Aalto University
School of Business

PREDICTING STOCK RETURNS BY NUMBER OF COMPANY MENTIONS IN TWEETS

Bachelor's Thesis
Kimi Päivärinta
Aalto University School of Business
Finance
Fall 2017

Predicting Stock Returns by Number of Company Mentions in Tweets

Department of Finance

Aalto University School of Business

Thesis instructor: Matthijs Lof

Fall 2017

Abstract

This study attempts to establish whether return or magnitude of return can be predicted by how many tweets mention a company by either its name or stock symbol. The sample data consists of 365 million tweets of which 706,700 mention a S&P 500 company between June 1st, 2016 and June 30th, 2017. It was found that tweets which mention a company by its stock symbol while stock markets are open have a positive impact on its return between 0 to 1%. No evidence was found of number of tweets holding a predictive value of the magnitude of return.

Kimi Päivärinta

Table of Contents

1 Introduction

Social media has grown to be a natural part of the society. Largest social media platforms such as Twitter and Facebook have come a long way from the days they were founded and broke through. Facebook is the older of these two and was founded in 2003 and Twitter a few years later in 2006. At the beginning of 2010 Facebook and Twitter had 40 million active users combined (Statista, 2017a, 2018b). The growth has been fast after that, and just Twitter had around 330 million active users in Q3 of 2017. Users tweet roughly 500 million tweets a day which amounts to 200 billion tweets in a year (Twitter, 2011). The vast number of tweets contain information that can be utilized in many different industries, including finance.

Although Facebook and Twitter – the platforms which we conceive as social media – are rather young, there have been other social media platforms such as discussion forums on the Internet for a lot longer. The potential of the information on the social media was noticed already in the late 1990s and early 2000s. One of the earliest studies was conducted by Wysocki in 1998. He studied over 3000 posts on Yahoo! Finance forum and found evidence of a correlation between number of messages and the trading volume in NYSE. Twitter has made its way to be a de facto social media platform which has also lead to studies on its impact on Finance field. There are numerous studies on how tweets' sentiment have an impact on stock markets, but only a few studying the effect of the volume of tweets (Bollen, Mao & Zeng, 2011, Porshnev, Redkin & Schevchenko, 2013; Zhang, Fuehres & Gloor, 2011). A more recent study was conducted by Sanger & Warin (2016) where the positive correlation was found between number of tweets mentioning a company and overnight return of the stock.

I hypothesize that volume of tweets holds potential to become an essential part of predicting the stock markets via social media. For this reason, in this thesis, I aim to replicate the results of Sanger & Warin (2016) with a newer dataset of tweets. I also aim to test if number of tweets can be used to predict the magnitude of return. A proven predictive result would allow to use number of tweets together with sentiment analysis to improve prediction results. So, the research question is whether the number of tweets has an explanatory power for changes in company stock valuations.

Sanger & Warin (2016) found out that tweets, which are released while stock markets are not open and mention a company by its stock ticker, positively explain a return between 1 to 5% of the said stock. I did not find evidence for this, but evidence that tweets released while stock market is open positively explains a return between 0 to 1%. Furthermore, number of tweets does not explain the magnitude of return on a significant level.

The structure of this thesis is as follows: in Chapter 2, I present the motivation for the study as well as a short history of social media studies and review the existing literature. Hypotheses are also presented in Chapter 2 based on the literature and motivation. In Chapter 3 I present the datasets used in this thesis – tweet and stock market data – as well as how this data was obtained and what information it holds. In Chapter 4, I summarize and discuss the results which are preceded by the conclusion in Chapter 5.

2 Literature Review, Motivation, and Hypotheses

2.1 Literature Review and Motivation

Before Twitter was founded and become common social media platform, the discussions of stocks were held in message boards. In this Chapter, I will first go through literature studying the impact of postings on message boards (Wysocki, 1998; Antweiler & Frank, 2004; De Choudhury et al., 2008). Then moving onwards to Twitter studies which mostly study how tweets' sentimental can be used in prediction of stock returns (Bollen, Mao & Zeng, 2011; Porshnev, Redkin & Shevchenko, 2013; Zhang, Fuehres & Gloor, 2011; Schultz & Sheffer; 2010; Kouloumpis, 2011). Finally, reviewing the most recent study of Twitter that concentrates on the number of tweets mentioning a company by either its stock ticker or name and the differences between these mentioning styles (Sanger & Waring, 2016).

Wysocki (1998) was one of the first to study postings on message boards and their possible impact on stock markets. He studied around 3,000 stocks listed on Yahoo! Finance message boards between January 1998 and August 1998. Although the Internet – or World Wide Web as it was mostly referred to then – was in its early stages, thousands of new messages of the most popular stocks were released daily. However, variation in number of number of messages per company a day was high; the peak was 1,740 messages about Dell Computer, but at the same time only two messages were written about USX – U.S. Steel Group. Interestingly, it was found that with overnight messages it was possible to predict next day changes in volume and returns. The time between stock close and its opening is regarded as the Overnight period. Similarly, Intraday is the period when the stock market is open. Evidence was found that when number of messages was between 10 to 100 times above the average, then the trading volume of the following stock day will be 15.6% higher and there will be 0.7% abnormal return (Wysocki, 1998).

In 2004 another study on Yahoo! Finance message board was conducted by Antweiler & Frank. In addition to Yahoo! Finance message board, the data source was broadened to take into account messages from Raging bull message board. The dataset consisted of impressive 1.5 million postings regarding 45 companies in the Dow Jones Industrial Average and the Dow Jones Internet Index. The study did not only focus on the number of messages, but it also analyzed the meaning of the messages and what information could be extracted from them about the stocks. Antweiler & Frank (2004) were first to report that from high number of messages it is possible to predict negative subsequent returns for the given stocks. The result was found to be robust, but economically small and most likely not usable as a trading strategy when transaction costs are taken into account. This result is on the contrary to Wysocki's (1998) result that a high number of messages would result in a positive abnormal return in the subsequent days. However, both studies find significant evidence that a high number of messages predict higher trading volumes in the subsequent days (Wysocki, 1998; Antweiler & Frank, 2004). Furthermore, De Choudhury et al. (2008) created a forecasting model based on messages on technological-savvy messaging board called Engadget. The results of the forecasting model proved to be promising, and it predicted the direction of the movement of stock markets correct 87% of the time and had 78% accuracy in predicting the magnitude of the movement.

Number of Twitter studies have grown steadily at the same time when Twitter has grown. These studies are mostly focusing on the sentiment of tweets and try to predict the stock market returns and volumes based on the sentiment. Studies assume that large-scale mood changes in Twitter mood mean that the whole public's mood is also changing. However, proving that the Twitter mood is indeed global mood has not yet been done due to its complex nature and lack of mood studies. (Bollen, Mao & Zeng, 2011). With the utilization of the presumed public mood changes, existing DJIA prediction models can be improved to predict the daily up and down changes with 87.6% accuracy (Bollen, Mao & Zeng, 2011). However, the study acquired the public mood from Twitter messages via two different methods that identify the mood. Only one of the series showed significant results in the prediction model. Furthermore, the emotional word or mood with the most correlation between stock market returns was "calm" (Bollen, Mao & Zeng, 2011). These results have been replicated by Porshnev, Redkin & Shevchenko (2013), although, the prediction precision was not as good with as 87.6%, their model resulted in precision of 70%.

Zhang, Fuehres, & Gloor (2011) also found correlation between emotional words in tweets and DJIA. The correlation was highly negative. Therefore, when emotional words are mentioned in tweets, the DJIA index will decrease. Both of the studies agree on the correlation between emotional words and stock returns, but the words with the most correlation are very different. Bollen, Mao & Zeng (2011) found the most correlation with word "calm" which is neutral and usually positive. On the contrary, Zhang, Fuehres & Gloor (2011) found the most correlation with words "fear," "worry" and "hope" which are not deemed neutral, but slightly negative.

Although a few papers are studying the correlation between Tweets' sentiment and stock returns, there are only a few studies on the impact of number of tweets on stock returns and trading volumes (De Choudhury et al. 2008; Sanger & Warin, 2016). Also, Twitter sentimental analyses discussed above were measured the whole publics mood and predicted an Index return with it. The same methods can be applied to company specific tweets which could lead to improved accuracy also. Based this literature review, there is a need for a new study on the impact of a number of tweets. Studies above have proven that Twitter can be utilized in predicting stock markets and I hypothesize that also using the number of tweets could drastically improve the prediction models. Furthermore, when more data sources are used it becomes possible to predict the stock market with more narrow scope, such as on company level.

I chose to study the potential of number of tweets in predicting the magnitude of stock markets, and also to replicate Sanger's & Warin's (2016) paper "High Frequency and Unstructured Data in Finance: An Exploratory Study of Twitter." This paper is one of most the recent studies focusing on number of tweets and not on the sentiment. The scope of the paper was on tweets mentioning S&P500 companies and how these mentions impact the corresponding stock's overnight return. The tweets were collected for 251 business days between May 1st, 2012 and May 1st, 2013, but they did not enclose the total number of tweets. Sanger & Waring (2016) found a positive correlation between tweets and stock return of 1 to 5% when specific conditions were met.

The results of Sanger & Warin (2016) indicated that there is a positive correlation between tweets by financially literate people and stock return between 1 to 5%. They divided the tweets into two different categories: Tweets by financially literate, and financially illiterate. The division was done based on how the company was mentioned in the tweet. Furthermore, overnight return was defined as $R_{overnight} = \frac{Price\ open_t - Price\ close_{t-1}}{Price\ close_{t-1}}$. A logistic model was used to execute the analysis which means aggregating the returns to a binary value 0 or 1. Sanger & Waring (2016) defined three different Logistic models. In Model if result is positive it returns a value of 1 otherwise 0. In Model

2 if result is between 0 and 1% it returns 1. and in Model 3 if result is between 1 and 5% it returns 1, otherwise 0. (Sanger & Waring, 2016).

2.2 Hypotheses

The research question is whether number of tweets mentioning a company has a predictive value of corresponding stocks' return or magnitude of return. Hence, I focus on the number of tweets and their impact on stocks returns and magnitude of returns. Overall, the hypothesis follows Sanger's & Warin's (2016) hypotheses, which were

H1. Professionals tweet after they are done at work and their comments may be interesting for investment decisions implemented on the next day.

H2. Tweets written by professionals (using tickers) provide more useful financial information than tweets written by the layman (using company names).

About the first Hypothesis, I think it is safe to assume that finance professionals do not have time to tweet about markets while doing their job, especially if they are, e.g., traders. Furthermore, it is also logical to assume that finance professionals do have more information about stocks and hence the impact of their tweets is more significant than other tweets. In addition to the given hypotheses I define two new hypotheses and the first one is as follows:

H3. An overnight tweet mentioning a company provide more useful financial information than an intraday tweet and can be used to predict overnight returns.

Hypothesis 4 relies on the efficient market theory where public information is included in the price immediately. Information of Intraday tweet should, therefore, transfer to the stock price immediately which makes is harder to observe as we are only looking at the opening and closing price. However, the Overnight tweet's information should transfer to the stock's price when the stock opens. Hence it should be easy to observe this by calculating an overnight return.

The sentiment Twitter analysis has already proven that it is possible to predict the sign of the return. I assume that a raw number of tweets does not hold information of return's sign, but only magnitude of it. Hence, the fourth Hypothesis is as follows:

H4. The magnitude of return can be predicted by number of tweets.

3 Data and Methods

3.1 Sample data

The vast number of tweets written each day enables the possibility to study many things without having to worry about too few observations. However, the so-called "big data" also requires extra attention and work effort from researchers as the dataset may require a lot of storage space and even specialized skills with computers. In addition, big data requires computational power which may not be readily available. There are multiple different ways to obtain tweets for research. These methods are all suitable for different situations, and I explain them briefly below.

Via free Twitter Application Programming Interface (API) it is possible to search and download for maximum one-week old tweets. It is not possible to download all tweets within the week, but specific keywords have to be used. Generally speaking one week's tweets about a specific topic is not enough for a study. For this study, I want to acquire tweets for one year period. Hence this search functionality cannot be used. Twitter API also provides a so-called "Streaming API" where it is possible to download tweets when they are released with or without specific rules. The free version, so-called "Twitter Spritzer," enables collection of around 1% of the tweets live when they are released. The next collection levels are "Garden hose" (10%) and "Firehose" (100%) which are behind a paywall (Twitter, 2013; Twitter, 2017). However, Twitter has not enclosed how the Tweets are chosen for the streaming API. Hence it is not sure how well this 1% collected sample represents all the tweets. These versions of the streaming API also require the collector to have a dedicated server for collecting the tweets around the clock with enough storage space. I opted to download a set of tweets collected via Twitter Spritzer for a period of one year from Archive.org which allows to expedite the process by one year. Tweets are released on this site in monthly packages and at the time of this study, the most recent package was June 2017. To use the most recent tweets available the scope of the paper is from July 1st, 2016 to June 30th, 2017.

Monthly tweet packages include a file for every minute containing all tweets in that minute in JSON format. The combined size of these packages is over 1 TB which complicates the handling of the material substantially. Every tweet has a language variable which is estimated by Twitter by not disclosed algorithm. The scope of this study is S&P500 companies which are traded in NYSE and NASDAQ in the United States, so it can be assumed that tweets in English would have the largest impact. Based on this assumption all tweets not in English were discarded leaving a set of 365 million tweets for the given period. This is roughly 0.2% of all 200 billion tweets released yearly (Twitter, 2011). Tweet message, Tweet's ID, user's ID, timestamp of creation (GMT+0) and list of used hashtags were saved for every tweet. Two additional variables were needed for the analysis and had to be determined for each tweet. First one was the type of the tweet, meaning whether the tweet was released when the stock market was open or when it was closed, these cases are referred to as Intraday and Overnight respectively. The second variable to be determined was the business day on which the possible impact could be seen. Assuming Stock markets are efficient in its semi-strong form, all public information should be included in the price immediately. For an Intraday type, the business day would hence be the day the tweet was released. For an Overnight tweet, the business day would be the day whenever stock markets open for the next time. A tweet released on Friday evening 9 pm GMT -4 would have the following Monday as its business day. From this point on, tweets in these two categories are referred to as "Overnight tweets" and "Intraday tweets." Further information on tweet type and business day in Chapter 3.2. Unfortunately, Twitter does not enclose the geolocation of tweet in the free version of Streaming API, so impact between tweets from different locations cannot be studied.

From the 365 million tweets in English, all tweets that mentioned an S&P500 in either of the following ways were searched for. The common way of marking a topic of a tweet is to use the hashtag (#) character before the topic or keyword. Tweets mentioning a company name was searched for in two different ways: 1) hashtag and the official name of the company, e.g., #netflixltd and 2) hashtag and the spoken name of the company, e.g., #netflix. These tweets are later referenced to as name tweets and are not by the Hypothesis two to have much information on stocks.

Finance professionals and literate people have adopted their own way of talking about a specific stock. A dollar sign ($) is used instead of a hashtag, and it is preceded by the stock ticker. Therefore, it is assumed that professionals tweeting about Netflix Ltd. would include $NFLX in their tweet. By Hypothesis 1 and 2, this kind of tweet should include valuable information about the stock which should transfer to the price. These tweets are later referred to as ticker tweets. A full list of name and ticker search words are presented in Appendix 1. A combined dataset of name and ticker tweets are later on called "all tweets" in this paper.

With search words presented in Table 1, 706,700 unique tweets are found (unique by tweet's Id). When uniqueness is defined by having both unique tweet's Id and search type – name or ticker – the total number of tweets found is 931,937. This means that 275,238 tweets mention a company by both name and ticker method.

Table 1 List of companies with at least 15 tweets per day on average.

List of companies and search words

Company			Search words		
No.	Name	Stock ticker	Ticker	Official name	Spoken name
9	AdvanceAutoParts	AAP	$AAP	#advanceautoparts	#advanceauto
14	AgilentTechnologiesInc	A	$A	#agilenttechnologiesinc	#agilenttechnologies
28	AlphabetIncClassC	GOOG	$GOOG	#alphabetincclassc	#alphabetincclassc
30	AmazoncomInc	AMZN	$AMZN	#amazoncominc	#amazoncom
52	AppleInc	AAPL	$AAPL	#appleinc	#apple
58	ATTInc	T	$T	#attinc	#att
73	BestBuyCoInc	BBY	$BBY	#bestbuycoinc	#bestbuy
77	BoeingCompany	BA	$BA	#boeingcompany	#boeing
86	CAInc	CA	$CA	#cainc	#ca
97	CBSCorp	CBS	$CBS	#cbscorp	#cbs
109	ChubbLimited	CB	$CB	#chubblimited	#chubblimited
116	CitigroupInc	C	$C	#citigroupinc	#citi
122	CoachInc	COH	$COH	#coachinc	#coach
158	DominionEnergy	D	$D	#dominionenergy	#dominionenergy
169	eBayInc	EBAY	$EBAY	#ebayinc	#ebay
183	EsteeLauderCos	EL	$EL	#esteelaudercos	#esteelaudercos
193	FacebookInc	FB	$FB	#facebookinc	#facebook
206	FordMotor	F	$F	#fordmotor	#fordmotor
215	GeneralElectric	GE	$GE	#generalelectric	#generalelectric
227	HarleyDavidson	HOG	$HOG	#harleydavidson	#harleydavidson
252	IntelCorp	INTC	$INTC	#intelcorp	#intel
271	KelloggCo	K	$K	#kelloggco	#kellogg
272	KeyCorp	KEY	$KEY	#keycorp	#key
281	L3CommunicationsHoldings	LLL	$LLL	#l3communicationsholdings	#l3communicationsholdings
292	LoewsCorp	L	$L	#loewscorp	#loews
297	MacysInc	M	$M	#macysinc	#macys
317	MicrosoftCorp	MSFT	$MSFT	#microsoftcorp	#microsoft
325	MorganStanley	MS	$MS	#morganstanley	#morganstanley
329	NasdaqInc	NDAQ	$NDAQ	#nasdaqinc	#nasdaq
333	NetflixInc	NFLX	$NFLX	#netflixinc	#netflix
341	Nike	NKE	$NKE	#nike	#nike
344	Nordstrom	JWN	$JWN	#nordstrom	#nordstrom
361	PayPal	PYPL	$PYPL	#paypal	#paypal
364	PepsiCoInc	PEP	$PEP	#pepsicoinc	#pepsi
392	QuintilesIMSHoldingsInc	Q	$Q	#quintilesimsholdingsinc	#quintilesimsholdings
396	RealtyIncomeCorporation	O	$O	#realtyincomecorporation	#realtyincome
409	Salesforcecom	CRM	$CRM	#salesforcecom	#salesforcecom
423	SouthernCo	SO	$SO	#southernco	#southern
427	StarbucksCorp	SBUX	$SBUX	#starbuckscorp	#starbucks
443	TiffanyCo	TIF	$TIF	#tiffanyco	#tiffany
477	VisaInc	V	$V	#visainc	#visa
486	WellsFargo	WFC	$WFC	#wellsfargo	#wellsfargo

If a company is mentioned in tweets very seldom or the number of tweets in total is low, the number of tweets is likely to not have any valuable information of the stock which would affect the stock's return. To make the model more robust, a threshold of at least 15 tweets a day on average was set. From a total of 505 S&P500 companies, there are 42 companies that were mentioned at least 15 times per day in the studied dataset. In total, these 42 companies were tweeted about 585,323 times.

Companies used in the study are presented in Table 1. A full list of the companies presented in the Appendix 1.

Stock data for the S&P500 companies were downloaded from Quandl by using R. The data consisted of the following adjusted variables for each day: Open price, Day high price, Day low price, Close price, and volume for the day. The adjusted stock data does not include stock price changes that are only nominal and does not affect the stock's real value (Thomas, 2016). Event such as dividend payout may cause more tweets about the company, but being able to "predict" these events based on Twitter analysis does not offer benefit value to investors. These events are also known beforehand, hence it would not be predicting in scope of this thesis. Intraday and Overnight returns were calculated from these values by the following formulas:

$$R_{overnight} = \frac{Price\ open_t - Price\ close_{t-1}}{Price\ close_{t-1}} \qquad (1)$$

$$R_{intraday} = \frac{Price\ close_t - Price\ open_t}{Priceo\ open_t} \qquad (2)$$

where t presents the business day for which the return is calculated for. Descriptive statistics of the variables are shown in Table 2 and 3.

Table 2 Descriptive statistics of Overnight returns and corresponding Overnight tweets in different subcategories.

Descriptive Statistics of Overnight Tweets

	Returns	All tweets	Ticker tweets	Name tweets
Number of observations	9767	9767	9767	9767
Number of companies	42	42	42	42
Max	16.95 %	2608	1273	2608
Min	-14.39 %	1	0	0
Average	0.06 %	60	26	35
Median	0.04 %	23	6	6
1st Q.	-0.21 %	9	2	1
3rd Q.	0.32 %	70	22	26

Table 3 Descriptive statistics of Intraday returns and corresponding Intraday tweets in different subcategories.

Descriptive Statistics of Intraday Tweets

	Returns	All tweets	Ticker tweets	Name tweets
Number of observations	9263	9263	9263	9263
Number of companies	42	42	42	42
Max	19.47 %	424	424	325
Min	-8.37 %	1	0	0
Average	0.02 %	25	15	11
Median	0.01 %	12	4	3
1st Q.	-0.54 %	5	2	1
3rd Q.	0.58 %	35	14	10

3.1 Methods

Methods for testing the hypotheses are adapted from Sanger & Warin (2016) study which use Logistic model estimations. In logistic models, the dependent variable is reduced to a binary variable, 0 or 1. A same null hypothesis can be applied to both OLS and logistic models. The underlying method, for determining if a null hypothesis is true and on what level of significance, is the same for both methods. Hence the result acquired from either method will apply also in the other (Hosmer et al. 1997). However, the benefit of the logistic model is that it enables the possibility to calculate odds-ratio which is a ratio of an occurrence happening and not happening compared to a reference value or point. It should be noted that odds-ratio is not by itself a probability of how a variable explains a model, but it is a probability compared to some predetermined reference value or point (Hosmer et al. 1997).

By the hypothesis, the regression equation below is ran for two different types of tweets which are further divided into two more categories. The main tweet types are Intraday, and Overnight which indicate whether a tweet was released while the stock market is open (Intraday) or closed (Overnight). The motivation for dividing the tweets into two different categories are as follows: the impact of an Overnight tweet should be more easily observable than Intraday tweet impact. This is because information of Intraday tweet is transferred to the price immediately, but in this paper, I only have price information on day open and close level. Therefore, observing the price change after tweet's release is not possible. Information of Overnight tweet is, however, transferred to the price immediately when the stock opens for the next time. The change of the price by the information is therefore observable as it happens at once. The metric for observing this is the Overnight return explained later in this chapter.

Tweets are further divided into two sub-categories depending on how the company was mentioned in the tweet: name or ticker. This division follows Hypothesis 1: Ticker mentions are made by professionals, and their tweets should carry more information. In total, there are six different types of tweets with which the regression is run: Overnight ticker, Overnight all tweets, Overnight name, Intraday ticker, Intraday all tweets and Intraday name.

The predictor variable of regression is the number of tweets in an aforementioned category a day. Tweets are aggregated by company, tweet type and business day to acquire the sum of total number of tweets and sum of total number of ticker tweets for each business day. As earlier explained, a business day of a tweet is the day when the information of the tweet can and will be transferred to the price. For an Intraday tweet, the business day is the same day. For an Overnight tweet, the business day is the next business day the stock opens. In the calculation of a business day, I have taken into account the DST periods and holidays of NYSE and NASDAQ.

Three logistic models used in this study are the same as in Sanger & Warin (2016), as follows:

$$Model\ 1 = \begin{cases} 1\ if\ return >\ 0\ \% \\ 0\ otherwise \end{cases} \quad (3)$$

$$Model\ 2 = \begin{cases} 1\ if\ 0\% < return \leq 1\% \\ 0\ otherwise \end{cases} \quad (4)$$

$$Model\ 3 = \begin{cases} 1\ if\ 1\% < return \leq 5\% \\ 0\ otherwise \end{cases} \quad (5)$$

Where the first Model captures correlation between tweets and any positive return. In Model 2 the threshold is a moderate return between 0 and 1%. In the third Model, the variable is 1 if the return is between 1 and 5%.

The regression formula used is shown below. A 5% significance level is needed for disproving a null hypothesis.

$$\theta_{t\ \varepsilon} = a_0 + a_1 \log\left(\beta_{t\ \varepsilon}\right) + \alpha_2 \omega \qquad (6)$$

Where θ is the model used for a business day and tweet type, θ = {Model 1; Model 2; Model 3}. A business day is presented as t, ε = {intraday; overnight} is the type of return and also tweet. Number of tweets per type for a business day is presented by β. A natural logarithm is taken from the number of tweets for two reasons. First, the number of tweets is not normally distributed as is seen from the Figure 1. Taking a logarithm of the variable does make the variable more normally distributed which will make it better suitable for linear regression (Benoit, 2011). Secondly, the average number of tweets a day differs between companies, but are put into the same model. Some companies have 1,000 tweets a day on average whereas some have only 30. For the latter, an increase of 30 tweets per day would mean 100% increase, but in the first one, it is only a 3% increase. Taking a logarithm of the variable will scale these changes' impact closer to each other. It can be assumed that if the average number of tweets per day about a company is high, then an extraordinary event will also result into higher number of tweets. Sanger & Warin (2016) do not mention in their study whether these concerns are taken into account or not in their model. For controlling fixed effects, weekday dummies ω are added to the equation that absorb the weekday effect to make the results more robust (Hardy, 1993). In need of more control variables, then industry control variables can be added.

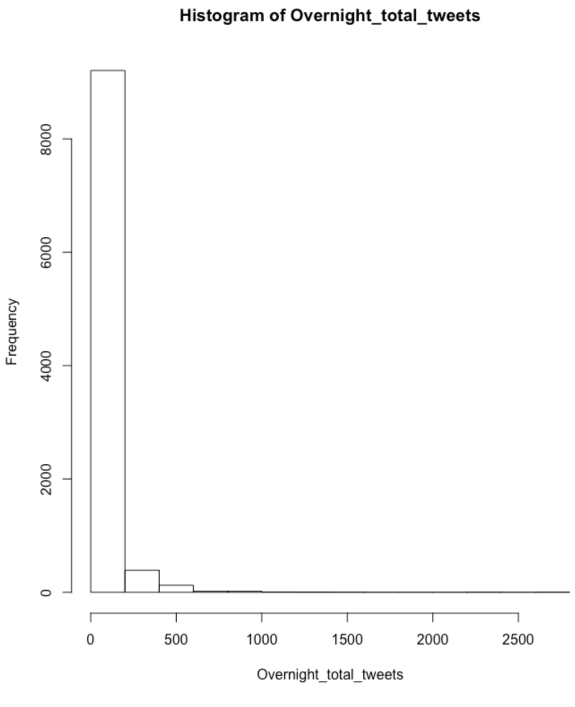

Figure 1 Overnight tweets distribution.

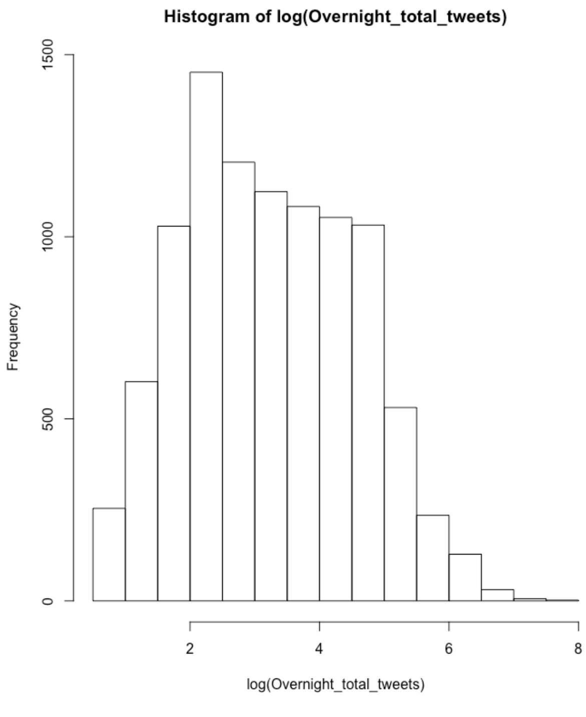

Figure 2 Overnight tweets distribution after logarithm.

For testing Hypothesis 4: Number of tweets do not correlate with the magnitude of stock return, a regression equation as below is used. This regression aims to explain the magnitude of return by number of tweets:

$$Abs(R_{t\,\varepsilon}) = a_0 + a_1 \log(\beta_{t\,\varepsilon}) + \alpha_2 \omega \qquad (5)$$

R is the return of a given business day and tweet type for a company. An absolute percentage value of the return is used in the regression, so a correlation between number of tweets and the magnitude of return can be tested. The motivation for taking an absolute of the return is that number of tweets does not include information about the sign of the return, but only the magnitude of the return. Rest of the variables are as explained above.

4 Results and Discussion

4.1 Results

I was not able to replicate the results of Sanger & Warin (2016) where number of overnight ticker tweets had a significant explanatory value on Model 3 - positive return between 1 to 5%. Regression results for overnight ticker tweets are in Table 4. The number of tweets does not explain any model at significant level. Interestingly, in Model 1 and 2, the number of tweets negatively correlate with the dependent variable which means that the possibility of positive return or return between 0 and 1% would decrease when number of tweets increase. This effect is opposite to Sanger & Warin (2016) findings.

Table 4 Regression output. Dependent variable Model 1 to 3 is calculated based on Overnight return. The number of tweets is tweets mentioning a company by its ticker.

Regression output: Dependent variable Model 1 to 3 is based on Overnight returns. Predictor value is Overnight ticker tweets.

	Model 1			Model 2			Model 3		
	Estimate	Std, Error	T value	Estimate	Std, Error	T value	Estimate	Std, Error	T value
No. Of tweets	-0.002	0.003	-0.658	-0.003	0.003	-0.825	0.001	0.001	0.474
Monday	-0.036	0.016	-2.235 *	-0.028	0.016	-1.711 .	-0.011	0.006	-1.791 .
Tuesday	-0.008	0.016	-0.501	0.005	0.016	0.282 .	-0.014	0.006	-2.247 *
Wednesday	-0.041	0.016	-2.523 *	-0.029	0.016	-1.755 .	-0.017	0.006	-2.717 **
Thursday	0.002	0.016	0.148	0.015	0.016	0.911	-0.014	0.006	-2.29 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Overnight all tweets also do not explain any model on significant level, these results can be seen in Table 5. In Model 1 the effect is nearly zero and in Model 3, the relationship is positive, but non-significant. Again, in Model 2, the relationship is negative. Both in Overnight ticker and all tweets regressions, some of the dummy weekdays are significant at 10% level on explaining the returns.

Table 5 Regression output. Dependent variable Model 1 to 3 is calculated based on Overnight return. The number of tweets is tweets mentioning a company by either its ticker, official name or spoken name.

Regression output: Dependent variable Model 1 to 3 is based on Overnight returns. Predictor value is Overnight all tweets.

	Model 1			Model 2			Model 3		
	Estimate	Std, Error	T value	Estimate	Std, Error	T value	Estimate	Std, Error	T value
No. of tweets	0.000	0.004	0.031	-0.001	0.004	-0.367	0.001	0.001	0.925
Monday	-0.035	0.016	-2.161 *	-0.027	0.016	-1.687 .	-0.010	0.006	-1.674 .
Tuesday	-0.007	0.016	-0.411	0.005	0.016	0.294	-0.013	0.006	-2.069 *
Wednesday	-0.040	0.017	-2.413 *	-0.028	0.017	-1.717 .	-0.016	0.006	-2.542 *
Thursday	0.004	0.017	0.221	0.015	0.017	0.911	-0.013	0.006	-2.119 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

In three regressions, Intraday tweets were found to be significant at 10% level, and one at 5% level. Intraday ticker tweets explain Model 1 dependent variable at 10% significance level, meaning that larger number of tweets would predict positive return for the stock, however, a 5% significance level is needed for the results. A more significant result at 5% level is acquired from the regression between Model 2 - positive return between 0 to 1% - and intraday ticker tweets. In accordance to the regression equation, the estimated value of 0.009 for predictor value is not an increase in return per tweet. The dependent variable is a binary value of 0 and 1 and not a return. It should also be noted that the predictor variable is not an absolute number of tweets, but a logarithm of the given value. Hence, relatively larger movement than 1 tweet is required for the impact to realize. Taken into account that other regressions are not providing any significant results, this result can be seen quite robust. As discussed earlier, I did not assume that Intraday tweet's impact could be observed when return is measured only once for a day. This result suggests that by collecting and analyzing tweets live, it could be possible to predict intraday price changes in short-term. However, this requires further studies and different techniques which are further discussed in following chapter. Also, some of the weekday dummies are positively correlating with the model variables at 10% significance level. See Table 6 for full results.

Intraday all tweets positively explain Model 1 variable at 10% significance level which is below the 5% level expected. In addition, weekday dummies Monday and Thursday explain the return at 5% significance level. For Model 2 and 3, there are positive impact visible, but not at any significance level.

Table 6 Regression output. Dependent variable Model 1 to 3 is calculated based on Intraday return. The number of tweets is tweets mentioning a company by its ticker.

Regression output: Dependent variable Model 1 to 3 is based on Intraday returns. Predictor value is Intraday ticker tweets.

	Model 1			Model 2			Model 3		
	Estimate	Std, Error	T value	Estimate	Std, Error	T value	Estimate	Std, Error	T value
No. of tweets	0.007	0.004	1.822 .	0.009	0.004	2.468 *	-0.002	0.003	-0.686
Monday	-0.040	0.017	-2.442 *	-0.021	0.016	-1.308	-0.020	0.011	-1.761 .
Tuesday	-0.003	0.017	-0.185	-0.018	0.016	-1.118	0.012	0.012	1.055
Wednesday	-0.024	0.017	-1.414	-0.006	0.016	-0.398	-0.020	0.012	-1.728 .
Thursday	-0.033	0.017	-1.993 *	-0.030	0.016	-1.885 .	-0.005	0.011	-0.447

Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Regarding the Overnight and Intraday name tweets, I found they had no impact to returns by them at 5% significance level. Intraday name tweets have a positive impact in Model 3 at significance level of 10%. These results are in line with Sanger & Waring (2016) findings. The full regression results are presented in Appendix 1 and Appendix 2.

Table 7 Regression output. Dependent variable Model 1 to 3 is calculated based on Intraday return. The number of tweets is tweets mentioning a company by either its ticker, official name or spoken name.

Regression output: Dependent variable Model 1 to 3 is based on Intraday returns. Predictor value is Intraday all tweets.

	Model 1			Model 2			Model 3		
	Estimate	Std, Error	T value	Estimate	Std, Error	T value	Estimate	Std, Error	T value
No. of tweets	0.009	0.005	1.939 .	0.005	0.004	1.167	0.004	0.003	1.301
Monday	-0.041	0.017	-2.451 *	-0.021	0.016	-1.294	-0.021	0.011	-1.793 .
Tuesday	-0.003	0.017	-0.201	-0.018	0.016	-1.112	0.012	0.012	1.023
Wednesday	-0.024	0.017	-1.426	-0.006	0.016	-0.382	-0.020	0.012	-1.768 .
Thursday	-0.033	0.017	-2.009 *	-0.030	0.016	-1.882 .	-0.005	0.011	-0.473

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hypothesis 1 states: professionals tweet after the workday and their tweets may have important information. Based on the results in this study this cannot be confirmed. The results do not indicate that tweets mentioning a company by a ticker has an impact or significant explanatory value on the overnight return of the stock. Hypothesis 2: tweets by professional have more valuable information than tweets by laymans (company name) can be confirmed. When the regressions are compared between Intraday and Overnight tweets, it seems that on the contrary to Hypothesis 4, the Intraday tweets are better to explain returns. Intraday ticker tweets positively explain the return between 0 to 1% at 5% significant level. Also, two more Intraday tweet regressions were significant at 10% level. A 5% significance level is required for the results in this study. Hence, these results are not taken into account. Overall, I did not hypothesize this kind of results which are also in contrary to Sanger & Warin (2016) who did not find any significant results from Intraday tweets.

For testing Hypothesis 4: The magnitude of return can be predicted by number of tweets, the dependent variable is absolute return percentage. I did not find evidence that number of tweets would explain the magnitude of the return. Overall, the estimated impact is very close to zero for all predictor values as can be seen in Table 8. Results suggests that neither number of tweets nor the weekday dummies have any impact on the absolute percentage of return. Dummies also play a lesser role in these regressions when compared to logistic model regressions above. Only in Intraday ticker and all tweets regressions did Tuesday have a significant explanatory value at 10% significance level, although, the impact is nearly zero. Based on these results number of tweets does not hold any predictive value on a magnitude of return and Hypothesis 4 cannot be accepted. As a sole information source, number of tweets does not fulfil the potential as a predictor variable in Finance as I assumed.

Table 8 Regression output. Tweets explaining magnitude of return.

Regression output: Dependent variable based on Overnight or Intraday returns. Predictor value is the number number of tweets in given categories.

	Overnight						Intraday					
	Ticker			All tweets			Ticker			All tweets		
	Estimate	Std, Error	T value	Estimate	Std, Error	T value	Estimate	Std, Error	T value	Estimate	Std, Error	T value
No. Of tweets	0.000	0.000	-0.727	0.000	0.000	0.962	0.000	0.000	-0.708	0.000	0.000	0.653
Monday	0.000	0.000	0.717	0.000	0.000	0.913	0.000	0.000	0.331	0.000	0.000	0.31
Tuesday	0.000	0.000	0.922	0.000	0.000	1.195	0.001	0.000	1.894 .	0.000	0.000	1.873 .
Wednesday	0.000	0.000	-0.135	0.000	0.000	0.132	0.000	0.000	0.228	0.000	0.000	0.201
Thursday	0.000	0.000	0.7	0.000	0.000	0.959	0.000	0.000	0.616	0.000	0.000	0.599

Signif.codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

4.2 Discussion and Further Research

In this chapter, the results are discussed in more speculative manner. Observations and possible reasons for the large difference in the results between this paper and those in the study by Sanger & Warin (2016) are discussed. Possible further research topics and limitations of methods in this study are also discussed.

I found that Intraday ticker tweets have a significant explanatory value on stock returns between 0 to 1%. I think to be able to utilize this finding in a trading strategy; then price changes should be observed in shorter period of time. Preferably the tweets would be analyzed in real-time. However, this proposes a problem: there are not many companies that are tweeted often enough. In this study, I found 42 companies that were tweeted on average at least 15 times a day, although, my data present only around 1% of the total tweets in English in a year.

The relatively small amount of companies that are actively tweeted about is a limitation of this twitter analysis method. If a predictive model can be built from tweet data, there is only a limited set of companies to which it can be used - 42 S&P500 companies based on this study. To obtain more than only 1% of the tweets, then one has to pay for the data. In case of a trading strategy, it means that the excess returns from a Twitter based prediction model should surpass the price of the data and computational power which are required to process these vast amounts of data. The excess returns acquired from this kind of Twitter analysis should hence be notable.

Sanger & Warin (2016) do not reveal much information about their sample data and its source, so most of the reasons for the different results are speculative. Firstly, the studied tweets have taken place at different time. Their tweets are from May 1st, 2012 to May 1st, 2013 whereas the tweets studied in this paper were released between July 1st, 2016 and June 30th, 2017. It is possible that the culture of tweeting about companies has changed in some way, e.g., informative tweets are made while working and not afterward.

Furthermore, they studied 71 companies that were tweeted at least 30 times a day on average; I studied 41 companies that were tweeted at least 15 times a day. Without further knowledge of their data, it cannot be said if these daily tweet numbers are comparable or not. In total, we had 25 companies in common. Tweets about different companies may very have a different amount of valuable information. Descriptive statistics of their variables implies that their sample data consisted of around 300 million tweets in which a S&P500 company was mentioned – 16,043 observations, average value of 16,732 and maximum value of 1,091,212. This is notably more than the 706,700 tweets studied in this paper. This data was acquired from Twitter Spritzer and should, therefore, present around 1% of all tweets in English mentioning an S&P500 company – which would imply that there are 70 million tweets in English mentioning S&P500 companies in total. However, the language of Sanger & Warin (2016) dataset is not known.

As the results significantly differ from each other, there is a need for a new replication study with also some new techniques on how the possible effect should be measured. It seems that the number of tweets as an absolute value or as a logarithmic value does not propose definite prediction value that could be used in all situations. One method would be to not use absolute number of tweets nor logarithmic value, but to calculate relative position compared to average number of tweets per day per the given company. The formula for this variable is as follows:

$$Relative\ number\ of\ tweets = \frac{Number\ of\ tweets_{t\ \varepsilon}}{Average\ number\ of\ tweets\ per\ day_{\epsilon}} \quad (7)$$

where t is business day and ε is a company. This equation would scale all companies' number of tweets to same scale while still preserving differences between often tweeted and seldom tweeted companies. Companies that on average are tweeted more daily would still require a larger change in the number of tweets to have an impact on the variable. Both of these studies have also focused on S&P 500 companies which are very large which makes them absorb effect on many different levels.

Hypothesis 4: The magnitude of return can be predicted by number of tweets, was not confirmed by this data and study. I assumed that number of tweets could have had more potential in predicting the magnitude of stock returns. This prediction value would have been a good fit with Twitter sentiment analyses that have been conducted and have provided predictive value on stock returns.

Overall, there have not been many studies looking into the potential of number of tweets and other social media messages. Further studies are required on the number of tweets analysis to determine whether it could help predict some movements in the financial market, either alone or in a model with other variables.

Furthermore, twitter is only one platform of the many in the field of social media. Due to its semi-open APIs and mainly text format it has been an easy target for research. However, many other social media platforms such as Instagram and Flickr also possess a huge amount of information but are mainly in the unstructured format: images and videos. These formats are harder to process than text with technology available at the moment (Xing et al., 2015). Processing large amounts of Tweets with a normal office computer was not possible ten years ago, and is barely possible nowadays. Although the growth of the computational power is still fast, I see that it will not be possible to study these image/video social media platforms in the near-future. In addition, before moving to these platforms, I presume that text-based social media and its impact on stock markets have to be thoroughly studied.

4.3 Acknowledgements

Some of the used search words such as "#apple" are common words and do not only refer to the company under research. No action due to these tweets was taken as it can be assumed that the amount of these tweets stay relatively stable on a daily level. The prediction variable used in this study is the number of tweets on a logarithmic scale which downscales the problem of having some number of unrelated tweets in the dataset.

5 Conclusion

Social media has become an essential part of everyday life and it contains of vast amount of information. This information flood is not utilized to its full potential in Finance. In this thesis, I studied over 700 thousand tweets that mentioned a S&P 500 company between June 1st, 2016 and June 30th, 2017 to find an answer whether number of tweets mentioning a company has a predictive value of corresponding stocks' return or magnitude of return.

I found that Intraday ticker tweets have a positive impact on stock's Intraday return between 0 to 1%. Intraday ticker tweets are tweets that are released when stock markets are open and mention a S&P500 company by its stock ticker. Intraday return is the stock's return calculated from open and closing price. Sanger's & Warin (2016) have studied the same phenomenon. They did not find any significant results in Intraday tweets. They found Overnight ticker tweets explaining the overnight return of 1 to 5%. Furthermore, I worked on a hypothesis that number of tweets mentioning a company can predict its stock's magnitude of return – absolute percentage of return. The logic for this hypothesis was based on assumption that the number of tweets does not include information whether message and mood is positive or negative, but it holds information of the event's magnitude. I tested this hypothesis with both Overnight and Intraday tweets and their subcategories ticker, name and all tweets. These regression results did not produce any significant results and the hypothesis cannot be confirmed.

Benefits of using information on social media has been proven by, e.g., Twitter sentiment studies such as Bollen, Mao & Zeng (2011) and Porshnev, Redkin & Shevchenko (2013). Further research is still needed to acquire consistent results between methods and data samples. The amount of information is growing all the time and it is getting more complex. It is increasingly hard to discern beneficial information from noise and fake information. To be able to benefit from all the new data, I think text based social media platforms must be first studied thoroughly.

6 References

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259-1294.

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, *22*(1), 23-36.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8.

De Choudhury, M., Sundaram, H., John, A., & Seligmann, D. D. (2008, June). Can blog communication dynamics be correlated with stock market activity?. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia* (pp. 55-60). ACM.

Hardy, M. A. (1993). *Regression with dummy variables* (Vol. 93). Sage.

Hosmer, D. W., Hosmer, T., Le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine*, *16*(9), 965-980.

Internet Live Stats. (2017). *Twitter Usage Statistics - Internet Live Stats*. *Internetlivestats.com*. Retrieved 4 December 2017, from http://www.internetlivestats.com/twitter-statistics/

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg!. *Icwsm*, *11*(538-541), 164.

Pohlman, J. T., & Leitner, D. W. (2003). A comparison of ordinary least squares and logistic regression.

Porshnev, A., Redkin, I., & Shevchenko, A. (2013). Improving prediction of stock market indices by analyzing the psychological states of twitter users.

Sanger, W., & Warin, T. (2016). High Frequency and Unstructured Data in Finance: An Exploratory Study of Twitter. *Journal of Global Research in Computer Science*, *7*(4).

Schultz, B., & Sheffer, M. L. (2010). An exploratory study of how Twitter is affecting sports journalism. *International Journal of Sport Communication*, *3*(2), 226-239.

Statista. (2017). Facebook users worldwide 2017 | Statista. Statista. Retrieved 29 November 2017, from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Statista. (2017). Twitter: number of active users 2010-2017 | Statista. Statista. Retrieved 29 November 2017, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Thomas, A. (2016). *The Comprehensive Guide to Stock Price Calculation - Quandl Blog*. *Quandl Blog*. Retrieved 4 December 2017, from https://blog.quandl.com/guide-to-stock-price-calculation

Twitter. (2017). Overview. Developer.twitter.com. Retrieved 1 December 2017, from https://developer.twitter.com/en/docs/tweets/filter-realtime/overview

Twitter. (2011). *200 million Tweets per day*. *Blog.twitter.com*. Retrieved 4 December 2017, from https://blog.twitter.com/official/en_us/a/2011/200-million-tweets-per-day.html

Wysocki, P. D. (1999). Cheap talk on the web: The determinants of postings on stock message boards.

Xing, E. P., Ho, Q., Dai, W., Kim, J. K., Wei, J., Lee, S., ... & Yu, Y. (2015). Petuum: A new platform for distributed machine learning on big data. *IEEE Transactions on Big Data*, *1*(2), 49-67.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter “I hope it is not as bad as I fear”. *Procedia-Social and Behavioral Sciences*, *26*, 55-62.

7 Appendix

Appendix 1. Regression output. Dependent variable Model 1 to 3 is calculated based on Intraday return. The number of tweets is tweets mentioning a company by its name.

Regression output: Dependent variable Model 1 to 3 is based on Intraday returns. Predictor value is Intraday name tweets.

	Model 1			Model 2			Model 3		
	Estimate	**Std, Error**	**T value**	**Estimate**	**Std, Error**	**T value**	**Estimate**	**Std, Error**	**T value**
No. of tweets	0.001	0.004	0.176	-0.004	0.004	-1.161	0.005	0.003	1.81 .
Monday	-0.040	0.017	-2.417 *	-0.020	0.016	-1.271	-0.020	0.011	-1.777 .
Tuesday	-0.003	0.017	-0.166	-0.017	0.016	-1.082	0.012	0.012	1.035
Wednesday	-0.023	0.017	-1,384	-0.006	0.016	-0.348	-0.020	0.012	-1,753 .
Thursday	-0.033	0.017	-1.981 *	-0.030	0.016	-1.853 .	-0.005	0.011	-0.471

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix 2. Regression output. Dependent variable Model 1 to 3 is calculated based on Overnight return. The number of tweets is number of tweets mentioning a company by its name.

Regression output: Dependent variable Model 1 to 3 is based on Overnight returns. Predictor value is Overnight name tweets.

	Model 1			Model 2			Model 3		
	Estimate	**Std, Error**	**T value**	**Estimate**	**Std, Error**	**T value**	**Estimate**	**Std, Error**	**T value**
No. Of tweets	0.003	0.003	1.161	0.001	0.003	0.442	0.002	0.001	1.428
Monday	-0.033	0.016	-2.081 *	-0.026	0.016	-1.606	-0.010	0.006	-1.704 .
Tuesday	-0.005	0.016	-0.281	0.007	0.016	0.432	-0.013	0.006	-2.127 *
Wednesday	-0.038	0.016	-2.314 *	-0.026	0.016	-1.614	-0.016	0.006	-2.594 **
Thursday	0.006	0.016	0.353	0.017	0.016	1.049	-0.013	0.006	-2.171 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix 3. List of all S&P companies, search words and number of tweets found with search words.

List of companies and used search words

Company			Search words			Number of tweets				
No.	Name	Stock ticker	Ticker	Official name	Spoken name	Ticker	Official name	Spoken name	Total	Total unique
1	3MCompany	MMM	$MMM	#3mcompany	#3m	264	1	352	617	476
2	AbbottLaboratories	ABT	$ABT	#abbottlaboratories	#abbottlaboratories	267	2	14	283	13
3	AbbVieInc	ABBV	$ABBV	#abbvieinc	#abbvie	267	0	23	290	12
4	Accentureplc	ACN	$ACN	#accentureplc	#accenture	196	0	329	525	281
5	ActivisionBlizzard	ATVI	$ATVI	#activisionblizzard	#activisionblizzard	283	13	226	522	205
6	AcuityBrandsInc	AYI	$AYI	#acuitybrandsinc	#acuitybrands	52	0	7	59	5
7	AdobeSystemsInc	ADBE	$ADBE	#adobesystemsinc	#adobesystems	294	0	1772	2066	1651
8	AdvancedMicroDevicesInc	AMD	$AMD	#advancedmicrodevicesinc	#advancedmicrodevices	942	0	1076	2018	941
9	AdvanceAutoParts	AAP	$AAP	#advanceautoparts	#advanceautoparts	8530	9	11	8550	8
10	AESCorp	AES	$AES	#aescorp	#aes	35	0	56	91	89
11	AetnaInc	AET	$AET	#aetnainc	#aetna	279	0	200	479	359
12	AffiliatedManagersGroupInc	AMG	$AMG	#affiliatedmanagersgroupinc	#affiliatedmanagers	450	0	0	450	377
13	AFLACInc	AFL	$AFL	#aflacinc	#aflac	115	0	39	154	92
14	AgilentTechnologiesInc	A	$A	#agilenttechnologiesinc	#agilenttechnologies	78649	0	34	78683	71193
15	AirProductsChemicalsInc	APD	$APD	#airproductschemicalsinc	#airproductschemicals	126	0	0	126	NA
16	AkamaiTechnologiesInc	AKAM	$AKAM	#akamaitechnologiesinc	#akamaitechnologies	91	0	34	125	33
17	AlaskaAirGroupInc	ALK	$ALK	#alaskaairgroupinc	#alaskaair	249	0	46	295	38
18	AlbemarleCorp	ALB	$ALB	#albemarlecorp	#albemarle	105	0	92	197	90
19	AlexandriaRealEstateEquitiesInc	ARE	$ARE	#alexandriarealestateequitiesinc	#alexandriarealestateequities	214	0	0	214	NA
20	AlexionPharmaceuticals	ALXN	$ALXN	#alexionpharmaceuticals	#alexionpharmaceuticals	152	1	6	159	4
21	AlignTechnology	ALGN	$ALGN	#aligntechnology	#aligntechnology	41	1	35	77	34
22	Allegion	ALLE	$ALLE	#allegion	#allegion	24	2	2	28	NA
23	AllerganPlc	AGN	$AGN	#allerganplc	#allergan	504	0	97	601	34
24	AllianceDataSystems	ADS	$ADS	#alliancedatasystems	#alliancedatasystems	257	4	5	266	2
25	AlliantEnergyCorp	LNT	$LNT	#alliantenergycorp	#alliantenergy	71	0	0	71	57
26	AllstateCorp	ALL	$ALL	#allstatecorp	#allstate	272	0	170	442	146
27	AlphabetIncClassA	GOOGL	$GOOGL	#alphabetincclassa	#alphabetincclassa	1434	0	468	1902	1012
28	AlphabetIncClassC	GOOG	$GOOG	#alphabetincclassc	#alphabetincclassc	3314	0	468	3782	893
29	AltriaGroupInc	MO	$MO	#altriagroupinc	#altria	1224	0	6	1230	3
30	AmazoncomInc	AMZN	$AMZN	#amazoncominc	#amazoncom	3724	2	62964	66690	61504
31	AmerenCorp	AEE	$AEE	#amerencorp	#ameren	27	0	5	32	2
32	AmericanAirlinesGroup	AAL	$AAL	#americanairlinesgroup	#americanairlines	290	1	611	902	575
33	AmericanElectricPower	AEP	$AEP	#americanelectricpower	#americanelectricpower	281	5	5	291	NA
34	AmericanExpressCo	AXP	$AXP	#americanexpressco	#americanexpress	302	0	367	669	260
35	AmericanInternationalGroupInc	AIG	$AIG	#americaninternationalgroupinc	#americaninternational	213	0	0	213	NA
36	AmericanTowerCorpA	AMT	$AMT	#americantowercorpa	#americantowercorpa	190	0	3	193	1
37	AmericanWaterWorksCompanyInc	AWK	$AWK	#americanwaterworkscompanyinc	#americanwaterworks	77	0	0	77	NA
38	AmeripriseFinancial	AMP	$AMP	#ameripriseFinancial	#ameriprisefinancial	197	1	1	199	NA
39	AmerisourceBergenCorp	ABC	$ABC	#amerisourcebergencorp	#amerisourcebergen	274	0	8	282	2
40	AMETEKInc	AME	$AME	#ametekinc	#ametek	92	0	7	99	2
41	AmgenInc	AMGN	$AMGN	#amgeninc	#amgen	399	0	56	455	41
42	AmphenolCorp	APH	$APH	#amphenolcorp	#amphenol	475	0	10	485	4
43	AnadarkoPetroleumCorp	APC	$APC	#anadarkopetroleumcorp	#anadarkopetroleum	299	0	0	299	NA
44	AnalogDevicesInc	ADI	$ADI	#analogdevicesinc	#analogdevices	148	0	16	164	2
45	Andeavor	ANDV	$ANDV	#andeavor	#andeavor	NA	NA	NA	NA	NA
46	ANSYS	ANSS	$ANSS	#ansys	#ansys	38	27	27	92	21
47	AnthemInc	ANTM	$ANTM	#antheminc	#anthem	164	2	835	1001	787
48	Aonplc	AON	$AON	#aonplc	#aon	39	1	106	146	105
49	AOSmithCorp	AOS	$AOS	#aosmithcorp	#aosmith	46	0	0	46	NA
50	ApacheCorporation	APA	$APA	#apachecorporation	#apache	142	1	738	881	716
51	ApartmentInvestmentManagement	AIV	$AIV	#apartmentinvestmentmanagement	#apartmentinvestmentmanagement	54	0	0	54	NA
52	AppleInc	AAPL	$AAPL	#appleinc	#apple	8421	75	40489	48985	35968
53	AppliedMaterialsInc	AMAT	$AMAT	#appliedmaterialsinc	#appliedmaterials	242	0	5	247	1
54	ArcherDanielsMidlandCo	ADM	$ADM	#archerdanielsmidlandco	#archerdanielsmidland	285	0	0	285	NA
55	ArconicInc	ARNC	$ARNC	#arconicinc	#arconic	39	0	22	61	21
56	ArthurJGallagherCo	AJG	$AJG	#arthurjgallagherco	#arthurjgallagher	43	0	1	44	1
57	AssurantInc	AIZ	$AIZ	#assurantinc	#assurant	45	0	9	54	7
58	ATTInc	T	$T	#attinc	#att	29247	1	898	30146	16640
59	AutodeskInc	ADSK	$ADSK	#autodeskinc	#autodesk	122	0	369	491	357
60	AutomaticDataProcessing	ADP	$ADP	#automaticdataprocessing	#automaticdataprocessing	217	1	1	219	NA
61	AutoZoneInc	AZO	$AZO	#autozoneinc	#autozone	70	0	27	97	23
62	AvalonBayCommunitiesInc	AVB	$AVB	#avalonbaycommunitiesinc	#avalonbaycommunities	100	0	0	100	NA
63	AveryDennisonCorp	AVY	$AVY	#averydennisoncorp	#averydennison	42	0	12	54	9
64	BakerHughesaGECompany	BHGE	$BHGE	#bakerhughesagecompany	#bakerhughesage	NA	NA	NA	NA	NA
65	BallCorp	BLL	$BLL	#ballcorp	#ball	81	0	1202	1283	1246
66	BankofAmericaCorp	BAC	$BAC	#bankofamericacorp	#bankofamerica	893	0	237	1130	524
67	TheBankofNewYorkMellonCorp	BK	$BK	#thebankofnewyorkmelloncorp	#thebankofnewyorkmellon	2548	0	2	2550	1134
68	BardCRInc	BCR	$BCR	#bardcrinc	#bardcr	158	0	0	158	105
69	BaxterInternationalInc	BAX	$BAX	#baxterinternationalinc	#baxterinternational	113	0	0	113	42
70	BBTCorporation	BBT	$BBT	#bbtcorporation	#bbt	157	0	417	574	469
71	BectonDickinson	BDX	$BDX	#bectondickinson	#bectondickinson	115	7	7	129	44
72	BerkshireHathaway	BRK.B	$BRK.B	#berkshirehathaway	#berkshirehathaway	75	65	65	205	96
73	BestBuyCoInc	BBY	$BBY	#bestbuycoinc	#bestbuy	212	0	6034	6246	5944
74	BiogenInc	BIIB	$BIIB	#biogeninc	#biogen	304	0	62	366	164
75	BlackRock	BLK	$BLK	#blackrock	#blackrock	288	378	378	1044	508

Company			Search words			Number of tweets				
No.	Name	Stock ticker	Ticker	Official name	Spoken name	Ticker	Official name	Spoken name	Total	Total unique
76	BlockHR	HRB	$HRB	#blockhr	#blockhr	105	0	0	105	43
77	BoeingCompany	BA	$BA	#boeingcompany	#boeing	3561	1	2151	5713	3344
78	BorgWarner	BWA	$BWA	#borgwarner	#borgwarner	69	6	6	81	40
79	BostonProperties	BXP	$BXP	#bostonproperties	#bostonproperties	36	1	1	38	28
80	BostonScientific	BSX	$BSX	#bostonscientific	#bostonscientific	134	13	13	160	67
81	BrighthouseFinancialInc	BHF	$BHF	#brighthousefinancialinc	#brighthousefinancial	NA	NA	NA	NA	NA
82	BristolMyersSquibb	BMY	$BMY	#bristolmyerssquibb	#bristolmyerssquibb	469	42	42	553	225
83	Broadcom	AVGO	$AVGO	#broadcom	#broadcom	271	27	27	325	22
84	BrownFormanCorp	BF.B	$BF.B	#brownformancorp	#brownforman	23	0	6	29	16
85	CHRobinsonWorldwide	CHRW	$CHRW	#chrobinsonworldwide	#chrobinsonworldwide	45	0	0	45	NA
86	CAInc	CA	$CA	#cainc	#ca	3709	7	21489	25205	21345
87	CabotOilGas	COG	$COG	#cabotoilgas	#cabotoilgas	90	0	0	90	NA
88	CadenceDesignSystems	CDNS	$CDNS	#cadencedesignsystems	#cadencedesignsystems	44	1	1	46	NA
89	CampbellSoup	CPB	$CPB	#campbellsoup	#campbellsoup	55	8	8	71	8
90	CapitalOneFinancial	COF	$COF	#capitalonefinancial	#capitalonefinancial	177	1	1	179	NA
91	CardinalHealthInc	CAH	$CAH	#cardinalhealthinc	#cardinalhealth	125	9	8	142	12
92	CBOEHoldings	CBOE	$CBOE	#cboeholdings	#cboeholdings	51	0	0	51	NA
93	CarmaxInc	KMX	$KMX	#carmaxinc	#carmax	89	0	30	119	28
94	CarnivalCorp	CCL	$CCL	#carnivalcorp	#carnival	252	0	1381	1633	1376
95	CaterpillarInc	CAT	$CAT	#caterpillarinc	#caterpillar	639	0	415	1054	287
96	CBREGroup	CBG	$CBG	#cbregroup	#cbre	73	19	135	227	137
97	CBSCorp	CBS	$CBS	#cbscorp	#cbs	250	0	4891	5141	4798
98	CelgeneCorp	CELG	$CELG	#celgenecorp	#celgene	351	0	22	373	7
99	CenteneCorporation	CNC	$CNC	#centenecorporation	#centene	154	1	12	167	5
100	CenterPointEnergy	CNP	$CNP	#centerpointenergy	#centerpointenergy	41	12	12	65	41
101	CenturyLinkInc	CTL	$CTL	#centurylinkinc	#centurylink	164	0	123	287	265
102	Cerner	CERN	$CERN	#cerner	#cerner	91	26	26	143	111
103	CFIndustriesHoldingsInc	CF	$CF	#cfindustriesholdingsinc	#cfindustriesholdings	436	0	0	436	435
104	CharlesSchwabCorporation	SCHW	$SCHW	#charlesschwabcorporation	#charlesschwab	126	0	15	141	130
105	CharterCommunications	CHTR	$CHTR	#chartercommunications	#chartercommunications	159	3	3	165	160
106	ChesapeakeEnergy	CHK	$CHK	#chesapeakeenergy	#chesapeakeenergy	529	11	11	551	525
107	ChevronCorp	CVX	$CVX	#chevroncorp	#chevron	475	1	365	841	794
108	ChipotleMexicanGrill	CMG	$CMG	#chipotlemexicangrill	#chipotlemexicangrill	510	17	769	1296	1267
109	ChubbLimited	CB	$CB	#chubblimited	#chubblimited	3634	0	0	3634	3568
110	ChurchDwight	CHD	$CHD	#churchdwight	#churchdwight	89	14	14	117	98
111	CIGNACorp	CI	$CI	#cignacorp	#cigna	760	0	90	850	806
112	CimarexEnergy	XEC	$XEC	#cimarexenergy	#cimarexenergy	58	0	0	58	57
113	CincinnatiFinancial	CINF	$CINF	#cincinnatifinancial	#cincinnatifinancial	44	0	0	44	NA
114	CintasCorporation	CTAS	$CTAS	#cintascorporation	#cintas	59	0	15	74	63
115	CiscoSystems	CSCO	$CSCO	#ciscosystems	#ciscosystems	738	65	1962	2765	2484
116	CitigroupInc	C	$C	#citigroupinc	#citi	42739	0	266	43005	35223
117	CitizensFinancialGroup	CFG	$CFG	#citizensfinancialgroup	#citizensfinancial	56	0	684	740	683
118	CitrixSystems	CTXS	$CTXS	#citrixsystems	#citrixsystems	67	13	522	602	499
119	TheCloroxCompany	CLX	$CLX	#thecloroxcompany	#theclorox	51	0	0	51	NA
120	CMEGroupInc	CME	$CME	#cmegroupinc	#cme	142	1	426	569	416
121	CMSEnergy	CMS	$CMS	#cmsenergy	#cmsenergy	47	0	0	47	NA
122	CoachInc	COH	$COH	#coachinc	#coach	327	2	4672	5001	4652
123	CocaColaCompanyThe	KO	$KO	#cocacolacompanythe	#cocacolacompanythe	1485	0	0	1485	1187
124	CognizantTechnologySolutions	CTSH	$CTSH	#cognizanttechnologysolutions	#cognizanttechnologysolutions	189	1	1	191	1
125	ColgatePalmolive	CL	$CL	#colgatepalmolive	#colgatepalmolive	3160	2	163	3325	157
126	ComcastCorp	CMCSA	$CMCSA	#comcastcorp	#comcast	307	0	922	1229	871
127	ComericaInc	CMA	$CMA	#comericainc	#comerica	82	0	11	93	7
128	ConagraBrands	CAG	$CAG	#conagrabrands	#conagrabrands	89	2	15	106	14
129	ConchoResources	CXO	$CXO	#conchoresources	#conchoresources	70	2	2	74	NA
130	ConocoPhillips	COP	$COP	#conocophillips	#conocophillips	286	27	139	452	128
131	ConsolidatedEdison	ED	$ED	#consolidatededison	#consolidatededison	651	0	0	651	562
132	ConstellationBrands	STZ	$STZ	#constellationbrands	#constellationbrands	108	1	597	706	685
133	TheCooperCompanies	COO	$COO	#thecoopercompanies	#thecoopercompanies	121	0	0	121	NA
134	CorningInc	GLW	$GLW	#corninginc	#corning	166	1	168	335	256
135	CostcoWholesaleCorp	COST	$COST	#costcowholesalecorp	#costcowholesale	345	0	362	707	345
136	CotyInc	COTY	$COTY	#cotyinc	#coty	67	0	66	133	60
137	CrownCastleInternationalCorp	CCI	$CCI	#crowncastleinternationalcorp	#crowncastleinternational	77	1	1	79	2
138	CSRAInc	CSRA	$CSRA	#csrainc	#csra	29	0	31	60	24
139	CSXCorp	CSX	$CSX	#csxcorp	#csx	163	0	84	247	70
140	CumminsInc	CMI	$CMI	#cumminsinc	#cummins	128	0	118	246	110
141	CVSHealth	CVS	$CVS	#cvshealth	#cvshealth	333	25	318	676	295
142	DRHorton	DHI	$DHI	#drhorton	#drhorton	54	8	8	70	58
143	DanaherCorp	DHR	$DHR	#danahercorp	#danaher	139	1	14	154	82
144	DardenRestaurants	DRI	$DRI	#dardenrestaurants	#dardenrestaurants	143	2	2	147	116
145	DaVitaInc	DVA	$DVA	#davitainc	#davita	155	0	9	164	129
146	DeereCo	DE	$DE	#deereco	#deere	837	13	43	893	680
147	DelphiAutomotivePLC	DLPH	$DLPH	#delphiautomotiveplc	#delphiautomotive	52	0	1	53	46
148	DeltaAirLinesInc	DAL	$DAL	#deltaairlinesinc	#deltaairlines	328	0	505	833	795
149	DentsplySirona	XRAY	$XRAY	#dentsplysirona	#dentsplysirona	40	10	10	60	38
150	DevonEnergyCorp	DVN	$DVN	#devonenergycorp	#devonenergy	74	0	6	80	47
151	DigitalRealtyTrustInc	DLR	$DLR	#digitalrealtytrustinc	#digitalrealtytrust	84	0	0	84	57
152	DiscoverFinancialServices	DFS	$DFS	#discoverfinancialservices	#discoverfinancialservices	109	1	2069	2179	2116
153	DiscoveryCommunicationsA	DISCA	$DISCA	#discoverycommunicationsa	#discoverycommunicationsa	66	0	0	66	57
154	DiscoveryCommunicationsC	DISCK	$DISCK	#discoverycommunicationsc	#discoverycommunicationsc	20	0	0	20	14
155	DishNetwork	DISH	$DISH	#dishnetwork	#dishnetwork	149	593	593	1335	675
156	DollarGeneral	DG	$DG	#dollargeneral	#dollargeneral	977	72	72	1121	854
157	DollarTree	DLTR	$DLTR	#dollartree	#dollartree	122	74	74	270	160
158	DominionEnergy	D	$D	#dominionenergy	#dominionenergy	22962	0	0	22962	13505
159	DoverCorp	DOV	$DOV	#dovercorp	#dover	54	1	870	925	868
160	DowDuPont	DWDP	$DWDP	#dowdupont	#dowdupont	0	1	1	2	1

Company			Search words			Number of tweets				
No.	Name	Stock ticker	Ticker	Official name	Spoken name	Ticker	Official name	Spoken name	Total	Total unique
161	DrPepperSnappleGroup	DPS	$DPS	#drpeppersnapplegroup	#drpeppersnapple	100	3	1	104	1
162	DTEEnergyCo	DTE	$DTE	#dteenergyco	#dteenergy	105	0	13	118	2
163	DukeRealtyCorp	DRE	$DRE	#dukerealtycorp	#dukerealty	36	0	3	39	NA
164	DukeEnergy	DUK	$DUK	#dukeenergy	#dukeenergy	170	61	61	292	23
165	DXCTechnology	DXC	$DXC	#dxctechnology	#dxctechnology	70	25	25	120	25
166	ETrade	ETFC	$ETFC	#etrade	#etrade	57	655	655	1367	536
167	EastmanChemical	EMN	$EMN	#eastmanchemical	#eastmanchemical	59	0	0	59	39
168	EatonCorporation	ETN	$ETN	#eatoncorporation	#eaton	51	0	232	283	273
169	eBayInc	EBAY	$EBAY	#ebayinc	#ebay	286	0	79435	79721	79372
170	EcolabInc	ECL	$ECL	#ecolabinc	#ecolab	90	0	7	97	43
171	EdisonIntl	EIX	$EIX	#edisonintl	#edisonintl	91	1	1	93	41
172	EdwardsLifesciences	EW	$EW	#edwardslifesciences	#edwardslifesciences	810	4	4	818	553
173	ElectronicArts	EA	$EA	#electronicarts	#electronicarts	829	85	85	999	444
174	EmersonElectricCompany	EMR	$EMR	#emersonelectriccompany	#emersonelectric	151	0	15	166	82
175	EntergyCorp	ETR	$ETR	#entergycorp	#entergy	253	0	26	279	174
176	EnvisionHealthcare	EVHC	$EVHC	#envisionhealthcare	#envisionhealthcare	40	2	2	44	32
177	EOGResources	EOG	$EOG	#eogresources	#eogresources	166	20	20	206	86
178	EQTCorporation	EQT	$EQT	#eqtcorporation	#eqt	72	0	344	416	394
179	EquifaxInc	EFX	$EFX	#equifaxinc	#equifax	50	0	49	99	87
180	Equinix	EQIX	$EQIX	#equinix	#equinix	74	32	32	138	75
181	EquityResidential	EQR	$EQR	#equityresidential	#equityresidential	39	0	0	39	24
182	EssexPropertyTrustInc	ESS	$ESS	#essexpropertytrustinc	#essexpropertytrust	81	0	1	82	64
183	EsteeLauderCos	EL	$EL	#esteelaudercos	#esteelaudercos	6983	0	0	6983	2068
184	EversourceEnergy	ES	$ES	#eversourceenergy	#eversourceenergy	2755	3	3	2761	1978
185	EverestReGroupLtd	RE	$RE	#everestregroupltd	#everestre	1331	0	0	1331	914
186	ExelonCorp	EXC	$EXC	#exeloncorp	#exelon	193	0	21	214	79
187	ExpediaInc	EXPE	$EXPE	#expediainc	#expedia	164	0	178	342	272
188	ExpeditorsInternational	EXPD	$EXPD	#expeditorsinternational	#expeditorsinternational	32	0	0	32	29
189	ExpressScripts	ESRX	$ESRX	#expressscripts	#expressscripts	166	8	8	182	4
190	ExtraSpaceStorage	EXR	$EXR	#extraspacestorage	#extraspacestorage	45	4	4	53	34
191	ExxonMobilCorp	XOM	$XOM	#exxonmobilcorp	#exxonmobil	838	0	917	1755	1270
192	F5Networks	FFIV	$FFIV	#f5networks	#f5networks	55	5	5	65	50
193	FacebookInc	FB	$FB	#facebookinc	#facebook	5503	25	62676	68204	65456
194	FastenalCo	FAST	$FAST	#fastenalco	#fastenal	79	0	4	83	54
195	FederalRealtyInvestmentTrust	FRT	$FRT	#federalrealtyinvestmenttrust	#federalrealtyinvestmenttrust	110	0	0	110	93
196	FedExCorporation	FDX	$FDX	#fedexcorporation	#fedex	231	0	728	959	865
197	FidelityNationalInformationServices	FIS	$FIS	#fidelitynationalinformationservices	#fidelitynationalinformationservices	199	0	0	199	94
198	FifthThirdBancorp	FITB	$FITB	#fifththirdbancorp	#fifththirdban	72	0	0	72	53
199	FirstEnergyCorp	FE	$FE	#firstenergycorp	#firstenergy	550	0	13	563	401
200	FiservInc	FISV	$FISV	#fiservinc	#fiserv	99	0	26	125	22
201	FLIRSystems	FLIR	$FLIR	#flirsystems	#flirsystems	20	5	5	30	20
202	FlowserveCorporation	FLS	$FLS	#flowservecorporation	#flowserve	830	1	23	854	90
203	FluorCorp	FLR	$FLR	#fluorcorp	#fluor	58	0	9	67	57
204	FMCCorporation	FMC	$FMC	#fmccorporation	#fmc	335	0	59	394	319
205	FootLockerInc	FL	$FL	#footlockerinc	#footlocker	1560	0	138	1698	583
206	FordMotor	F	$F	#fordmotor	#fordmotor	22162	139	7182	29483	16350
207	FortiveCorp	FTV	$FTV	#fortivecorp	#fortive	29	0	0	29	NA
208	FortuneBrandsHomeSecurity	FBHS	$FBHS	#fortunebrandshomesecurity	#fortunebrandshomesecurity	35	0	0	35	NA
209	FranklinResources	BEN	$BEN	#franklinresources	#franklinresources	113	0	0	113	51
210	FreeportMcMoRanInc	FCX	$FCX	#freeportmcmoraninc	#freeportmcmoran	345	0	26	371	4
211	GapInc	GPS	$GPS	#gapinc	#gap	142	2	957	1101	1015
212	GarminLtd	GRMN	$GRMN	#garminltd	#garmin	80	0	943	1023	995
213	GartnerInc	IT	$IT	#gartnerinc	#gartner	1165	0	476	1641	1267
214	GeneralDynamics	GD	$GD	#generaldynamics	#generaldynamics	1506	23	23	1552	849
215	GeneralElectric	GE	$GE	#generalelectric	#generalelectric	5118	29	29	5176	4040
216	GeneralGrowthPropertiesInc	GGP	$GGP	#generalgrowthpropertiesinc	#generalgrowthproperties	80	0	0	80	65
217	GeneralMills	GIS	$GIS	#generalmills	#generalmills	137	65	65	267	147
218	GeneralMotors	GM	$GM	#generalmotors	#generalmotors	1276	223	1878	3377	2537
219	GenuineParts	GPC	$GPC	#genuineparts	#genuineparts	48	11	11	70	42
220	GileadSciences	GILD	$GILD	#gileadsciences	#gileadsciences	854	21	21	896	512
221	GlobalPaymentsInc	GPN	$GPN	#globalpaymentsinc	#globalpayments	40	0	2	42	34
222	GoldmanSachsGroup	GS	$GS	#goldmansachsgroup	#goldmansachs	2631	57	713	3401	2650
223	GoodyearTireRubber	GT	$GT	#goodyeartirerubber	#goodyeartirerubber	450	0	0	450	283
224	GraingerWWInc	GWW	$GWW	#graingerwwinc	#graingerww	49	0	0	49	39
225	HalliburtonCo	HAL	$HAL	#halliburtonco	#halliburton	371	0	41	412	239
226	HanesbrandsInc	HBI	$HBI	#hanesbrandsinc	#hanesbrands	68	0	2	70	51
227	HarleyDavidson	HOG	$HOG	#harleydavidson	#harleydavidson	147	1979	1979	4105	2041
228	HarrisCorporation	HRS	$HRS	#harriscorporation	#harris	51	0	347	398	386
229	HartfordFinancialSvcGp	HIG	$HIG	#hartfordfinancialsvcgp	#hartfordfinancialsvcgp	49	0	0	49	35
230	HasbroInc	HAS	$HAS	#hasbroinc	#hasbro	176	0	846	1022	923
231	HCAHoldings	HCA	$HCA	#hcaholdings	#hcaholdings	151	0	0	151	61
232	HCPInc	HCP	$HCP	#hcpinc	#hcp	43	0	93	136	115
233	HelmerichPayne	HP	$HP	#helmerichpayne	#helmerichpayne	768	0	0	768	520
234	HenrySchein	HSIC	$HSIC	#henryschein	#henryschein	54	10	10	74	48
235	TheHersheyCompany	HSY	$HSY	#thehersheycompany	#thehershey	168	0	0	168	82
236	HessCorporation	HES	$HES	#hesscorporation	#hess	88	0	30	118	84
237	HewlettPackardEnterprise	HPE	$HPE	#hewlettpackardenterprise	#hewlettpackardenterprise	210	6	45	261	43
238	HiltonWorldwideHoldingsInc	HLT	$HLT	#hiltonworldwideholdingsinc	#hiltonworldwideholdings	86	0	0	86	62
239	Hologic	HOLX	$HOLX	#hologic	#hologic	54	12	12	78	45
240	HomeDepot	HD	$HD	#homedepot	#homedepot	623	423	423	1469	785
241	HoneywellIntlInc	HON	$HON	#honeywellintlinc	#honeywellintl	170	0	0	170	105
242	HormelFoodsCorp	HRL	$HRL	#hormelfoodscorp	#hormelfoods	72	0	3	75	51
243	HostHotelsResorts	HST	$HST	#hosthotelsresorts	#hosthotelsresorts	81	0	0	81	69
244	HPInc	HPQ	$HPQ	#hpinc	#hp	260	6	1974	2240	1873
245	HumanaInc	HUM	$HUM	#humanainc	#humana	204	0	96	300	142

Company			Search words			Number of tweets				
No.	Name	Stock ticker	Ticker	Official name	Spoken name	Ticker	Official name	Spoken name	Total	Total unique
246	HuntingtonBancshares	HBAN	$HBAN	#huntingtonbancshares	#huntingtonbancshares	52	0	0	52	42
247	IDEXXLaboratories	IDXX	$IDXX	#idexxlaboratories	#idexxlaboratories	58	5	5	68	50
248	IHSMarkitLtd	INFO	$INFO	#ihsmarkitltd	#ihsmarkit	43	0	18	61	51
249	IllinoisToolWorks	ITW	$ITW	#illinoistoolworks	#illinoistoolworks	113	2	2	117	1
250	IlluminaInc	ILMN	$ILMN	#illuminainc	#illumina	132	0	19	151	73
251	IngersollRandPLC	IR	$IR	#ingersollrandplc	#ingersollrand	362	0	4	366	267
252	IntelCorp	INTC	$INTC	#intelcorp	#intel	971	2	3681	4654	3952
253	IntercontinentalExchange	ICE	$ICE	#intercontinentalexchange	#intercontinentalexchange	138	0	0	138	67
254	InternationalBusinessMachines	IBM	$IBM	#internationalbusinessmachines	#internationalbusinessmachines	884	1	1	886	522
255	Incyte	INCY	$INCY	#incyte	#incyte	188	7	7	202	84
256	InternationalPaper	IP	$IP	#internationalpaper	#internationalpaper	604	7	7	618	413
257	InterpublicGroup	IPG	$IPG	#interpublicgroup	#interpublic	78	0	1	79	NA
258	IntlFlavorsFragrances	IFF	$IFF	#intlflavorsfragrances	#intlflavorsfragrances	48	0	0	48	42
259	IntuitInc	INTU	$INTU	#intuitinc	#intuit	147	0	76	223	140
260	IntuitiveSurgicalInc	ISRG	$ISRG	#intuitivesurgicalinc	#intuitivesurgical	147	0	1	148	80
261	InvescoLtd	IVZ	$IVZ	#invescoltd	#inves	53	0	26	79	60
262	IronMountainIncorporated	IRM	$IRM	#ironmountainincorporated	#ironmountainincorporated	52	0	0	52	NA
263	JacobsEngineeringGroup	JEC	$JEC	#jacobsengineeringgroup	#jacobsengineering	53	1	0	54	39
264	JBHuntTransportServices	JBHT	$JBHT	#jbhunttransportservices	#jbhunttransportservices	36	1	1	38	24
265	JMSmucker	SJM	$SJM	#jmsmucker	#jmsmucker	76	0	0	76	54
266	JohnsonJohnson	JNJ	$JNJ	#johnsonjohnson	#johnsonjohnson	653	60	60	773	367
267	JohnsonControlsInternational	JCI	$JCI	#johnsoncontrolsinternational	#johnsoncontrolsinternational	93	0	0	93	69
268	JPMorganChaseCo	JPM	$JPM	#jpmorganchaseco	#jpmorganchase	882	0	260	1142	636
269	JuniperNetworks	JNPR	$JNPR	#junipernetworks	#junipernetworks	78	32	32	142	76
270	KansasCitySouthern	KSU	$KSU	#kansascitysouthern	#kansascitysouthern	54	3	3	60	39
271	KelloggCo	K	$K	#kelloggco	#kellogg	7428	0	176	7604	4396
272	KeyCorp	KEY	$KEY	#keycorp	#key	150	2	4733	4885	4725
273	KimberlyClark	KMB	$KMB	#kimberlyclark	#kimberlyclark	124	12	12	148	5
274	KimcoRealty	KIM	$KIM	#kimcorealty	#kimcorealty	49	2	2	53	NA
275	KinderMorgan	KMI	$KMI	#kindermorgan	#kindermorgan	208	1216	1216	2640	1196
276	KLATencorCorp	KLAC	$KLAC	#klatencorcorp	#klatencor	60	0	1	61	NA
277	KohlsCorp	KSS	$KSS	#kohlscorp	#kohls	138	0	189	327	164
278	KraftHeinzCo	KHC	$KHC	#kraftheinzco	#kraftheinz	167	1	42	210	43
279	KrogerCo	KR	$KR	#krogerco	#kroger	557	1	296	854	252
280	LBrandsInc	LB	$LB	#lbrandsinc	#lbrands	521	0	13	534	403
281	L3CommunicationsHoldings	LLL	$LLL	#l3communicationsholdings	#l3communicationsholdings	2606	0	0	2606	2153
282	LaboratoryCorpofAmericaHolding	LH	$LH	#laboratorycorpofamericaholding	#laboratorycorpofamericaholding	96	0	0	96	79
283	LamResearch	LRCX	$LRCX	#lamresearch	#lamresearch	90	4	4	98	58
284	LeggettPlatt	LEG	$LEG	#leggettplatt	#leggettplatt	42	2	2	46	38
285	LennarCorp	LEN	$LEN	#lennarcorp	#lennar	69	0	34	103	68
286	Level3Communications	LVLT	$LVLT	#level3communications	#level3communications	64	3	3	70	39
287	LeucadiaNationalCorp	LUK	$LUK	#leucadianationalcorp	#leucadianational	43	0	0	43	26
288	LillyEliCo	LLY	$LLY	#lillyelico	#lillyeli	281	0	0	281	146
289	LincolnNational	LNC	$LNC	#lincolnnational	#lincolnnational	115	0	0	115	50
290	LKQCorporation	LKQ	$LKQ	#lkqcorporation	#lkq	27	0	4	31	24
291	LockheedMartinCorp	LMT	$LMT	#lockheedmartincorp	#lockheedmartin	426	0	112	538	335
292	LoewsCorp	L	$L	#loewscorp	#loews	15001	0	11	15012	6645
293	LowesCos	LOW	$LOW	#lowescos	#lowescos	285	0	0	285	NA
294	LyondellBasell	LYB	$LYB	#lyondellbasell	#lyondellbasell	68	5	5	78	2
295	MTBankCorp	MTB	$MTB	#mtbankcorp	#mtbank	207	0	16	223	123
296	Macerich	MAC	$MAC	#macerich	#macerich	187	4	4	195	140
297	MacysInc	M	$M	#macysinc	#macys	36555	0	451	37006	22251
298	MarathonOilCorp	MRO	$MRO	#marathonoilcorp	#marathonoil	129	0	9	138	NA
299	MarathonPetroleum	MPC	$MPC	#marathonpetroleum	#marathonpetroleum	95	2	2	99	NA
300	MarriottIntl	MAR	$MAR	#marriottintl	#marriottintl	258	0	0	258	NA
301	MarshMcLennan	MMC	$MMC	#marshmclennan	#marshmclennan	90	1	1	92	NA
302	MartinMariettaMaterials	MLM	$MLM	#martinmariettamaterials	#martinmariettamaterials	50	0	0	50	NA
303	MascoCorp	MAS	$MAS	#mascocorp	#masco	110	0	10	120	3
304	MastercardInc	MA	$MA	#mastercardinc	#mastercard	1844	0	840	2684	816
305	MattelInc	MAT	$MAT	#mattelinc	#mattel	174	1	1091	1266	1068
306	McCormickCo	MKC	$MKC	#mccormickco	#mccormick	59	0	66	125	62
307	McDonaldsCorp	MCD	$MCD	#mcdonaldscorp	#mcdonalds	492	6	2271	2769	2208
308	McKessonCorp	MCK	$MCK	#mckessoncorp	#mckesson	132	0	36	168	31
309	Medtronicplc	MDT	$MDT	#medtronicplc	#medtronic	213	0	96	309	60
310	MerckCo	MRK	$MRK	#merckco	#merck	442	2	99	543	79
311	MetLifeInc	MET	$MET	#metlifeinc	#metlife	212	0	321	533	308
312	MettlerToledo	MTD	$MTD	#mettlertoledo	#mettlertoledo	70	4	4	78	NA
313	MGMResortsInternational	MGM	$MGM	#mgmresortsinternational	#mgmresortsinternational	149	1	1	151	1
314	MichaelKorsHoldings	KORS	$KORS	#michaelkorsholdings	#michaelkorsholdings	118	0	0	118	NA
315	MicrochipTechnology	MCHP	$MCHP	#microchiptechnology	#microchiptechnology	65	0	0	65	NA
316	MicronTechnology	MU	$MU	#microntechnology	#microntechnology	815	3	3	821	1
317	MicrosoftCorp	MSFT	$MSFT	#microsoftcorp	#microsoft	2219	3	13407	15629	12327
318	MidAmericaApartments	MAA	$MAA	#midamericaapartments	#midamericaapartments	33	0	0	33	NA
319	MohawkIndustries	MHK	$MHK	#mohawkindustries	#mohawkindustries	35	6	6	47	5
320	MolsonCoorsBrewingCompany	TAP	$TAP	#molsoncoorsbrewingcompany	#molsoncoorsbrewing	135	1	0	136	97
321	MondelezInternational	MDLZ	$MDLZ	#mondelezinternational	#mondelezinternational	179	24	24	227	3
322	MonsantoCo	MON	$MON	#monsantoco	#monsanto	291	15	2498	2804	2485
323	MonsterBeverage	MNST	$MNST	#monsterbeverage	#monsterbeverage	123	0	0	123	NA
324	MoodysCorp	MCO	$MCO	#moodyscorp	#moodys	341	0	69	410	67
325	MorganStanley	MS	$MS	#morganstanley	#morganstanley	3723	90	90	3903	40
326	TheMosaicCompany	MOS	$MOS	#themosaiccompany	#themosaic	139	0	575	714	565
327	MotorolaSolutionsInc	MSI	$MSI	#motorolasolutionsinc	#motorolasolutions	48	0	5	53	4
328	MylanNV	MYL	$MYL	#mylannv	#mylannv	441	0	0	441	NA
329	NasdaqInc	NDAQ	$NDAQ	#nasdaqinc	#nasdaq	64	0	7338	7402	5361
330	NationalOilwellVarcoInc	NOV	$NOV	#nationaloilwellvarcoinc	#nationaloilwellvar	138	0	0	138	83

Company			Search words			Number of tweets				
No.	Name	Stock ticker	Ticker	Official name	Spoken name	Ticker	Official name	Spoken name	Total	Total unique
331	Navient	NAVI	$NAVI	#navient	#navient	27	21	21	69	34
332	NetApp	NTAP	$NTAP	#netapp	#netapp	97	244	244	585	291
333	NetflixInc	NFLX	$NFLX	#netflixinc	#netflix	1872	3	12029	13904	12399
334	NewellBrands	NWL	$NWL	#newellbrands	#newellbrands	100	0	0	100	47
335	NewfieldExplorationCo	NFX	$NFX	#newfieldexplorationco	#newfieldexploration	58	1	1	60	47
336	NewmontMiningCorporation	NEM	$NEM	#newmontminingcorporation	#newmontmining	181	0	7	188	86
337	NewsCorpClassA	NWSA	$NWSA	#newscorpclassa	#newscorpclassa	26	0	104	130	121
338	NewsCorpClassB	NWS	$NWS	#newscorpclassb	#newscorpclassb	64	0	104	168	19
339	NextEraEnergy	NEE	$NEE	#nexteraenergy	#nexteraenergy	143	12	12	167	69
340	NielsenHoldings	NLSN	$NLSN	#nielsenholdings	#nielsenholdings	32	0	0	32	23
341	Nike	NKE	$NKE	#nike	#nike	605	24365	24365	49335	23475
342	NiSourceInc	NI	$NI	#nisourceinc	#nisource	381	0	4	385	288
343	NobleEnergyInc	NBL	$NBL	#nobleenergyinc	#nobleenergy	61	0	6	67	54
344	Nordstrom	JWN	$JWN	#nordstrom	#nordstrom	227	1369	1369	2965	1468
345	NorfolkSouthernCorp	NSC	$NSC	#norfolksoutherncorp	#norfolksouthern	63	0	42	105	86
346	NorthernTrustCorp	NTRS	$NTRS	#northerntrustcorp	#northerntrust	85	0	15	100	62
347	NorthropGrummanCorp	NOC	$NOC	#northropgrummancorp	#northropgrumman	132	0	18	150	68
348	NRGEnergy	NRG	$NRG	#nrgenergy	#nrgenergy	52	0	0	52	33
349	NucorCorp	NUE	$NUE	#nucorcorp	#nucor	84	0	15	99	53
350	NvidiaCorporation	NVDA	$NVDA	#nvidiacorporation	#nvidia	1344	0	2212	3556	2736
351	OReillyAutomotive	ORLY	$ORLY	#oreillyautomotive	#oreillyautomotive	111	0	0	111	41
352	OccidentalPetroleum	OXY	$OXY	#occidentalpetroleum	#occidentalpetroleum	169	0	0	169	72
353	OmnicomGroup	OMC	$OMC	#omnicomgroup	#omnicom	77	2	7	86	72
354	ONEOK	OKE	$OKE	#oneok	#oneok	73	3	3	79	34
355	OracleCorp	ORCL	$ORCL	#oraclecorp	#oracle	402	0	2579	2981	2714
356	PACCARInc	PCAR	$PCAR	#paccarinc	#paccar	89	0	11	100	43
357	PackagingCorporationofAmerica	PKG	$PKG	#packagingcorporationofamerica	#packagingcorporationofamerica	268	0	0	268	152
358	ParkerHannifin	PH	$PH	#parkerhannifin	#parkerhannifin	701	7	7	715	287
359	PattersonCompanies	PDCO	$PDCO	#pattersoncompanies	#pattersoncompanies	27	0	0	27	19
360	PaychexInc	PAYX	$PAYX	#paychexinc	#paychex	59	0	10	69	47
361	PayPal	PYPL	$PYPL	#paypal	#paypal	222	8443	8443	17108	7433
362	PentairLtd	PNR	$PNR	#pentairltd	#pentair	145	0	7	152	109
363	PeoplesUnitedFinancial	PBCT	$PBCT	#peoplesunitedfinancial	#peoplesunitedfinancial	55	0	0	55	43
364	PepsiCoInc	PEP	$PEP	#pepsicoinc	#pepsi	452	1	3812	4265	3963
365	PerkinElmer	PKI	$PKI	#perkinelmer	#perkinelmer	36	7	7	50	33
366	Perrigo	PRGO	$PRGO	#perrigo	#perrigo	94	11	11	116	61
367	PfizerInc	PFE	$PFE	#pfizerinc	#pfizer	479	1	213	693	364
368	PGECorp	PCG	$PCG	#pgecorp	#pge	50	0	84	134	108
369	PhilipMorrisInternational	PM	$PM	#philipmorrisinternational	#philipmorrisinternational	948	1	1	950	440
370	Phillips66	PSX	$PSX	#phillips66	#phillips66	170	35	35	240	101
371	PinnacleWestCapital	PNW	$PNW	#pinnaclewestcapital	#pinnaclewestcapital	32	1	1	34	28
372	PioneerNaturalResources	PXD	$PXD	#pioneernaturalresources	#pioneernaturalresources	65	1	1	67	45
373	PNCFinancialServices	PNC	$PNC	#pncfinancialservices	#pncfinancialservices	154	0	0	154	62
374	PoloRalphLaurenCorp	RL	$RL	#poloralphlaurencorp	#poloralphlauren	362	0	806	1168	956
375	PPGIndustries	PPG	$PPG	#ppgindustries	#ppgindustries	75	13	13	101	62
376	PPLCorp	PPL	$PPL	#pplcorp	#ppl	121	0	687	808	732
377	PraxairInc	PX	$PX	#praxairinc	#praxair	193	9	11	213	100
378	PricelinecomInc	PCLN	$PCLN	#pricelinecominc	#pricelinecom	482	0	21	503	175
379	PrincipalFinancialGroup	PFG	$PFG	#principalfinancialgroup	#principalfinancial	61	10	1	72	49
380	ProcterGamble	PG	$PG	#proctergamble	#proctergamble	1244	27	27	1298	485
381	ProgressiveCorp	PGR	$PGR	#progressivecorp	#progressive	54	0	2585	2639	2580
382	Prologis	PLD	$PLD	#prologis	#prologis	52	4	4	60	27
383	PrudentialFinancial	PRU	$PRU	#prudentialfinancial	#prudentialfinancial	134	30	30	194	82
384	PublicServEnterpriseInc	PEG	$PEG	#publicserventerpriseinc	#publicserventerprise	98	0	0	98	72
385	PublicStorage	PSA	$PSA	#publicstorage	#publicstorage	121	4	4	129	44
386	PulteHomesInc	PHM	$PHM	#pultehomesinc	#pultehomes	102	0	1	103	1
387	PVHCorp	PVH	$PVH	#pvhcorp	#pvh	100	1	14	115	62
388	Qorvo	QRVO	$QRVO	#qorvo	#qorvo	73	3	3	79	49
389	QuantaServicesInc	PWR	$PWR	#quantaservicesinc	#quantaservices	32	0	2	34	24
390	QUALCOMMInc	QCOM	$QCOM	#qualcomminc	#qualcomm	511	0	274	785	425
391	QuestDiagnostics	DGX	$DGX	#questdiagnostics	#questdiagnostics	49	10	10	69	4
392	QuintilesIMSHoldingsInc	Q	$Q	#quintilesimsholdingsinc	#quintilesimsholdings	4371	0	0	4371	1715
393	RangeResourcesCorp	RRC	$RRC	#rangeresourcescorp	#rangeresources	60	0	5	65	44
394	RaymondJamesFinancialInc	RJF	$RJF	#raymondjamesfinancialinc	#raymondjamesfinancial	48	0	0	48	39
395	RaytheonCo	RTN	$RTN	#raytheonco	#raytheon	212	0	66	278	138
396	RealtyIncomeCorporation	O	$O	#realtyincomecorporation	#realtyincome	21223	0	3	21226	11029
397	RedHatInc	RHT	$RHT	#redhatinc	#redhat	141	0	811	952	874
398	RegencyCentersCorporation	REG	$REG	#regencycenterscorporation	#regencycenters	306	0	5	311	1
399	Regeneron	REGN	$REGN	#regeneron	#regeneron	240	7	7	254	4
400	RegionsFinancialCorp	RF	$RF	#regionsfinancialcorp	#regionsfinancial	164	0	0	164	99
401	RepublicServicesInc	RSG	$RSG	#republicservicesinc	#republicservices	49	0	6	55	42
402	ResMed	RMD	$RMD	#resmed	#resmed	44	12	12	68	48
403	RobertHalfInternational	RHI	$RHI	#roberthalfinternational	#roberthalfinternational	33	0	0	33	27
404	RockwellAutomationInc	ROK	$ROK	#rockwellautomationinc	#rockwellautomation	112	0	11	123	86
405	RockwellCollins	COL	$COL	#rockwellcollins	#rockwellcollins	283	10	10	303	3
406	RoperTechnologies	ROP	$ROP	#ropertechnologies	#ropertechnologies	36	0	0	36	29
407	RossStores	ROST	$ROST	#rossstores	#rossstores	112	2	2	116	41
408	RoyalCaribbeanCruisesLtd	RCL	$RCL	#royalcaribbeancruisesltd	#royalcaribbeancruises	79	0	0	79	49
409	Salesforcecom	CRM	$CRM	#salesforcecom	#salesforcecom	675	4	3061	3740	2924
410	SBACommunicationsCorp	SBAC	$SBAC	#sbacommunicationscorp	#sbacommunications	30	0	2	32	28
411	SCANACorp	SCG	$SCG	#scanacorp	#scana	57	0	5	62	40
412	SchlumbergerLtd	SLB	$SLB	#schlumbergerltd	#schlumberger	133	0	30	163	98
413	ScrippsNetworksInteractiveInc	SNI	$SNI	#scrippsnetworksinteractiveinc	#scrippsnetworksinteractive	64	0	2	66	47
414	SeagateTechnology	STX	$STX	#seagatetechnology	#seagatetechnology	159	1	1	161	65
415	SealedAir	SEE	$SEE	#sealedair	#sealedair	115	6	6	127	62

Company			Search words			Number of tweets				
No.	Name	Stock ticker	Ticker	Official name	Spoken name	Ticker	Official name	Spoken name	Total	Total unique
416	SempraEnergy	SRE	$SRE	#sempraenergy	#sempraenergy	124	0	0	124	59
417	SherwinWilliams	SHW	$SHW	#sherwinwilliams	#sherwinwilliams	99	23	23	145	57
418	SignetJewelers	SIG	$SIG	#signetjewelers	#signetjewelers	132	2	2	136	93
419	SimonPropertyGroupInc	SPG	$SPG	#simonpropertygroupinc	#simonproperty	222	0	1	223	71
420	SkyworksSolutions	SWKS	$SWKS	#skyworkssolutions	#skyworkssolutions	120	1	1	122	58
421	SLGreenRealty	SLG	$SLG	#slgreenrealty	#slgreenrealty	53	1	1	55	39
422	SnapOnInc	SNA	$SNA	#snaponinc	#snapon	1523	0	52	1575	940
423	SouthernCo	SO	$SO	#southernco	#southern	2907	1	1410	4318	3780
424	SouthwestAirlines	LUV	$LUV	#southwestairlines	#southwestairlines	245	241	241	727	210
425	SPGlobalInc	SPGI	$SPGI	#spglobalinc	#spglobal	95	0	5	100	1
426	StanleyBlackDecker	SWK	$SWK	#stanleyblackdecker	#stanleyblackdecker	179	4	4	187	50
427	StarbucksCorp	SBUX	$SBUX	#starbuckscorp	#starbucks	678	0	5101	5779	5097
428	StateStreetCorp	STT	$STT	#statestreetcorp	#statestreet	134	0	39	173	89
429	StericycleInc	SRCL	$SRCL	#stericycleinc	#stericycle	37	0	5	42	34
430	StrykerCorp	SYK	$SYK	#strykercorp	#stryker	78	0	30	108	73
431	SunTrustBanks	STI	$STI	#suntrustbanks	#suntrustbanks	73	8	8	89	53
432	SymantecCorp	SYMC	$SYMC	#symanteccorp	#symantec	147	0	154	301	211
433	SynchronyFinancial	SYF	$SYF	#synchronyfinancial	#synchronyfinancial	112	3	3	118	49
434	SynopsysInc	SNPS	$SNPS	#synopsysinc	#synopsys	51	0	10	61	46
435	SyscoCorp	SYY	$SYY	#syscocorp	#sysco	99	0	7	106	40
436	TRowePriceGroup	TROW	$TROW	#trowepricegroup	#troweprice	58	0	4	62	46
437	TargetCorp	TGT	$TGT	#targetcorp	#target	620	0	2251	2871	2410
438	TEConnectivityLtd	TEL	$TEL	#teconnectivityltd	#teconnectivity	125	0	4	129	88
439	TechnipFMC	FTI	$FTI	#technipfmc	#technipfmc	37	2	2	41	1
440	TexasInstruments	TXN	$TXN	#texasinstruments	#texasinstruments	177	42	42	261	103
441	TextronInc	TXT	$TXT	#textroninc	#textron	269	0	18	287	96
442	ThermoFisherScientific	TMO	$TMO	#thermofisherscientific	#thermofisherscientific	134	4	4	142	69
443	TiffanyCo	TIF	$TIF	#tiffanyco	#tiffany	174	18	6780	6972	6857
444	TimeWarnerInc	TWX	$TWX	#timewarnerinc	#timewarner	380	2	185	567	380
445	TJXCompaniesInc	TJX	$TJX	#tjxcompaniesinc	#tjxcompanies	101	1	0	102	49
446	TorchmarkCorp	TMK	$TMK	#torchmarkcorp	#torchmark	29	0	2	31	20
447	TotalSystemServices	TSS	$TSS	#totalsystemservices	#totalsystemservices	38	1	1	40	33
448	TractorSupplyCompany	TSCO	$TSCO	#tractorsupplycompany	#tractorsupply	213	0	27	240	218
449	TransDigmGroup	TDG	$TDG	#transdigmgroup	#transdigm	64	0	3	67	53
450	TheTravelersCompaniesInc	TRV	$TRV	#thetravelerscompaniesinc	#thetravelerscompanies	170	0	0	170	122
451	TripAdvisor	TRIP	$TRIP	#tripadvisor	#tripadvisor	151	453	453	1057	531
452	TwentyFirstCenturyFoxClassA	FOXA	$FOXA	#twentyfirstcenturyfoxclassa	#twentyfirstcenturyfoxclassa	158	0	0	158	NA
453	TwentyFirstCenturyFoxClassB	FOX	$FOX	#twentyfirstcenturyfoxclassb	#twentyfirstcenturyfoxclassb	306	0	0	306	NA
454	TysonFoods	TSN	$TSN	#tysonfoods	#tysonfoods	145	17	17	179	90
455	UDRInc	UDR	$UDR	#udrinc	#udr	31	0	16	47	37
456	UltaSalonCosmeticsFragranceInc	ULTA	$ULTA	#ultasaloncosmeticsfragranceinc	#ultasaloncosmeticsfragrance	175	0	0	175	102
457	USBancorp	USB	$USB	#usbancorp	#usban	168	4	9	181	92
458	UnderArmourClassC	UA	$UA	#underarmourclassc	#underarmourclassc	1113	0	0	1113	767
459	UnderArmourClassA	UAA	$UAA	#underarmourclassa	#underarmourclassa	116	0	0	116	NA
460	UnionPacific	UNP	$UNP	#unionpacific	#unionpacific	166	13	13	192	84
461	UnitedContinentalHoldings	UAL	$UAL	#unitedcontinentalholdings	#unitedcontinentalholdings	361	1	1	363	1
462	UnitedHealthGroupInc	UNH	$UNH	#unitedhealthgroupinc	#unitedhealth	233	0	21	254	116
463	UnitedParcelService	UPS	$UPS	#unitedparcelservice	#unitedparcelservice	178	18	18	214	84
464	UnitedRentalsInc	URI	$URI	#unitedrentalsinc	#unitedrentals	76	0	3	79	54
465	UnitedTechnologies	UTX	$UTX	#unitedtechnologies	#unitedtechnologies	166	11	11	188	86
466	UniversalHealthServicesInc	UHS	$UHS	#universalhealthservicesinc	#universalhealthservices	61	1	0	62	39
467	UnumGroup	UNM	$UNM	#unumgroup	#unum	36	0	27	63	54
468	VFCorp	VFC	$VFC	#vfcorp	#vf	124	0	144	268	196
469	ValeroEnergy	VLO	$VLO	#valeroenergy	#valeroenergy	154	0	0	154	59
470	VarianMedicalSystems	VAR	$VAR	#varianmedicalsystems	#varianmedicalsystems	56	0	0	56	43
471	VentasInc	VTR	$VTR	#ventasinc	#ventas	56	0	88	144	125
472	VerisignInc	VRSN	$VRSN	#verisigninc	#verisign	31	0	18	49	39
473	VeriskAnalytics	VRSK	$VRSK	#veriskanalytics	#veriskanalytics	36	2	2	40	31
474	VerizonCommunications	VZ	$VZ	#verizoncommunications	#verizoncommunications	642	5	5	652	384
475	VertexPharmaceuticalsInc	VRTX	$VRTX	#vertexpharmaceuticalsinc	#vertexpharmaceuticals	119	0	0	119	62
476	ViacomInc	VIAB	$VIAB	#viacominc	#viacom	110	1	139	250	206
477	VisaInc	V	$V	#visainc	#visa	13231	0	1710	14941	8254
478	VornadoRealtyTrust	VNO	$VNO	#vornadorealtytrust	#vornadorealtytrust	49	2	2	53	NA
479	VulcanMaterials	VMC	$VMC	#vulcanmaterials	#vulcanmaterials	54	2	2	58	1
480	WalMartStores	WMT	$WMT	#walmartstores	#walmartstores	776	1	1	778	333
481	WalgreensBootsAlliance	WBA	$WBA	#walgreensbootsalliance	#walgreensbootsalliance	263	10	10	283	126
482	TheWaltDisneyCompany	DIS	$DIS	#thewaltdisneycompany	#thewaltdisney	1009	5	765	1779	762
483	WasteManagementInc	WM	$WM	#wastemanagementinc	#wastemanagement	1217	0	255	1472	496
484	WatersCorporation	WAT	$WAT	#waterscorporation	#waters	134	0	154	288	236
485	WecEnergyGroupInc	WEC	$WEC	#wecenergygroupinc	#wecenergy	37	0	0	37	28
486	WellsFargo	WFC	$WFC	#wellsfargo	#wellsfargo	1050	1474	1474	3998	1842
487	WelltowerInc	HCN	$HCN	#welltowerinc	#welltower	101	0	0	101	37
488	WesternDigital	WDC	$WDC	#westerndigital	#westerndigital	251	67	67	385	158
489	WesternUnionCo	WU	$WU	#westernunionco	#westernunion	59	0	52	111	88
490	WestRockCompany	WRK	$WRK	#westrockcompany	#westrock	40	0	4	44	32
491	WeyerhaeuserCorp	WY	$WY	#weyerhaeusercorp	#weyerhaeuser	486	0	4	490	245
492	WhirlpoolCorp	WHR	$WHR	#whirlpoolcorp	#whirlpool	52	2	150	204	183
493	WilliamsCos	WMB	$WMB	#williamscos	#williamscos	152	0	0	152	NA
494	WillisTowersWatson	WLTW	$WLTW	#willistowerswatson	#willistowerswatson	32	9	9	50	39
495	WyndhamWorldwide	WYN	$WYN	#wyndhamworldwide	#wyndhamworldwide	347	0	0	347	NA
496	WynnResortsLtd	WYNN	$WYNN	#wynnresortsltd	#wynnresorts	302	0	9	311	NA
497	XcelEnergyInc	XEL	$XEL	#xcelenergyinc	#xcelenergy	80	0	12	92	63
498	XeroxCorp	XRX	$XRX	#xeroxcorp	#xerox	58	0	163	221	190
499	XilinxInc	XLNX	$XLNX	#xilinxinc	#xilinx	101	0	31	132	93
500	XLCapital	XL	$XL	#xlcapital	#xlcapital	1511	0	0	1511	610
501	XylemInc	XYL	$XYL	#xyleminc	#xylem	45	0	10	55	47
502	YumBrandsInc	YUM	$YUM	#yumbrandsinc	#yumbrands	220	0	8	228	131
503	ZimmerBiometHoldings	ZBH	$ZBH	#zimmerbiometholdings	#zimmerbiometholdings	95	0	0	95	29
504	ZionsBancorp	ZION	$ZION	#zionsbancorp	#zionsban	67	0	0	67	53
505	Zoetis	ZTS	$ZTS	#zoetis	#zoetis	85	6	6	97	48

Yrityksen arvon kehityksen ennustaminen tviittauksien lukumäärällä

Sosiaalisen median rooli maailmassa kasvaa jatkuvasti ja se sisältää suuria määriä informaatiota. Tätä informaatiota ja sen metatietoja voidaan hyödyntää liiketoiminnassa. Twitteristä on saatavilla taloustietoa, kun finanssialan henkilöt tviittaavat huomaamistaan ilmiöistä sekä mielipiteistään.

Tviittien lukumäärä selittää yritysten tuottoa

Tutkimuksen tulosten mukaan finanssialan ammattilaisten työpäivän kuluessa lähettämien tviittien lukumäärän kasvaessa myös todennäköisyys yrityksen arvonnousun 0 ja 1 % välillä kasvaa. Kvantitatiivisella analyysilla voidaan siis ennustaa osakkeen arvostusta.

Aiemmin sentimenttianalyysi eli tunteen mittaaminen tviiteistä on osoittanut seuraavaa: mitä enemmän tunnetta tviiteissä on, niin sitä todennäköisemmin tuotot ovat negatiivisia seuraavana päivänä. Merkittävimpiä vaikuttajia ovat sana ”pelko”, ”toivo” ja ”rauhallisuus”. Tutkimuksissa on myös onnistuttu ennustamaan sentimenttianalyysillä seuraavan päivän arvon kehityksen suunta jopa 87.6 % tarkkuudella.

Tulosten merkittävyys ja näkymät

Tämä tutkimus ei vahvista aiempaa kvantitatiivisen tutkimuksen tulosta, jossa havaittiin työajan ulkopuolella lähetettyjen tviittien selittävän seuraavana päivänä tapahtunutta arvon nousua 1-5 prosentilla.

Päivätviittien kvantitatiivista analyysia voi hyödyntää automaattisessa kaupankäynnissä. Se vaatii kuitenkin ”big datan” käsittelyä reaaliajassa, mikä vaatii merkittävää laskentatehoa ja tallennustilaa. Analyysin metodeja täytyy vielä kehittää ennen kuin sen tuotot ylittävät vaadittavalla tasolla sen kustannukset.

Muiden kuin finanssialan henkilöiden lähettämien tviittien lukumäärällä ei havaittu olevan vaikutusta yritysten markkina-arvoon.

Tutkimuksen metodeista

Tutkimuksessa kerättiin 365 miljoonaa englanninkielistä tviittiä, jotka julkaistiin 1.6.2016 ja 30.6.2017 välillä. Otos koostui 706700 tviitistä joissa mainittiin S&P 500 yritys. Tutkimukseen valittiin 42 yritystä, joista tviitattin keskimäärin yli 15 kertaa päivässä, jolloin otokseen jäi 506368 tviittiä.

Tviitit jaettiin niiden julkaisuajankohdan mukaan kahteen kategoriaan. Päivätviitit on lähetetty pörssin ollessa auki ja yötviitit pörssin ollessa kiinni. Lisäksi tviitit jaettiin kuinka yritys mainittiin tviitissä: 1) osakesymbolin avulla 2) yrityksen virallinen nimi tai 3) yrityksen puhekielinen nimi. Ensimmäisen ryhmän viestien oletettiin tulevan finanssialan ammattilaisilta.

Vaikutus tutkittiin regressioanalyysillä jossa selittävänä tekijänä oli tviittien lukumäärä päivässä ja selitettävänä logistinen malli, joka sai arvon yrityksen tuoton perusteella. Tutkimuksessa oli kolme logistista mallia. Ensimmäisessä malli sai arvon 1, jos tuotto oli positiivinen. Toinen malli sai arvon 1, jos tuotto oli 0 ja 1 % välillä. Kolmas malli sai arvon 1, jos tuotto oli 1-5 %. Muuten mallit saivat 0.