.503794	-1.5153065	-0.2435101	10	0.40924321	0.7593898	0.52446212	10	-0.4816735	-1.7930143	0.3628455
25	0.16464787	0.21191786	0.00378499	25	0.13890025	-0.0067355	-0.3955832	25	0.12578233	0.19926853	0.10011061
50	0.39059824	0.40752014	0.11987488	50	0.22972247	-0.0138153	-0.4533952	50	0.26407862	0.39416436	0.10649701
75	0.69568667	0.78297809	0.24624912	75	0.23338748	0.03100907	-0.4782442	75	0.68606427	0.86379169	0.20094308
90	0.84616782	0.90285861	0.2547523	90	0.27956212	0.07094538	-0.4292714	90	0.88674652	0.98605241	0.04920366
95	0.9692484	0.95949836	0.40678542	95	0.33189705	0.12214009	-0.3789705	95	1.1656612	1.0701014	0.19944847
99	2.0830573	2.0783415	1.3222317	99	0.70262146	0.71482812	-0.1666568	99	3.0560143	2.4019625	0.48269893
sd	1.67	1.58	0.67	sd	0.39	0.14	-0.43	sd	2.12	1.78	0.11
	Return weighted by volatility				Return wieghted by # of tweets				Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-0.3049224	-0.5847541	1.4142587	1	-0.8945907	-0.7193188	0.04618527	1	-0.286823	-0.8840888	2.0066491
5	0.48760055	-0.3498044	1.001712	5	-0.855858	-0.2061215	0.23898252	5	0.83896687	-0.6045526	1.4729349
10	1.2855033	-0.6479983	0.76498994	10	-0.6093393	0.68872578	0.79864199	10	1.9625358	-0.9735803	1.0510044
25	0.20217801	0.29859774	0.04114528	25	0.35932998	1.12126	0.61178229	25	0.10514345	0.24320649	0.08691507
50	-0.1514488	0.18850277	-0.0866579	50	0.85071737	0.66777314	0.23954271	50	-0.2580572	0.18359127	-0.1805825
75	-0.0317259	0.25922657	-0.5012861	75	0.85071216	0.63334967	0.15173353	75	-0.1438534	0.2491621	-0.7013566
90	0.03476205	0.27072887	-0.8933969	90	0.88359069	0.60208915	0.10717885	90	-0.0307098	0.24612708	-1.2023689
95	0.21712507	0.35427439	-1.1291608	95	0.83511334	0.65431854	0.13261355	95	0.24234988	0.35349127	-1.5391007
99	-0.4705475	0.53131395	-2.97104	99	0.55787291	0.80917897	-0.1220699	99	-0.4243763	0.39083072	-3.0842807
sd	0.06	0.3	-1.69	sd	0.78	0.59	-0.06	sd	0.12	0.27	-2.29

Table 5: Accenture T-stats table.

However, despite significant correlation of certain sentiment percentiles with the stock price and return of a given company, no consistent pattern has been observed across the 10 companies. The rest of the results have been attached in the Appendix B, with highlighted t-values significant at a p smaller or equal to 5%. Hence, during the given analyzed period of time, there is no correlation between the tweet sentiment and stock performance of that company in the 3 days immediately following.

Lastly, python code and shell script for data acquisition and the Stata code for data analysis and regression are included in Appendix C.

5. Discussion

According to the Efficient Market Hypothesis, as stock market efficiency causes existing share price to incorporate and reflect all relevant information, investors cannot rely on elaborate techniques of security analysis to discover superior value opportunities. The “market is so efficient—prices move so quickly when information arises—that no one can buy or sell fast enough to benefit” (Malkiel, 184). As investor sentiment captures their beliefs about future cash flows and investment risks that are not justified by commercially available data, if sentiment correlates with stock performance significantly, it means that investors know something about the companies that

public information does not reveal. In this case, there is inefficiency in the market.

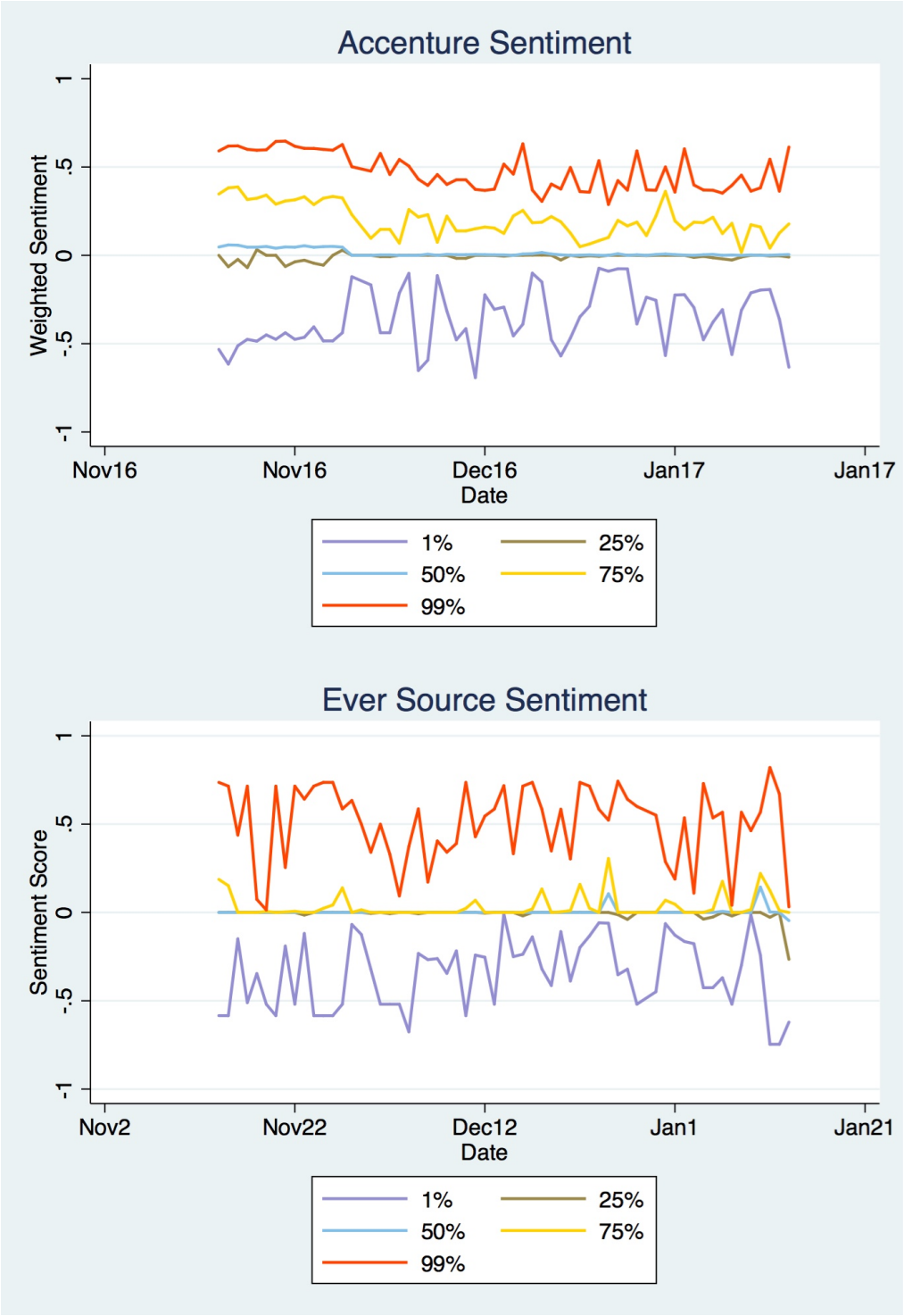
The absence of consistent predictive power between twitter's tweet sentiment and stock performance in both customer-facing and non-customer-facing companies suggests that the stock market, at least in my analyzed period, has been efficient.

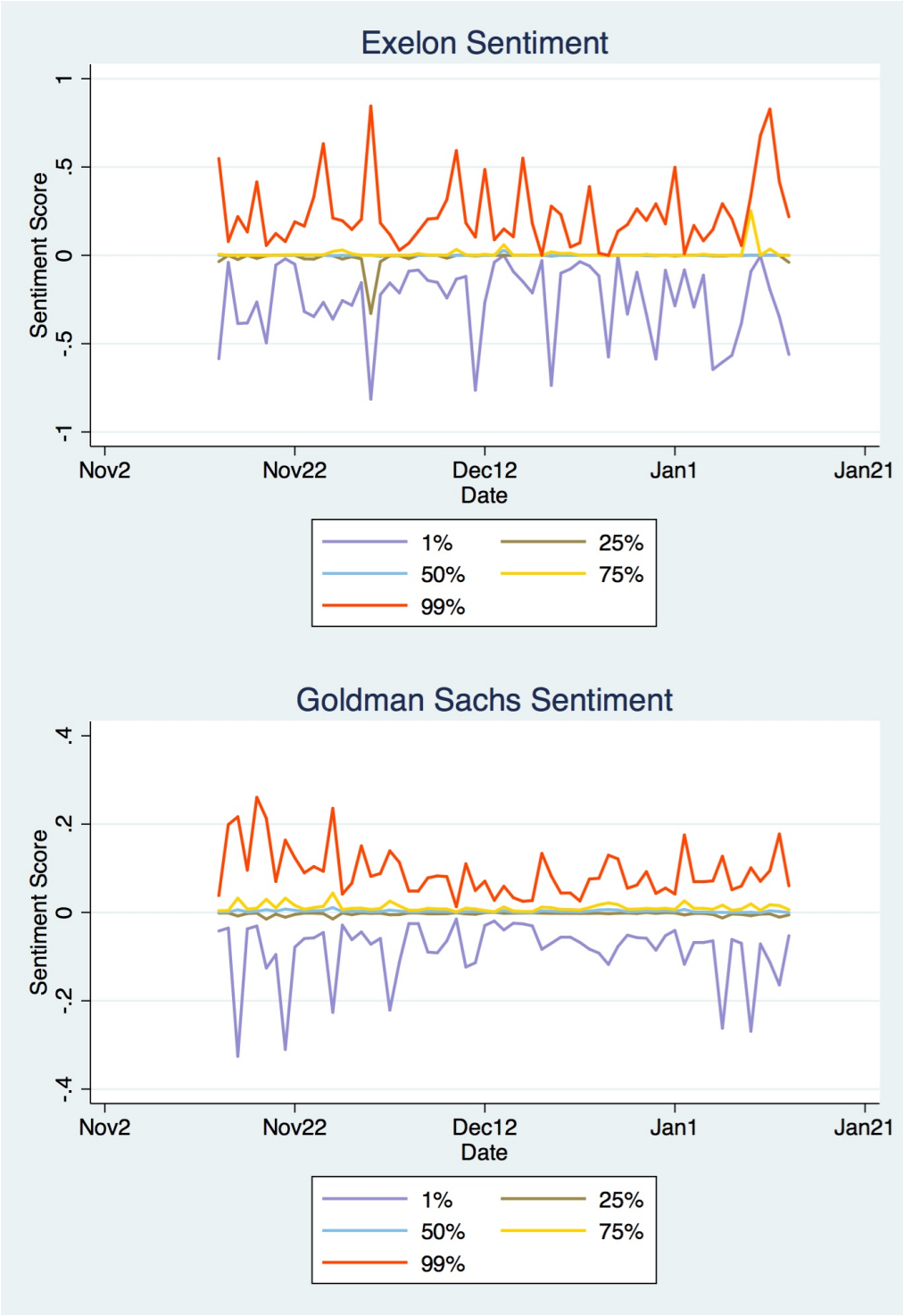
There are several limitations that can be addressed in future studies. First, the data span 62 days, a relatively short period. The past studies usually span from 7 months to 2 years, and longer and larger datasets will help to discover a more accurate relationship. Second, unlike Ranco et al (2015) who employ a group of experts manually reading and categorizing sentiment data, mine is entirely dependent on machines and algorithm. There are incidences that generate sentiment scores that are not accurate. For instance, one of the tweets addressing Michael Kors says "the new style @Michael Kors is so sick!" and is given a negative sentiment value. However, the slang word 'sick' here expresses an inverted meaning and has been used as a term of approval. The NLP algorithm fails to recognize the modern adaptation of words like this one, yielding inaccurate sentiment scores. Apart from inaccurate sentiment scores, there is noise in the data that will not be identified by programs but rather by people. Therefore, letting experts audit the sentiment scores will make the results more precise. Third, there are frequent retweets of the same content. In my research, I count each retweet a new tweet. However, rather than counting them independently, aggregating the total number of followers as the weight and counting all the retweets as one tweet may be a better solution to avoid repetition and skewing the sentiment data distribution. Lastly, currently all the tweets with keywords that match the company's name and stock label are taken into consideration. These tweets come from customer accounts, investor accounts and the company departments' accounts as well. Categorizing them and analyzing them separately will reduce the chance that the company's own advertisements influence the sentiment data.

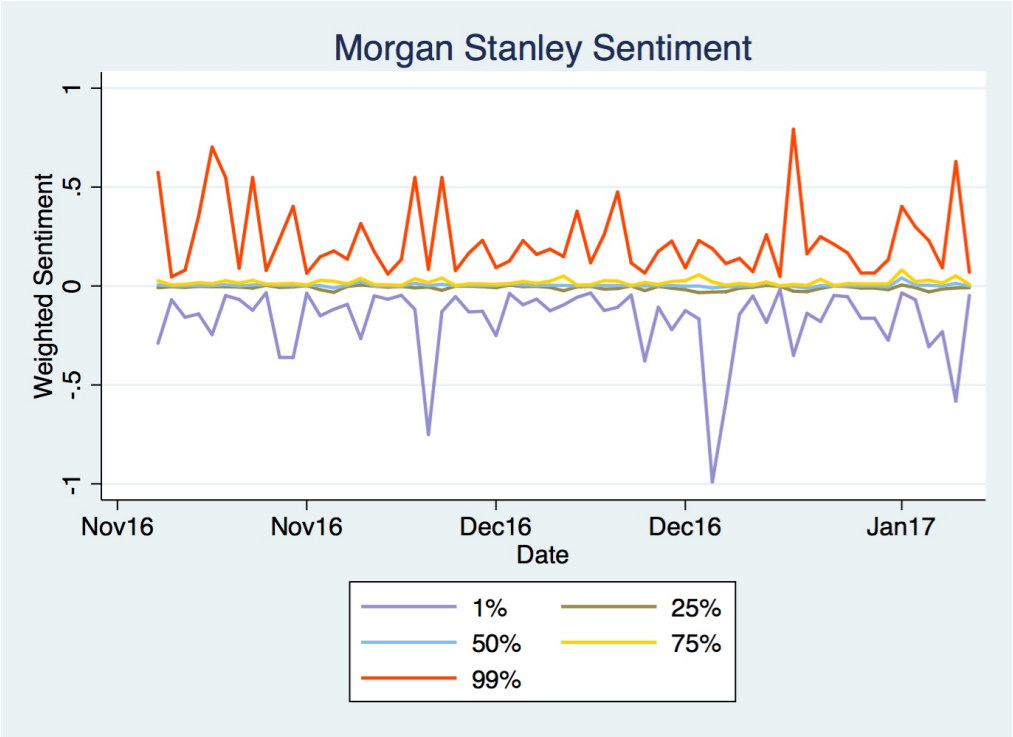
This paper finds little predictive power of tweet sentiment on stock performances at an individual firm level. It supports the Efficient Market Hypothesis at the analyzed period. The instrument used in this research could be applied to examine whether the market is efficient or not. In the short term, there is no consistent linear relationship between tweet sentiment and stock performances. However, in the long run, there may be a non-linear relationship. For instance, if there were a bubble in the market and stock price kept rising and sentiment fell significantly before the market crash, sentiment might potentially detect bubbles in the market and signify the imminent burst of the bubble. This could be an interesting topic to explore. Further research will be needed to test the validity of such method in detecting a bubble.

Appendix A

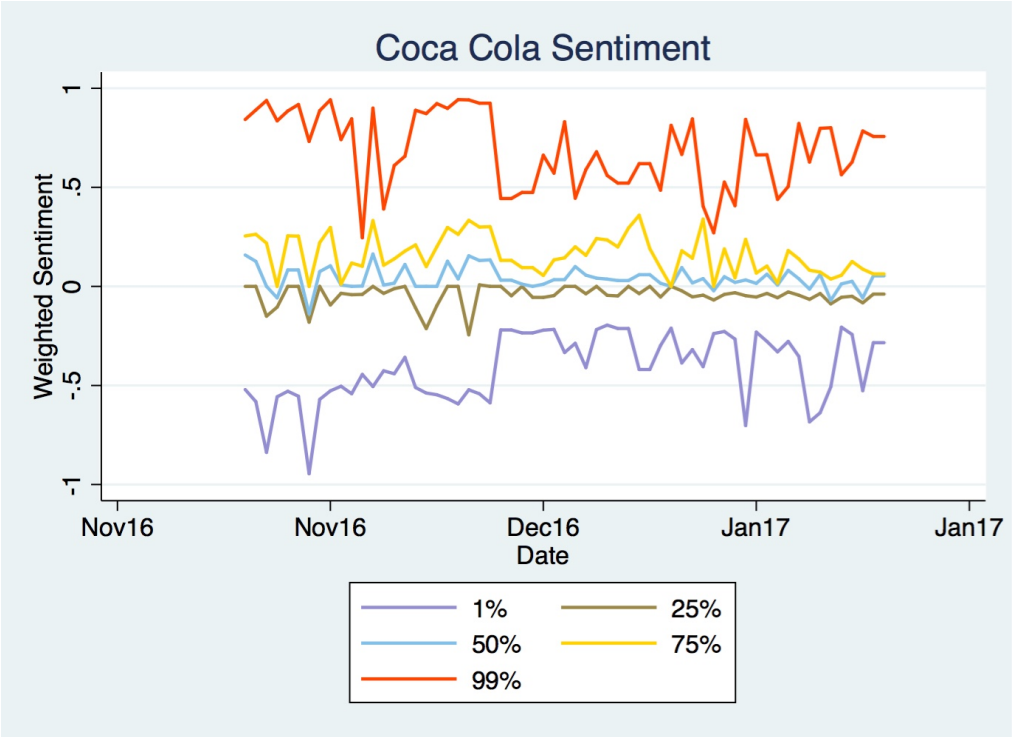
Non-customer Facing Companies

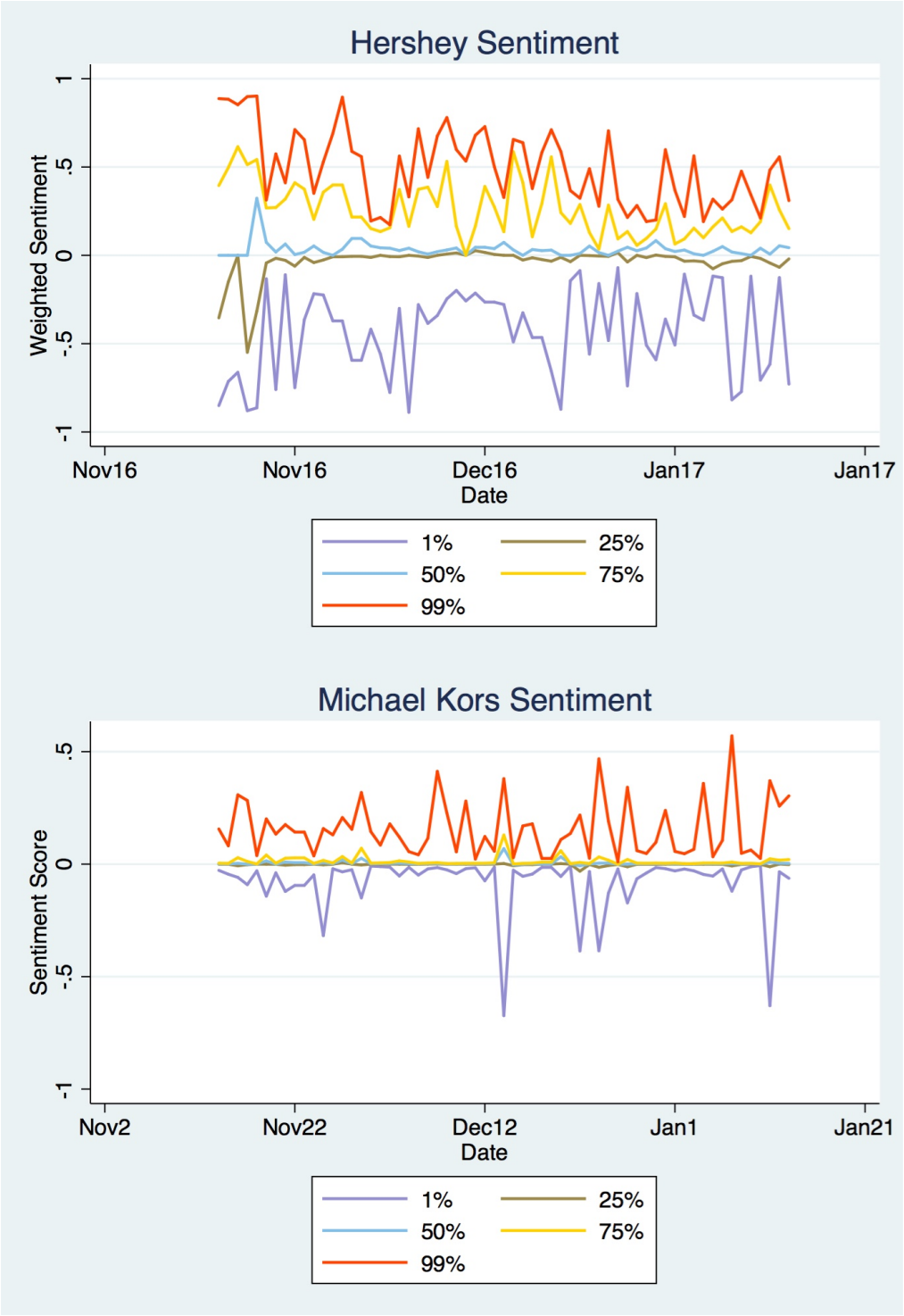


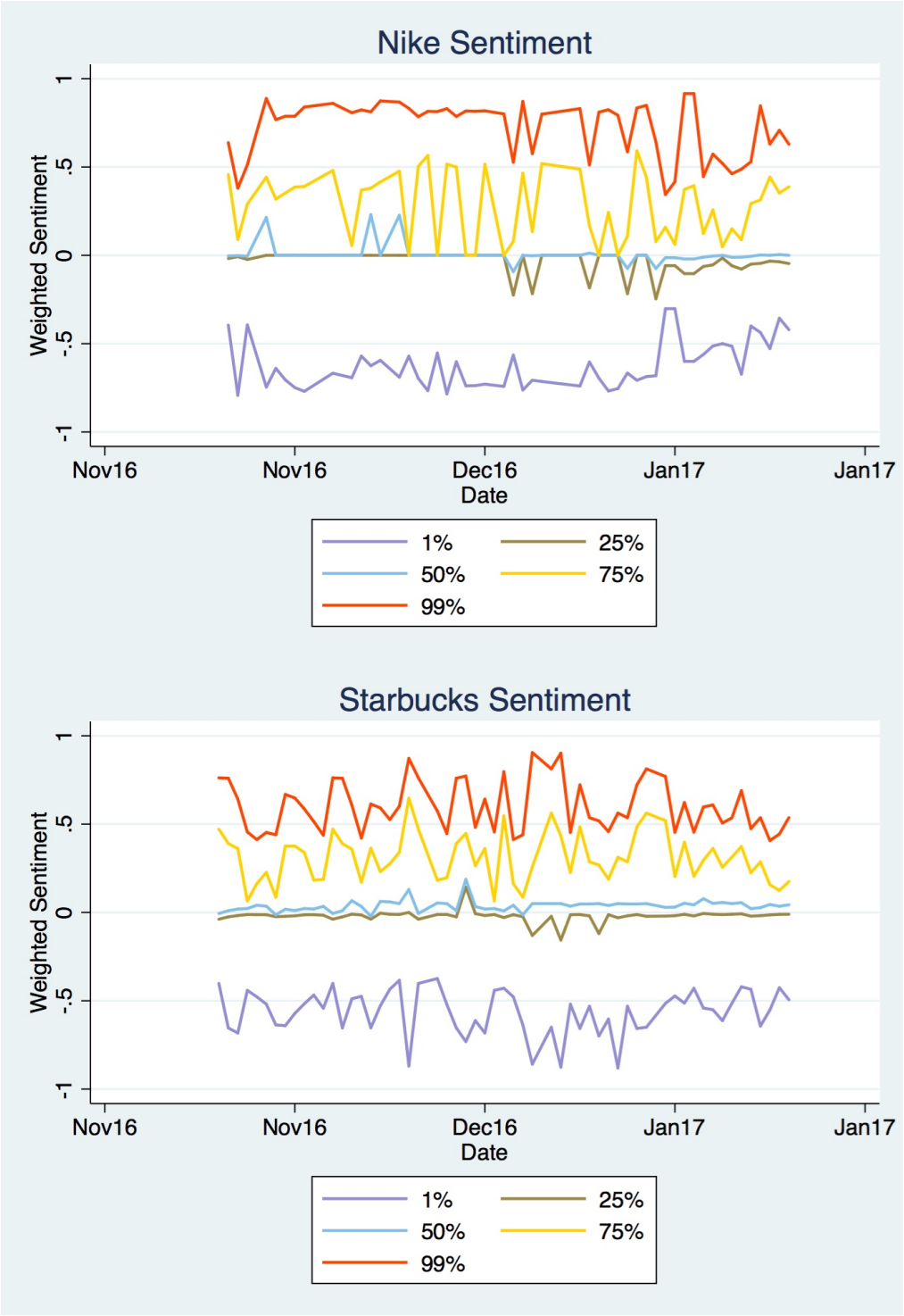

Exelon Sentiment
Sentiment Score
1
.5
0
-.5
-1
Nov2
Nov22
Dec12
Jan1
Jan21
Date
1%
25%
50%
75%
99%
Goldman Sachs Sentiment
Sentiment Score
.4
.2
0
-.2
-.4
Nov2
Nov22
Dec12
Jan1
Jan21
Date
1%
25%
50%
75%
99%



Customer-Facing Companies




Hershey Sentiment
Weighted Sentiment
1
.5
0
-.5
-1
Nov16
Nov16
Dec16
Jan17
Jan17
Date
1%
25%
50%
75%
99%
Michael Kors Sentiment
Sentiment Score
.5
0
-.5
-1
Nov2
Nov22
Dec12
Jan1
Jan21
Date
1%
25%
50%
75%
99%


Nike Sentiment
Weighted Sentiment
1
.5
0
-.5
-1
Nov16
Nov16
Dec16
Jan17
Jan17
Date
1%
25%
50%
75%
99%
Starbucks Sentiment
Weighted Sentiment
1
.5
0
-.5
-1
Nov16
Nov16
Dec16
Jan17
Jan17
Date
1%
25%
50%
75%
99%

Appendix B

Customer-Facing Companies

Coca Cola (mean: 1103 tweets per day) (sd: 634 tweets)

		Price weighted by volatility			Price wieghted by # of tweets			Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	1.0414699	3.7460298	0.92788362	1	2.1160002	5.2618099	4.2982521	1	0.5093831	4.1629368	0.6203275
5	0.8651465	4.3270018	0.5982953	5	1.7455889	4.7545472	4.2177477	5	0.3329977	4.7946939	0.1478355
10	0.7775865	4.7142341	0.34583657	10	1.6694539	4.2107537	3.4473216	10	0.3766824	5.1860525	-0.0934587
25	-0.1854471	0.2003792	0.19277271	25	1.1426789	1.2793855	1.1105975	25	0.0477422	-0.0629101	-0.063908
50	-3.4852414	-4.8402897	-0.16838824	50	-1.2120023	-2.1595489	-2.0150245	50	-3.5865571	-5.4380771	-0.092823
75	-0.9473833	-4.8393053	-0.27529061	75	-1.5959419	-4.1952585	-3.1398393	75	-0.5713815	-5.3261626	-0.0451628
90	-0.9662028	-4.1834692	-0.38551198	90	-1.712693	-5.1643587	-4.7537683	90	-0.5946743	-4.6730328	0.0261105
95	-1.114866	-3.9497661	-0.52067247	95	-1.7003411	-5.07106	-4.9512406	95	-0.6798102	-4.4201607	0.0243724
99	-1.100119	-3.6281592	-1.0333482	99	-2.0627881	-5.5599881	-4.6851807	99	-0.5891514	-4.1019251	-0.1514146
sd	-0.93	-4.39	-1.24	sd	-1.85	-5.19	-4.51	sd	-0.44	-4.91	-0.65

		Return weighted by volatility			Return wieghted by # of tweets			Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-4.0251753	-0.1428792	0.42416643	1	-2.1346309	1.0340863	0.5013731	1	-3.7665337	0.8484559	0.4628578
5	-3.9431297	0.2998346	0.1890167	5	-2.3145233	0.7078807	0.1763737	5	-3.7374136	1.2602616	0.1163838
10	-2.8087616	0.6786096	0.08994078	10	-1.9271253	0.3790406	0.2108253	10	-2.8457291	1.574038	-0.01004
25	2.0416521	0.0507371	-0.06000586	25	0.181036	-0.2723663	0.1884238	25	1.6968264	0.0863495	-0.1568411
50	-0.1724078	-1.829686	-0.04109772	50	1.0106426	-0.6581694	-0.5638249	50	-0.579136	-2.3793254	-0.0967511
75	2.6829579	-0.8777831	-0.04697584	75	2.2985164	-1.1822586	-0.0685088	75	2.6533099	-1.7239299	-0.0287295
90	2.6177532	-0.2109678	-0.06029189	90	2.2371383	-0.6374172	-0.1982137	90	2.5960895	-1.1781261	0.0204675
95	3.1503402	-0.0238761	-0.06263272	95	2.3580844	-0.4324759	-0.277515	95	2.9551529	-1.0103345	0.081556
99	3.7594531	0.1533754	0.22867897	99	2.2351998	-0.8979714	-0.6701004	99	3.47138	-0.8464855	0.6293227
sd	3.65	-0.35	-0.07	sd	2.37	-0.83	-0.43	sd	3.46	-1.34	-0.19

Hershey (mean: 2210 tweets per day) (sd: 1256 tweets)

		Price weighted by volatility			Price wieghted by # of tweets			Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-2.4382869	-0.6871055	1.0339592	1	1.1231341	1.3931902	1.5265871	1	-2.8578195	-1.5796615	1.2875404
5	-1.6413397	-1.0432097	0.70355053	5	1.2114808	1.45222	1.6535954	5	-1.9902124	-1.6541065	0.8361743
10	-1.2352691	-0.9624116	0.65061428	10	1.2010495	1.4749379	1.6801488	10	-1.5073545	-1.282605	0.711741
25	0.0092932	-0.2540863	0.22280132	25	0.9387582	1.4246352	1.3880358	25	-0.1556597	-0.3489469	0.1696189
50	-0.2178505	-0.9914448	-0.99952314	50	-0.8736011	-0.4294963	-0.8207522	50	-0.245947	-0.9531212	-0.8091678
75	0.7674699	0.4708875	-1.041498	75	-1.3402818	-1.5073603	-1.6711891	75	0.9284706	0.8407203	-0.9902013
90	0.7626881	1.1930695	-1.4126303	90	-1.4298976	-1.4864211	-1.6972661	90	0.833123	1.8742598	-1.364001
95	0.8312326	1.606604	-1.538271	95	-1.4556408	-1.4372901	-1.668862	95	0.851846	2.2704045	-1.535029
99	-0.0294029	0.8398742	-2.1048472	99	-1.404988	-1.3754014	-1.4894041	99	-0.498514	1.3104444	-2.3280873
sd	1.02	0.52	-1.61	sd	-1.27	-1.48	-1.61	sd	0.93	1.12	-1.67

		Return weighted by volatility			Return wieghted by # of tweets			Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-1.5732799	-2.5685321	2.2656565	1	0.9097013	0.4759402	-0.3436918	1	-1.9814238	-2.7968134	1.904433
5	-1.9212288	-2.2206485	1.5311118	5	0.7517893	0.628344	-0.417071	5	-2.1799561	-2.2714699	1.2333113
10	-1.6313249	-1.6048018	1.0958064	10	0.67024	0.6360325	-0.4387691	10	-1.8278783	-1.646542	0.8646349
25	-0.6124532	-0.0773438	0.28084571	25	0.2523642	0.7090518	-0.5173335	25	-0.7785781	-0.2938087	0.1835649
50	-0.8606521	-0.5012753	-1.3788578	50	-1.1404805	-0.3021124	-0.2991527	50	-0.7339649	-0.4535771	-1.0220879
75	0.4547254	0.9632207	-1.5250388	75	-0.9988644	-0.6904961	0.075774	75	0.7260391	1.1480758	-1.168151
90	-0.0012576	1.0484738	-1.935072	90	-0.8796288	-0.5992372	0.2512903	90	0.5444448	1.5477251	-1.5445518
95	-0.1232005	1.1966166	-1.9485585	95	-0.908453	-0.5324306	0.2560013	95	0.3851607	1.7783035	-1.6068437
99	-0.7457609	-0.308924	-2.1746612	99	-0.9609904	-0.4117112	0.2373613	99	-0.9931834	0.6207861	-2.0447944
sd	0.29	1.11	-2.19	sd	-0.82	-0.59	0.3	sd	0.5	1.61	-1.82

Michael Kors (mean: 264 tweets per day) (sd: 203 tweets)

		Price weighted by volatility			Price wieghted by # of tweets			Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	0.6374616	0.8692934	-0.71976804	1	0.9952723	0.1354964	0.0866675	1	1.0541249	-0.1163192	-0.8293452
5	0.0782504	1.3403003	-0.25865522	5	1.0806873	0.3783579	0.5249087	5	0.9160905	0.4325244	-0.359902
10	1.1117324	1.9063962	0.30707808	10	2.2923289	0.9158486	0.9306904	10	2.4190267	1.0076642	-0.0539322
25	1.1425393	2.2449431	0.52289507	25	2.2194668	1.6409	1.1664764	25	2.0286133	2.0979318	-0.0927747
50	0.4984656	0.12209	0.12983246	50	-0.2342328	0.5380373	0.4625681	50	0.2918099	0.6117206	-0.1694669
75	0.7404705	-0.0901397	-0.0329248	75	-0.1730772	0.4995213	0.3633925	75	0.5006902	0.4978662	-0.437539
90	0.795785	-0.0356684	0.21824359	90	-0.1458281	0.474656	0.4082041	90	0.4160043	0.4652499	-0.222389
95	1.1797729	-0.062182	0.36997279	95	0.049297	0.5165801	0.2916737	95	0.6949467	0.4632018	0.0074997
99	1.077592	1.9897255	1.0970509	99	-1.5594093	1.3093776	1.1176754	99	-0.905104	2.1724419	1.3016441
sd	0.59	0.6	-0.15	sd	-0.51	0.4	-0.04	sd	-0.67	0.73	0.16

		Return weighted by volatility			Return wieghted by # of tweets			Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	0.2454773	1.7717869	0.07821714	1	0.6078638	2.5954196	0.4436847	1	0.0274173	3.6299279	0.2422072
5	0.7098672	1.1516598	0.01066514	5	0.7209358	1.9929928	0.7641293	5	0.1845491	2.8506406	1.1421999
10	0.337492	0.6163226	0.27140468	10	0.5195429	1.4473012	1.1009236	10	-0.2285523	1.9793068	1.9110498
25	0.6961217	-0.7582541	0.9063697	25	-0.0734928	-0.998524	1.6635917	25	0.0377081	-1.433649	2.9590412
50	-0.4957524	-2.028817	0.31873088	50	-1.2008585	-2.9527686	0.2942726	50	-0.4953656	-3.9021862	1.5460606
75	-0.5992972	-2.1433883	0.08225462	75	-1.116836	-2.8773076	0.1580392	75	-0.4656696	-3.9896279	0.6785983
90	-0.6777588	-2.1652335	0.24115783	90	-0.9515134	-2.7364399	0.2500344	90	-0.4830448	-3.984049	0.6122583
95	-0.8170978	-1.9821295	0.14965977	95	-0.8918792	-2.4236438	0.1701153	95	-0.5762186	-3.8250567	0.490974
99	-2.5111762	-2.1455102	0.49548548	99	-1.2843274	-1.6200845	1.2015742	99	-1.0633566	-3.4596688	0.9944324
sd	-1.65	-1.45	0.66	sd	-1.13	-2.01	0.18	sd	-0.39	-3.37	0.87

Nike **(mean: 27622 tweets per day)** **(sd: 8914 tweets)**

		Price weighted by volatility			Price wieghted by # of tweets				Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-0.187307	-0.509586	0.0454749	1	6.4678533	4.5047062	2.8586865	1	-0.1904941	-0.5198691	0.0621371
5	0.2527193	-0.0993875	0.15322574	5	3.2405731	2.726583	2.3834219	5	0.2480438	-0.1115799	0.1617058
10	0.1852269	-0.1047985	0.09088851	10	2.1981186	1.6698461	1.0312841	10	0.1854357	-0.124801	0.1029653
25	-0.2573012	-0.5902407	-0.05654663	25	0.7090431	1.1079386	0.6792263	25	-0.2453933	-0.6669983	-0.0742269
50	-0.1185151	-0.0411776	-0.0000286	50	-0.1445444	-0.4685033	-0.7055119	50	-0.1322295	-0.038707	0.0112298
75	-0.4301118	-0.1809272	-0.12204225	75	-1.9347364	-2.473438	-2.2775774	75	-0.4349052	-0.1732616	-0.1038921
90	-0.4832237	-0.079203	-0.12260826	90	-6.1946105	-4.1222504	-2.7024592	90	-0.4953521	-0.0617129	-0.1051047
95	-0.5109507	0.0509944	-0.10166581	95	-6.3397534	-4.3674784	-2.8598148	95	-0.5340438	0.0746693	-0.0861561
99	-0.8962569	0.6268546	0.30615158	99	-6.7655051	-4.6269883	-2.9771466	99	-1.0039906	0.6717548	0.3295095
sd	-0.34	-0.05	-0.26	sd	-6.24	-4.4	-2.93	sd	-0.34	-0.07	-0.32

		Return weighted by volatility			Return wieghted by # of tweets				Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	0.0613504	0.2070672	0.58673613	1	1.1701391	1.5715929	1.730879	1	0.0650525	0.1590951	0.6775444
5	0.0517409	0.0632701	0.50386757	5	1.0094479	1.4883225	3.176774	5	0.0444672	0.060024	0.574981
10	0.0492016	0.0644236	0.43521762	10	1.1243028	1.4699357	1.8359169	10	0.0588579	0.0462247	0.4888507
25	0.0266178	0.0615989	0.23575342	25	0.074683	0.8275019	1.5173823	25	0.0366	0.0392467	0.2350201
50	0.0171421	0.0030508	-0.01195851	50	2.153648	0.9571307	0.0331745	50	0.0374024	0.0005576	0.0065188
75	-0.1194196	-0.1800226	-0.36206083	75	0.7295475	-0.8770733	-0.8996103	75	-0.1088671	-0.2015089	-0.3619017
90	-0.1748254	-0.1849603	-0.45235599	90	-1.1331183	-1.1483371	-1.0901626	90	-0.172211	-0.1872095	-0.4752504
95	-0.2043066	-0.1685073	-0.61720249	95	-1.0836796	-1.1960847	-1.1491593	95	-0.2077274	-0.1633475	-0.6514704
99	-0.5857677	-0.1669876	-1.3541335	99	-1.1321094	-1.2318291	-1.3005443	99	-0.6365493	-0.1529784	-1.4212897
sd	0.03	0.01	-0.99	sd	-1.18	-1.36	-1.61	sd	0.02	0.01	-1.12

Starbucks **(mean: 15374 tweets per day)** **(sd: 6547 tweets)**

		Price weighted by volatility			Price wieghted by # of tweets				Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-0.3698954	-0.2467288	1.1259973	1	-0.761746	-0.1578996	-0.1823649	1	-0.2731966	-0.1193297	1.9138456
5	-0.3467768	-0.2300524	1.3067308	5	-0.7232146	-0.1441565	-0.2082688	5	-0.2564201	-0.0845493	2.0941523
10	-0.3185388	-0.2269539	1.1999238	10	-0.683581	-0.1412863	-0.216875	10	-0.2310484	-0.0768189	1.9602574
25	0.1634691	-0.155584	0.83407152	25	0.0821112	-0.0812062	-0.2299285	25	0.2397062	-0.0409192	1.5069004
50	1.0544563	0.1790262	0.79003739	50	1.3946151	0.1313299	0.8763076	50	0.9634606	0.0749225	1.622412
75	0.604069	0.2486486	-0.06795818	75	1.0621738	0.1624603	0.248855	75	0.5186171	0.0896408	-0.2040374
90	0.4525303	0.2508505	-0.20875958	90	0.8508224	0.1625446	0.2441291	90	0.3659381	0.0916209	-0.6846703
95	0.4094759	0.2547948	-0.09536156	95	0.7924782	0.1653069	0.2455569	95	0.3223526	0.0969838	-0.6792653
99	0.4014788	0.2843209	0.54500659	99	0.7883469	0.1908575	0.2655713	99	0.3049499	0.1243565	-0.060706
sd	0.38	0.25	-0.1	sd	0.78	0.16	0.24	sd	0.29	0.1	-0.69

		Return weighted by volatility			Return wieghted by # of tweets				Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	0.9638617	0.5896077	0.23170154	1	0.9032301	0.4603715	0.193678	1	0.9974477	0.8595788	0.6989267
5	0.9304437	0.6380689	0.03538836	5	0.8708459	0.513917	0.1639224	5	0.9564259	0.9172486	0.3185743
10	0.9333837	0.668301	0.02772966	10	0.8756411	0.5489918	0.1622882	10	0.9581316	0.9488679	0.275876
25	0.9408225	0.9550491	0.00095083	25	0.8660834	0.9132614	0.1592105	25	0.9509536	1.1979343	0.1574137
50	-0.197892	0.2670821	-0.45364262	50	-0.2613623	0.3463468	-0.7154179	50	-0.2428923	0.2141371	-0.4315839
75	-0.8434545	-0.41899	-0.37937444	75	-0.765853	-0.2830496	-0.1676125	75	-0.8644631	-0.6853643	-0.5346734
90	-0.9061076	-0.5695121	-0.68225701	90	-0.8431008	-0.4401126	-0.1761699	90	-0.9342422	-0.8492998	-0.9394226
95	-0.9252553	-0.611589	-0.742212	95	-0.8629082	-0.484613	-0.1804956	95	-0.9544829	-0.8919908	-1.0171078
99	-0.9801511	-0.6382597	-0.73257847	99	-0.912553	-0.4997548	-0.1942802	99	-1.0162955	-0.9183329	-0.9422926
sd	-0.96	-0.64	-0.76	sd	-0.9	-0.5	-0.19	sd	-1	-0.92	-0.83

Non-Customer Facing Companies

Accenture **(mean: 441 tweets per day)** **(sd: 218 tweets)**

		Price weighted by volatility			Price wieghted by # of tweets				Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-1.6554776	-1.501662	0.10367976	1	-0.4816377	-0.0489468	0.56541674	1	-2.1582025	-1.9983661	0.47858593
5	-1.2776591	-1.2664121	-0.4507212	5	-0.2389525	0.31895419	0.61224908	5	-1.6216961	-1.5907218	0.23445161
10	-0.503794	-1.5153065	-0.2435101	10	0.40924321	0.7593898	0.52446212	10	-0.4816735	-1.7930143	0.3628455
25	0.16464787	0.21191786	0.00378499	25	0.13890025	-0.0067355	-0.3955832	25	0.12578233	0.19926853	0.10011061
50	0.39059824	0.40752014	0.11987488	50	0.22972247	-0.0138153	-0.4533952	50	0.26407862	0.39416436	0.10649701
75	0.69568667	0.78297809	0.24624912	75	0.23338748	0.03100907	-0.4782442	75	0.68606427	0.86379169	0.20094308
90	0.84616782	0.90285861	0.2547523	90	0.27956212	0.07094538	-0.4292714	90	0.88674652	0.98605241	0.04920366
95	0.9692484	0.95949836	0.40678542	95	0.33189705	0.12214009	-0.3789705	95	1.1656612	1.0701014	0.19944847
99	2.0830573	2.0783415	1.3222317	99	0.70262146	0.71482812	-0.1666568	99	3.0560143	2.4019625	0.48269893
sd	1.67	1.58	0.67	sd	0.39	0.14	-0.43	sd	2.12	1.78	0.11

		Return weighted by volatility			Return wieghted by # of tweets				Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-0.3049224	-0.5847541	1.4142587	1	-0.8945907	-0.7193188	0.04618527	1	-0.286823	-0.8840888	2.0066491
5	0.48760055	-0.3498044	1.001712	5	-0.855858	-0.2061215	0.23898252	5	0.83896687	-0.6045526	1.4729349
10	1.2855033	-0.6479983	0.76498994	10	-0.6093393	0.68872578	0.79864199	10	1.9625358	-0.9735803	1.0510044
25	0.20217801	0.29859774	0.04114528	25	0.35932998	1.12126	0.61178229	25	0.10514345	0.24320649	0.08691507
50	-0.1514488	0.18850277	-0.0866579	50	0.85071737	0.66777314	0.23954271	50	-0.2580572	0.18359127	-0.1805825
75	-0.0317259	0.25922657	-0.5012861	75	0.85071216	0.63334967	0.15173353	75	-0.1438534	0.2491621	-0.7013566
90	0.03476205	0.27072887	-0.8933969	90	0.88359069	0.60208915	0.10717885	90	-0.0307098	0.24612708	-1.2023689
95	0.21712507	0.35427439	-1.1291608	95	0.83511334	0.65431854	0.13261355	95	0.24234988	0.35349127	-1.5391007
99	-0.4705475	0.53131395	-2.97104	99	0.55787291	0.80917897	-0.1220699	99	-0.4243763	0.39083072	-3.0842807
sd	0.06	0.3	-1.69	sd	0.78	0.59	-0.06	sd	0.12	0.27	-2.29

Ever Source (mean: 100 tweets per day) (sd: 63 tweets)

	Price weighted by volatility				Price wieghted by # of tweets				Price weighted by volatility and # of tweets			
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	1.7041012	1.4810361	0.18383219	1	0.83517849	1.0120508	0.2784443	1	1.68281	1.4941899	0.44544048
5	-2.4235015	-2.305028	-1.596205	5	-2.1397828	-2.0173826	-1.3124824	5	-2.2268776	-1.5688852	-1.1185872
10	-1.5428598	-1.4536316	-1.3575507	10	-1.3246963	-1.8837806	-1.6065975	10	-1.5963959	-1.2235404	-0.9208919
25	-1.0147541	-1.1223191	-1.1840847	25	-0.7662252	-1.971445	-1.3362749	25	-0.6783478	-0.923309	-1.0825748
50	1.5731818	0.47991904	0.16407835	50	1.5982476	0.81433159	0.14929661	50	1.4741184	0.38107441	0.14855447
75	1.7910659	1.2863162	0.67788694	75	1.4225743	1.1018173	0.48925376	75	2.0462621	1.3667455	0.52078572
90	1.6722796	0.50850188	1.0851575	90	1.4763609	1.3262718	0.27292499	90	1.9173482	0.40595072	0.91649004
95	2.1237028	0.32107571	0.77922918	95	1.915137	1.1407795	0.15910552	95	2.514286	0.56999687	0.95441643
99	0.86183497	-0.5066586	1.5990022	99	1.3066754	0.06116278	1.5706991	99	1.3046668	-0.7476743	2.0253881
sd	0.89	0.25	1.3	sd	1.38	0.61	1.14	sd	1.09	-0.25	1.4

	Return weighted by volatility				Return wieghted by # of tweets				Return weighted by volatility and # of tweets			
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	0.48352951	3.1136625	-1.0487807	1	1.3535093	1.7866183	1.2965464	1	-1.7621353	-0.9543557	0.72093527
5	0.88077455	5.0570982	0.08638385	5	1.3939816	2.992646	2.4156959	5	0.43599474	-0.4272651	0.22539733
10	1.6611677	4.9271696	-0.1107514	10	1.309771	2.789431	2.2513315	10	1.3573898	0.6084657	0.30833015
25	1.4466635	2.9150918	0.65861866	25	1.2088714	2.4119953	1.8917218	25	-0.3865632	0.83793364	0.33809868
50	0.40741511	1.349193	0.70441609	50	1.1469689	2.1658921	1.656101	50	-0.0730104	0.27682064	-0.185064
75	1.5664099	1.2239211	1.1556962	75	0.87594689	1.7425405	1.403828	75	-0.2653329	0.59007438	0.08730581
90	-0.8350767	-0.8020142	0.54145214	90	-0.4995632	-1.5391553	-1.2009048	90	-1.0868304	-1.3660322	0.97317139
95	-0.3006473	-1.6696495	1.8635742	95	-0.3620528	-1.474264	-0.8053646	95	-0.2046781	-0.1439242	1.0220128
99	0.56342041	-2.277318	0.55769127	99	-0.8192981	-1.7452665	-0.6140077	99	-0.6010481	-1.1426152	0.7880327
sd	0.14	-0.3	0.58	sd	-0.36	-0.35	0.86	sd	-0.27	-0.27	0.74

Exelon (mean: 136 tweets per day) (sd: 141 tweets)

	Price weighted by volatility				Price wieghted by # of tweets				Price weighted by volatility and # of tweets			
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	1.7041012	1.4810361	0.18383219	1	0.83517849	1.0120508	0.2784443	1	1.5348287	5.1337474	-1.4729164
5	-2.4235015	-2.305028	-1.596205	5	-2.1397828	-2.0173826	-1.3124824	5	2.1554945	6.5554187	-0.1075305
10	-1.5428598	-1.4536316	-1.3575507	10	-1.3246963	-1.8837806	-1.6065975	10	2.3192514	5.9195458	-0.1228505
25	-1.0147541	-1.1223191	-1.1840847	25	-0.7662252	-1.971445	-1.3362749	25	2.2787661	2.4772806	0.37667534
50	1.5731818	0.47991904	0.16407835	50	1.5982476	0.81433159	0.14929661	50	0.48558567	2.3656203	0.68263019
75	1.7910659	1.2863162	0.67788694	75	1.4225743	1.1018173	0.48925376	75	1.3287973	1.9071354	1.8742874
90	1.6722796	0.50850188	1.0851575	90	1.4763609	1.3262718	0.27292499	90	-1.9356487	-1.1193043	1.2018921
95	2.1237028	0.32107571	0.77922918	95	1.915137	1.1407795	0.15910552	95	-1.7264322	-2.6621593	2.085613
99	0.86183497	-0.5066586	1.5990022	99	1.3066754	0.06116278	1.5706991	99	-1.5065351	-4.5227106	1.0773868
sd	-0.5	-3.02	0.96	sd	-1	-2.29	-1.51	sd	-1.96	-4.69	1.45

	Return weighted by volatility				Return wieghted by # of tweets				Return weighted by volatility and # of tweets			
Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0	Percentile	Day2	Day1	Day0
1	-1.0693987	-4.3411371	2.0568187	1	-1.7645244	0.0458489	5.2548819	1	-2.3163272	-4.9877896	3.3574849
5	-1.8683102	-5.6097257	0.50809409	5	-2.4039629	-0.377577	5.4262353	5	-3.463294	-6.5134535	2.3117208
10	-1.3590027	-4.6863689	0.36660971	10	-2.4304077	0.31458529	5.9570524	10	-3.3717638	-4.2515603	2.4060098
25	-1.408209	-1.8261259	0.76015058	25	-2.577039	1.2221563	6.1313964	25	-3.3999162	-0.2933755	1.7282627
50	-0.291726	0.03782748	-1.1190237	50	-0.2581778	0.03641083	1.1500148	50	-0.4437517	-0.4162981	0.74585601
75	-2.6725222	-0.706879	-2.2721141	75	0.12473411	0.1736077	0.01498807	75	-1.8804337	-0.6752981	-2.3435019
90	0.19576841	0.78794293	-2.4391614	90	2.7250975	-1.3355278	-5.7186663	90	3.0301225	-0.4109914	-2.935456
95	-0.2799096	2.931789	-2.4917937	95	2.7920924	-0.9333323	-5.2335477	95	2.8558453	1.8546003	-3.3344778
99	-1.7071694	4.8466725	-1.4633301	99	2.0680452	-0.1407042	-4.6692217	99	2.0461322	5.0777155	-2.716849
sd	0.63	3.62	-2.45	sd	2.59	-0.57	-6.57	sd	3.04	3.68	-4.34

Goldman Sachs (mean: 743 tweets per day) (sd: 678 tweets)

	Price weighted by volatility			Price wieghted by # of tweets			Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Day2	Day1	Day0	Day2	Day1	Day0
1	1.9830494	1.9428896	0.99797218	1.1044626	0.51217878	0.94006699	1.2961362	0.82116131	2.5504862
5	2.2047117	1.0808291	1.0262071	1.5965226	0.21582384	0.70752137	2.0150513	0.24717695	2.3863318
10	2.9940472	1.7173313	1.4895048	2.4561969	0.92103079	1.187539	3.0554243	0.80543862	2.603598
25	2.9617706	1.1923967	0.96225038	2.6203523	0.75058617	0.6015307	2.7588318	0.18941927	1.706678
50	-2.1970663	-2.3817277	-2.9087021	-1.9331654	-1.2585161	-3.2349179	-2.4452997	-1.0342096	-3.7178824
75	-2.5075575	-2.4149669	-2.3441912	-1.9108765	-0.9391045	-2.1531567	-2.8755697	-1.3784166	-3.515918
90	-2.373909	-2.6295101	-2.2565205	-1.9106868	-1.006725	-2.0717636	-2.5851011	-1.487928	-3.6550209
95	-2.693832	-3.0013314	-2.6161679	-2.1104659	-1.4161086	-2.4863998	-2.6043523	-1.8675289	-3.9891981
99	-4.2686619	-2.353073	-4.049012	-2.0621396	-2.1149657	-4.1118838	-2.4749793	-1.7015916	-4.9317046
sd	-3.06	-2	-2.42	-1.71	-1.53	-2.36	-3.01	-2.06	-3.87

	Return weighted by volatility			Return wieghted by # of tweets			Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Day2	Day1	Day0	Day2	Day1	Day0
1	-0.2154021	-0.5695561	-0.2808585	0.75017795	0.6447174	0.44960358	0.4022299	0.48082619	-1.3762041
5	-0.1519312	-0.6058613	-0.2262776	0.46333176	0.47559954	0.35985189	0.14028999	0.17519749	-1.2751504
10	-0.3185395	-0.4942512	-0.238836	-0.0198022	0.51234657	0.44399627	-0.293428	0.29066278	-1.3036229
25	-0.2566873	0.00945869	-0.1767687	-0.2007694	0.84559199	0.41250761	-0.3995916	1.016854	-0.9047944
50	-0.8000202	1.7198026	1.2337162	0.04485982	0.3394559	-0.1416026	-0.4386776	1.2892353	1.8713474
75	-0.5914899	1.4336483	0.57895821	-0.3113593	0.08591858	-0.4582649	-0.4680591	1.04826	1.4960564
90	-0.4704255	1.5277839	0.53095766	-0.3991468	0.14694569	-0.3255084	-0.4873042	1.1054109	1.5925579
95	-0.1728448	1.5514093	0.73341916	-0.1685578	0.34204615	-0.2323026	-0.2506832	1.1145468	1.7938071
99	0.94461809	2.2808963	0.78658614	0.58899022	1.8528154	0.15929509	0.35813814	2.5961462	1.5132055
sd	0.25	1.4	1.07	0.24	1.16	0.07	0.08	1.38	1.94

Morgan Stanley (mean: 143 tweets per day) (sd: 104 tweets)

	Price weighted by volatility			Price wieghted by # of tweets			Price weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Day2	Day1	Day0	Day2	Day1	Day0
1	-0.1844457	-0.9279196	-1.1547405	-0.349062	-0.3706884	-0.1603735	-0.1971849	-0.4959054	-1.0795858
5	-0.5643501	-0.2456685	-1.3436117	-0.790789	-0.8019345	-0.6483684	-0.7304202	-0.6597989	-1.0631035
10	-0.4132891	-0.3221229	-1.8821768	-0.8624595	-0.3260079	-1.1617074	-0.5255616	-0.4309695	-1.8351734
25	-0.6034893	-0.5231085	-2.7459034	-0.8830611	-0.8444575	-1.6676825	-1.2491115	-0.2472028	-2.5216199
50	-0.3677581	0.18726325	-1.6079069	-0.286611	-0.1488169	-1.1892485	-0.5890021	0.25741544	-1.616359
75	0.22226011	0.69268458	0.73193139	0.21622572	0.2757219	0.04509695	0.37738548	0.4920129	0.5887698
90	-0.0591243	-0.1747311	0.92983877	0.06740783	0.17606868	-0.3049949	0.12162546	-0.0076006	0.84906268
95	-0.2063585	-1.1035794	0.40638289	-0.1609124	-0.2010069	-0.6594252	0.08392647	-0.6570905	0.4690768
99	-0.2970661	-0.7673522	-0.1683158	-0.732409	-0.6857679	-0.8938958	-0.3182743	-0.8432155	0.79361241
sd	-0.96	-0.83	-0.46	-0.15	-0.62	-1.79	-0.59	-0.76	-0.42

	Return weighted by volatility			Return wieghted by # of tweets			Return weighted by volatility and # of tweets		
Percentile	Day2	Day1	Day0	Day2	Day1	Day0	Day2	Day1	Day0
1	-0.3990025	-1.3011252	-0.7111428	0.01295546	-1.1322702	-0.3037455	-1.3166076	-1.0036031	-0.4840201
5	-0.2900251	-0.3224353	-0.1409902	0.22237823	-0.4326705	1.2251658	-0.6914826	-0.2017823	0.13962739
10	-0.4826833	-0.101658	-0.1354952	0.28833582	-0.3301357	1.0908552	-1.0501796	0.0043435	0.120938
25	0.82582159	-0.4449192	-0.7189843	1.1274707	1.0064205	-0.5180973	0.77249069	0.11589257	-0.6878727
50	0.90565377	-0.5399458	-1.2279095	0.82658336	1.0401864	-1.8406361	1.2909046	0.35799947	-1.4494533
75	0.08481825	-0.1963041	-0.599502	-0.0972678	0.57416232	-1.0283034	0.04295565	0.71628519	-1.1189756
90	-0.0045504	0.77722536	-0.4489866	-0.4914629	1.0888537	-0.6339511	0.04325666	1.394794	-0.7625275
95	-0.4868222	0.67597271	-0.739527	-1.0987834	0.65114613	-0.9733988	-0.7559888	1.1513228	-0.8047815
99	0.69678136	0.72258738	-0.3948128	0.27480289	1.0670658	0.5766205	0.97495042	0.89126885	-0.3234247
sd	1.69	1.53	0.11	-0.03	1.69	0.35	0.95	1.94	0.31

Appendix C

Python Code

Code1: Twitter Data Acquisition

```
# Import the necessary package to process data in JSON format
try:
    import json
except ImportError:
    import simplejson as json

from twitter import Twitter, OAuth, TwitterHTTPError, TwitterStream

#!/usr/bin/python
import sys

#print 'Number of arguments:', len(sys.argv), 'arguments.'


ACCESS_TOKEN = '609610288-2a5WhAbYUkFeR3IWP6eSrsFFiles70YwUErrp4ea'
ACCESS_SECRET = 'OmDmdvCj8NXlRrg2iiMowgVP1dhzRbCGwYCluC46yBJYB'
CONSUMER_KEY = 'PuQvKjcGlbCN3H5W2X4nGgXtV'
CONSUMER_SECRET = 'n19eq6oaxsgJQt6z7PO22fdOuVJXqFr8rlndi2yQCXaYvZGnq9'


oauth = OAuth(ACCESS_TOKEN, ACCESS_SECRET, CONSUMER_KEY, CONSUMER_SECRET)
twitter_stream = TwitterStream(auth=oauth)


trackName = sys.argv[1] + ", " + sys.argv[2]
#print trackName
iterator = twitter_stream.statuses.filter(track=trackName, language="en")
#twitter_data = open('twitter %s.txt' %(trackName), 'a')

for tweet in iterator:

    print json.dumps(tweet)
#twitter_data.write(json.dumps(tweet))
```

Code 2: Shell Script to run parallel programs for consecutive days

```
#!/bin/bash
collect1c() {
    python twitter_streaming_cus1.py $*
}
collect2c() {
    python twitter_streaming_cus2.py $*
}
collect3c() {
    python twitter_streaming_cus3.py $*
}
collect1n() {
    python twitter_streaming_noncus1.py $*
}
collect2n() {
    python twitter_streaming_noncus2.py $*
}
collect3n() {
    python twitter_streaming_noncus3.py $*
}

delay=1200
logs=$PWD

company1c=MichaelKors
stock1c=\$KORS
company2c=Hershey
stock2c=\$HSY
company3c=Starbucks
stock3c=\$SBUX
company4c=Nike
stock4c=\$NKE
company5c=CocaCola
stock5c=\$KO
company1n=Accenture
stock1n=\$ACN
company2n=GoldmanSachs
stock2n=\$GS
company3n=Exelon
stock3n=\$EXC
company4n=Eversource
stock4n=\$ES
company5n=MorganStanley
stock5n=\$MS
```

```
44
45  # 60 days
46  # for (( i=1;i<=120;i++ )); do
47  for (( i=1;i<=2;i++ )); do
48      dt=$(date "+%y%m%d%H%M%S")
49      collect2c $company1c $stock1c > "$logs/$company1c$dt.txt" &
50      childprocess1=$!
51      collect2c $company2c $stock2c > "$logs/$company2c$dt.txt" &
52      childprocess2=$!
53      collect3c $company3c $stock3c > "$logs/$company3c$dt.txt" &
54      childprocess3=$!
55      collect3c $company4c $stock4c > "$logs/$company4c$dt.txt" &
56      childprocess4=$!
57      collect1n $company5c $stock5c > "$logs/$company5c$dt.txt" &
58      childprocess5=$!
59      collect1n $company1n $stock1n > "$logs/$company1n$dt.txt" &
60      childprocess1n=$!
61      collect2n $company2n $stock2n > "$logs/$company2n$dt.txt" &
62      childprocess2n=$!
63      collect2n $company3n $stock3n > "$logs/$company3n$dt.txt" &
64      childprocess3n=$!
65      collect3n $company4n $stock4n > "$logs/$company4n$dt.txt" &
66      childprocess4n=$!
67      collect3n $company5n $stock5n > "$logs/$company5n$dt.txt" &
68      childprocess5n=$!
69
70      sleep $delay
71      kill $childprocess1
72      kill $childprocess2
73      kill $childprocess3
74      kill $childprocess4
75      kill $childprocess5
76      kill $childprocess1n
77      kill $childprocess2n
78      kill $childprocess3n
79      kill $childprocess4n
80      kill $childprocess5n
81  done
```

Code 3: Sentiment Acquisition

```
try:
    import json
except ImportError:
    import simplejson as json
import csv
import httplib, urllib, base64
import unicodedata

#Please use your own API key when replicating the experiment
api_key1 = 'b8ec824e3b00343d78b9062e1c3da891a4db52a6'

headers = {
    # Request headers
    'Content-Type': 'application/json',
    'Ocp-Apim-Subscription-Key': api_key1,
}

class myData(object):
    text = ""
    date = ""
    keyword = ""
    sentiment_score = 0

    def __init__(self, text, date, keyword):
        self.text = text
        self.date = date
        self.keyword = keyword

def create_myData(text, date, keyword):
    mydata = myData(text, date, keyword)
    return mydata

# We use the file saved from last step as example
tweets_filename = 'Starbucks161031110859.txt'
tweets_file = open(tweets_filename, "r")

outputFile = open('Starbucks161031.csv', 'a')
outputWriter = csv.writer(outputFile)

```

```
40
41  for line in tweets_file:
42      # print ("ee")
43      try:
44          # Read in one line of the file, convert it into a json object
45          tweet = json.loads(line.strip())
46
47          if 'text' in tweet:
48
49              # clean up the data, break and combine
50              textString = tweet['text']
51              textArray = textString.split(" ")
52              count = 0
53              followers = tweet['user']['followers_count']
54              for word in textArray:
55                  if word[:4] == "http":
56                      del textArray[count]
57                      #print ("deleted")
58                  count = count + 1
59
60              textSt = unicodedata.normalize("NFKD", ' '.join(textArray)).encode('ascii','ignore')
61              params = urllib.urlencode({textSt})
62              conn = httplib.HTTPSConnection('westus.api.cognitive.microsoft.com')
63              conn.request("POST", "/text/analytics/v2.0/target=Starbucks/sentiment?%s" % params, "{body}", headers)
64              response = conn.getresponse()
65              data = response.read()
66
67              #jsonStr = json.dumps(alchemy_language.targeted_sentiment(text=" ".join(textArray), targets=['$ES']), indent=2)
68
69              print (jsonStr)
70              mydata = create_myData(textSt, "2016/10/31", "Starbucks")
71              #print (jsonStr)
72              jsonObj = json.loads(jsonStr)
73              if jsonObj['docSentiment']['type'] == "neutral":
74                  mydata.sentiment_score = 0;
75              else:
76                  mydata.sentiment_score = jsonObj['docSentiment']['score']
77              # print (type(str(mydata.text))
78              outputWriter.writerow([mydata.date, mydata.keyword, str(mydata.sentiment_score), mydata.text, str(followers)])
79              print ("added")
80              conn.close()
81      except:
82          print ('wrong')
83          # read in a line is not in JSON format (sometimes error occured)
84          continue
85
86  outputFile.close()
```

Stata Code

Code 4: Merge data and create weighted sentiment variables

```
clear all

cd /Users/daiyuhui/Desktop/CapProcess/done/Starbucks
local files : dir . files "*.csv"


foreach file in `files' {
    import delimited `file', bindquote(strict) varnames(nonames) clear
    /* retrieve part of the file name as date variable */
    capture drop if v3 == .

    if (_rc !=0){
        dir `file'
        }

    /* Put the code for generating and using the weights */
    capture gen srfoll = sqrt(v5)
    capture egen maxfoll = max(srfoll)
    capture gen weight = srfoll/maxfoll

    if (_rc != 0){
        gen weight = 1
        }

    /* Finally, compute the statistics/distribution.mean and variance */
    gen wsent = weight * v3
    egen mean_sent = mean(wsent)
    gen sq_dev = (wsent-mean_sent)^2

    /* add a variable obs that counts how many obserations you have each day */
    collapse (p1) p1=wsent (p5) p5=wsent (p10) p10=wsent (p25) p25=wsent (p50)
    p50=wsent (p75) p75=wsent (p90) p90=wsent (p95) p95=wsent (p99) p99=wsent
    (mean) wav=wsent (mean) vari=sq_dev (count) num_tweet = wsent, by(v1 v2)
    rename (v1 v2) (date company)
    save "`file'.dta", replace
}

clear all

set obs 1
gen date = ""
gen company = ""
gen p1 = .
gen p5 = .
gen p10 = .
gen p25 = .
gen p50 = .
gen p75 = .
gen p90 = .
gen p95 = .
gen p99 = .
gen wav = .
gen vari = .
gen num_tweet = .
local files : dir . files "*.dta"

foreach file in `files'{
    append using `file'
    }

drop if _n == 1
drop if p1 == .

gen IQR = p75 - p25
gen I10_90 = p90 - p10
gen newdate = date(date,"YMD")
drop if newdate == .
collapse p1-I10_90, by(newdate company)
tset newdate

save "Starbucks.dta", replace
```

Code 5: Merge with Stock Prices, Regression Analysis & File Writing

```
import excel "/Users/daiyuhui/Desktop/capstone/stocks.xlsx", sheet("return")
firstrow clear

local stocklist KORS HSY SBUX NKE KO ACN GS EXC ES MS R_KORS R_HSY R_SBUX R_NKE
R_KO R_ACN R_GS R_EXC R_ES R_MS

gen daten = A
tsset daten, d

// convert stock data from string to real
//foreach x of varlist `stocklist' {
// gen  n_`x' = real(`x')
// drop `x'
// gen `x' = n_`x'
//}

keep daten `stocklist'

sort daten

save stockdata, replace

/* 2. Load the sentiment data */
cd /Users/daiyuhui/Desktop/CapProcess/done/CocaCola
use "CocaCola", clear
cap gen daten = newdate
drop newdate
tsset daten, d
sort daten

/* Merge the two datasets */
merge 1:1 daten using stockdata

/* Play with the combined data */
sort daten

//tsline EXC p10

 gen var_w = 1/vari
 egen max_tweet = max(num_tweet)
 gen num_w = num_tweet/max_tweet
 gen comb_w = num_tweet/vari

//regression with new weight
file open myfile using "Starbucks.csv", write replace

file write myfile  "Percentile,Day1,Day2,Day0 "  _n
foreach i of numlist 1 5 10 25 50 75 90 95 99 {
 file write myfile  "`i',"
 regress R_SBUX L1.p`i' [aw = comb_w]
 local tstat : disp _b[L1.p`i'] / _se[L1.p`i']
 file write myfile  (`tstat') ","
 regress R_SBUX L2.p`i'[aw = comb_w]
 local tstat : disp _b[L2.p`i'] / _se[L2.p`i']
 file write myfile (`tstat') ","
 regress R_SBUX L0.p`i'[aw = comb_w]
 local tstat : disp _b[L0.p`i'] / _se[L0.p`i']
 file write myfile  (`tstat') "," _n
}
file close myfile
```

Notes*: Comment Areas is optional depending on the whether the processing files satisfy specific requirement. For instance, if data are written in string, converting to real number is important. Thus, user needs to comment out the string to real number conversion in Code 5.

Bibliography

AlchemyLanguageAPI. "Sentiment Analysis." June 22, 2016. Accessed November 14, 2016. http://www.alchemyapi.com/products/alchemylanguage/sentiment-analysis.

Antweiler, Werner, and Murray Z. Frank. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance* 59, no. 3 (2004): 1259-294.

Baker, Malcolm, and Jeffrey Wurgler. "Investor Sentiment and the Cross-Section of Stock Returns." *The Journal of Finance* 61, no. 4 (2006): 1645-680.

Baker, Malcolm, and Jeremy C. Stein. "Market Liquidity as a Sentiment Indicator." *Journal of Financial Markets* 7, no. 3 (2004): 271-99.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2, no. 1 (2011): 1-8.

Brown, Gregory W., and Michael T. Cliff. "Investor Sentiment and the Near-term Stock Market." *Journal of Empirical Finance* 11, no. 1 (2004): 1-27.

CanbaÅ, Serpil, and Serkan YÄlmaz KandÄr. "Investor Sentiment and Stock Returns: Evidence from Turkey." *Emerging Markets Finance and Trade* 45, no. 4 (2009): 36-52.

Chung, San-Lin, Chi-Hsiou Hung, and Chung-Ying Yeh. "When Does Investor Sentiment Predict Stock Returns?" *Journal of Empirical Finance* 19, no. 2 (2012): 217-40.

Corea, Francesco. "Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector." *Big Data Research* 4 (2016): 70-74.

Fisher, Kenneth L., and Meir Statman. "Customer Confidence and Stock Returns." *The Journal of Portfolio Management* 30, no. 1 (2003): 115-27.

Garcia, Diego. "Sentiment during Recessions." *The Journal of Finance* 68, no. 3 (2013): 1267-300.

Groß-Klußmann, Axel, and Nikolaus Hautsch. "When Machines Read the News: Using Automated Text Analytics to Quantify High Frequency News-implied Market Reactions." *Journal of Empirical Finance* 18, no. 2 (2011): 321-40.

He, Wu, Lin Guo, Jiancheng Shen, and Vasudeva Akula. "Social Media-Based Forecasting:." *Journal of Organizational and End User Computing* 28, no. 2 (2016): 74-91.

Hudson, Yawen, and Christopher J. Green. "Is Investor Sentiment Contagious? International Sentiment and UK Equity Returns." *Journal of Behavioral and Experimental Finance* 5 (2015): 46-59.

Jiang, Yumei, and Mingzhao Wang. "Investor Sentiment and the Near-Term Stock Returns: Evidence from Chinese Stock Market." *2009 International Conference on Management and Service Science*, 2009.

JSON. "Introducing JSON." Accessed November 16, 2016. http://www.json.org/.

Kadilli, Anjeza. "Predictability of Stock Returns of Financial Companies and the Role of Investor Sentiment: A Multi-country Analysis." *Journal of Financial Stability* 21 (2015): 26-45.

Karabulut, Yigitcan. "Can Facebook Predict Stock Market Activity?" *SSRN Electronic Journal*.

Kim, Soon-Ho, and Dongcheol Kim. "Investor Sentiment from Internet Message Postings and the Predictability of Stock Returns." *Journal of Economic Behavior & Organization* 107 (2014): 708-29.

Kim, Minhyuk, and Jinwoo Park. "Individual Investor Sentiment and Stock Returns: Evidence from the Korean Stock Market." *Emerging Markets Finance and Trade* 51, no. Sup5 (2015)

Kumari, Jyoti, and Jitendra Mahakud. "Does Investor Sentiment Predict the

Asset Volatility? Evidence from Emerging Stock Market India." *Journal of Behavioral and Experimental Finance* 8 (2015): 25-39.

Lemmon, Michael, and Evgenia Portniaguina. "Customer Confidence and Asset Prices: Some Empirical Evidence." *Rev. Financ. Stud. Review of Financial Studies* 19, no. 4 (2006): 1499-529.

Lohr, Steve. "The Age of Big Data." The New York Times. February 11, 2012. Accessed December 07, 2016. http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

Lux, Thomas. "Sentiment Dynamics and Stock Returns: The Case of the German Stock Market." *Empirical Economics* 41, no. 3 (2010): 663-679

Malhotra, A., and C. Malhotra. "How to Get Your Messages Retweeted." MIT Sloan Management Review RSS. 2012. Accessed November 14, 2016.

Malkiel, Burton Gordon. *A Random Walk down Wall Street: The Time-tested Strategy for Successful Investing*. New York: W. W. Norton, 2003. Print.

Mao, Huina, Scott Counts, and Johan Bollen. "Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data." Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data. December 5, 2011.

Mian, G. Mujtaba, and Srinivasan Sankaraguruswamy. "Investor Sentiment and Stock Market Response to Earnings News." *The Accounting Review* 87, no. 4 (2012): 1357-384.

OpinionFinder | MPQA. "OpinionFinder | MPQA." Accessed November 14, 2016. http://mpqa.cs.pitt.edu/opinionfinder/.

Ranco, Gabriele, Darko Aleksovski, Guido Caldarelli, Miha GrÄar, and Igor MozetiÄ. "The Effects of Twitter Sentiment on Stock Price Returns." *PLOS ONE PLoS ONE* 10, no. 9 (2015).

Root. "Customer Sentiment." Investopedia. October 25, 2010. Accessed November 18, 2016. http://www.investopedia.com/terms/c/customer-

sentiment.asp.

Sabherwal, Sanjiv, Salil K. Sarkar, and Ying Zhang. "Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News." *Journal of Business Finance & Accounting* 38, no. 9-10 (2011): 1209-237.

Schmeling, Maik. "Investor Sentiment and Stock Returns: Some International Evidence." *Journal of Empirical Finance* 16, no. 3 (2009): 394-408.

Simon, David P., and Roy A. Wiggins. "S&P Futures Returns and Contrary Sentiment Indicators." *Journal of Futures Markets* 21, no. 5 (2001): 447-62.

Smales, Lee A. "Time-variation in the Impact of News Sentiment." *International Review of Financial Analysis* 37 (2015): 40-50.

Smales, Lee A. "News Sentiment in the Gold Futures Market." *Journal of Banking & Finance*49 (2014): 275-86.

Sprenger, Timm O., Philipp G. Sandner, Andranik Tumasjan, and Isabell M. Welpe. "News or Noise? Using Twitter to Identify and Understand Company-specific News Flow." *Journal of Business Finance & Accounting* 41, no. 7-8 (2014): 791-830.

Spyrou, Spyros. "Sentiment Changes, Stock Returns and Volatility: Evidence from NYSE, AMEX and NASDAQ Stocks." *Applied Financial Economics* 22, no. 19 (2012): 1631-646.

Twitter Company. "Twitter Usage/Company Facts." Accessed November 14, 2016. https://about.twitter.com/company.

Twitter Developer Documentation. "API Overview." Accessed November 14, 2016. https://dev.twitter.com/overview/api.

Wang, Yaw-Huei, Aneel Keswani, and Stephen J. Taylor. "The Relationships between Sentiment, Returns and Volatility." *International Journal of Forecasting* 22, no. 1 (2006): 109-23.

Yu, Yang, Wenjing Duan, and Qing Cao. "The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach." *Decision Support Systems* 55, no. 4 (2013): 919-26.

Zhang, Xue, Hauke Fuehres, and Peter A. Gloor. "Predicting Stock Market Indicators Through Twitter "I Hope It Is Not as Bad as I Fear." *Procedia - Social and Behavioral Sciences*26 (2011): 55-62.