HØGSKOLEN I OSLO
OG AKERSHUS

Fredrik Benjaminsen

Fredrik Husevåg

The Effect of Twitter Attention on Earnings Announcements

An Event Study Investigating the Relation Between Twitter Attention, by Volume and Sentiment, and Market Efficiency

Masteroppgave i økonomi og administrasjon

Handelshøyskolen ved HiOA

2017

Abstract

In our thesis, we use event study to analyze relations between Twitter attention and technology companies on the Nasdaq stock exchange. We divide attention into Twitter sentiment and volume. The classification of sentiment is done with supervised machine learning algorithms. The effects of Twitter sentiment are consistent with the Efficient Market Hypothesis. Twitter volume on the other hand, indicates an attention effect related to abnormal market behavior, mainly in larger well-known companies. We believe an indicator as to how individual investors pick stocks, is to distinguish whether a firm's operations are known to the investors, and how abnormal tweet volume affect traders' attention towards them.

Handelshøyskolen ved HiOA

Oslo 2017

Table of content

1 Introduction

In today's market, searching for stocks to purchase is not an easy task. There are countless of investment opportunities, and whereas professional traders have more time and resources to monitor a wide range of stocks, individual investors must use other measures to decide. Barber and Odean (2008) argue that individual investors solve their search problem by purchasing stocks based on information that catches their attention.

In contribution to Barber and Odean's (2008) findings, we seek to develop the studies on how attention influences the market. When conducting this research, we hope to answer the question: Does Twitter, as an attention source, have an effect on market efficiency?

We apply event study to measure the financial impact of earnings announcements connected to Twitter attention. We separate Twitter attention into two types; abnormal tweet volume and abnormal tweet sentiment. The ten companies included in our research is divided into two groups. The first group consists of four famous companies experiencing a lot of attention naturally. The other six companies are less famous, and experience respectively less attention. All ten are represented on the Nasdaq stock exchange and gets mentioned hourly on Twitter. The data retrieved consists of stock prices and tweets mentioning our companies by ticker name from January 1. 2014 to December 31. 2016.

Our results show that abnormal Twitter volume attention results in significant abnormal returns up to several days after a negative event. These effects are experienced in connection to the group of the famous companies. We cannot find any effects of abnormal Twitter sentiment attention. For the less famous companies there are no effects of attention at all, the same goes for all positive events researched in our study.

We conclude that Twitter attention alone is not enough to experience an effect on abnormal returns in the days after an event. The attention provided by the famous companies in combination with Twitter volume is essential to be able to draw conclusions around negative events.

Our findings contribute towards the financial attention theory presented in Barber and Odean (2008), but we believe further studies are essential before we can conclude that Twitter attention influences the market.

The thesis starts with a summary of related literature, highlighting the most relevant theories for understanding the essential parts of our study and findings. The methods we use to retrieve and analyze Twitter information is described in the following chapter, before we provide a detailed overview of our data. The results chapter discuss our findings, and the thesis ends with our key results being summed up in the conclusion.

2 Related literature

In this chapter, we provide literature findings from related studies towards the analyses of Twitter. Afterwards we explain how the Efficient Market Hypothesis (EMH) is an essential part of our thesis used as the basis for measuring our findings through event study. The key deductions in EMH towards our issue will be presented in its own subchapter. When experiencing results deviating from EMH, we turn to the field of behavioral finance and how herd behavior and over- and underreaction can explain inefficiency in the market.

2.1 Twitter studies

Three major categories of data are considered when investigating the relations between web data and financial markets; web news, search engine queries and social media (Ranco, et al. 2005). The latter falls into our category of study. The research on social media's relations to financial forecasting has evolved in popularity alongside the evolution of machine power. Bollen, Mao and Zeng (2010) is recognized to be the first to investigate the connection between Twitter and the financial market, where their results show that it is possible to predict the market to some extent with mood indicators from daily Twitter feeds.

Mao, Counts and Bollen (2011) and Zheludev, Smith and Aste (2014) have published research on the same field by exploring the relations between mood indicators on Twitter and Dow Jones Industrial Average (DJIA), where they find significant relations. There have also been research studying tweet frequency in connection to S&P 500 returns (Mao, et al. 2012), where stock market prediction models achieve greater predictive results with Twitter frequency as a part of the model.

Gabrovsek, et al. (2017) research Twitter volume and sentiment of the 30 companies on DJIA focusing on earnings announcements. The day before the announcement, Twitter volume is low and sentiment values have weak predictive power. Yet, they find that Twitter users sentiment reflects the stock movements on the day of the earnings statement.

2.2 Twitter sentiment

The findings in chapter 2.1 provides evidence that sentiment analyses connected to Twitter is represented as an essential part. Volume data from Twitter are researched as well, but the key findings focus on the effect of sentiment. We wish to follow this trend and investigate the relations of twitter sentiment towards our attention issue. Sentiment is defined as the settled opinion reflective of one's feelings, and the use of computational methods to deal with sentiment in a text is called sentiment analysis, or opinion mining (Pang and Lee 2008). Sentiment analysis are not necessarily done by computers, but to be able to retrieve aggregated results based on thousands of lines of text, we need a computer to be able to do it in a reasonable amount of time. The computational study and understanding of text is called natural language processing (NLP), and the research on this field aim to understand how human beings process text to be able to make machines perform desired tasks with the natural language (Chowdhury 2003).

Since we perform sentiment analysis on tweets for classification purposes, it is useful to understand where the theory behind the process. NLP together with the use of machine learning (ML) forms the basis of sentiment analysis. An early definition of ML by Samuel (1959), one of the pioneers on the field, is that it gives computers the ability to learn without being explicitly programmed. It is a set of algorithms constructed to learn from and make prediction by using historical data (Mohri, Rostamizadeh and Talwalker 2014). They point out a key factor to achieve better results with ML algorithms is to obtain quality training data.

When using ML algorithms for sentiment analysis the training data is processed with the use of NLP tools to make the quality and the format of the input as good as possible. We want to classify tweets into positive, neutral and negative, and NLP tools are useful for processing text to highlight these features. The problem of classifying financial news is that a piece of news can be good or bad without being subjective, hence there will be no sentiment connected to the sentence (Pang and Lee 2008). An example sentence could be; *the stock price rose*, which for a company would be good news, but for a machine it would be read as facts if not trained specifically for the task.

The ML algorithms are divided into three groups; supervised, unsupervised and reinforced (Murphy 2012). They all serve different purposes, and which one is best suited for a task is an individual answer given as a balance between complexity and efficiency. In all our mentioned research on Twitter sentiment, the researchers use the supervised learning method, and this is also our preferred method. When training a supervised algorithm, it needs labeled input data to recognize patterns. The data a supervised algorithm learns from is called features.

There have been a lot of research regarding sentiment, especially for reviews and longer texts (Pang, Lee and Vaithyanathan 2002). These sources of sentiment are easier to classify because the text will repeat sentiment indicative words more. The texts we want to classify, consist of maximum 140 characters. This results in fewer sentiment loaded words, and makes tweets harder to classify. Despite the lack of word count for each tweet, tweets are more frequently written than reviews, and can therefore be aggregated easier. Another difference between reviews compared to tweets are their intent. Reviews are written to summarize the view of an author thoughts, while tweets are normally more casual and their purpose is not always to express feelings toward a subject (Go, Bhayani and Huang 2009). Especially, when looking at financial tweets, we observe a lot of facts regarding prices that are neither positive nor negative, only facts.

Gabrovsek, et al. (2017) use financial experts to label their Twitter training set to match their sentiment classes towards financial reactions and not subjectivity. Meaning that the example sentence; *The stock price rose*, will be regarded as positive instead of neutral by their algorithms. This differs from our subjective sentiment training set, which gives us an algorithm classifying subjective, and irrational, reflections instead of rational financial sentiment.

Go, Bhayani and Huang (2009) use three of the most common supervised ML algorithms in their studies. They use the Naive Bayes, Maximum Entropy and Support Vector Machines algorithms to create models and patterns to recognize sentiment in a text. In the methods chapter, we consider how these methods work in practice and how we use these in our study.

2.3 Efficient market hypothesis

Twitter's mission is to give everyone the power to create and share ideas and information instantly, without barriers.[1]

As our study focus on how the effect of Twitter attention affects the market, we use Fama's (1970) classical theory on the Efficient Market Hypothesis (EMH) together with the well-researched field of event study. A financial market is defined as efficient if the reaction to security prices after an announcement are immediate, accurate and in the right direction with no subsequent price trends (Fama 1970). Given this definition we apply event study,

[1] https://about.twitter.com/company - Twitters vision statement expressed on their web page.

standardized by MacKinlay (1997), restricted to only include events experiencing abnormal Twitter attention to measure if a set of events is acting according to EMH. If we observe reactions with longer consecutive return periods than one day, we might be able to infer a relation between abnormal returns after the event day and Twitter attention.

There exists a lot of research on EMH, and several methods and standards have been developed to test efficiency. One of the most common ways to test the hypothesis is using the event study method with earnings announcements as financial events. Earnings announcements are official public statement of a company's financial performance or profitability for a specific time period, and are used extensively when forecasting future performance and valuations of equity (Mlonzi, Kruger and Nthoesane 2011). As Aharony and Swary (1980) states, these announcements carry inside information about companies' prospects. Lonie, et al. (1996) argues that earnings announcements are one of the most important signaling devices used by managers to provide essential future information to the market. On basis of these articles we use earning announcements as financial events when conducting our study.

There have also been conducted studies with contradicting results to EMH. Bloomfield, Libby and Nelson (2000) finds that stock markets both under- and overreacts to information in different periods for different securities. Khadiyala and Rau (2004) claims investors tend to underreact to new information. A study by Bhana (1995/1996) states that negative earnings announcements attract more attention in the market, which is related findings to our results.

As there have been discovered inconsistency to EMH, researchers have tried to find explanations to why. Cognitive biases are believed to be one of these explanations leading to the emerge of the Behavioral Finance (BF). De Bondt and Thaler (1985) provide evidence that supports the hypothesis of cognitive biases in finance, and study overreaction. This contradicts assumptions in EMH, namely that all investors behave rational.

2.4 Behavioral Finance

Conventional economic theory, such as the EMH, states that the market, its participants and the world in general operates as consistently rational and self-interested. It portrays humans as homo economicus, which attempts to maximize wealth by decisions based on perfect rationality

(Rittenberg and Tregarthen 2008). However, in real life, humans are constructed to fail to meet the expectations of conventional theory due to for instance cognitive biases, psychology and irrationality because of emotions. Thus, conventional economic- and finance theory also fail to explain several phenomena observed in the market. Behavioral economics and -finance on the other hand examines actual economic behavior and account for issues and phenomena traditional theory is not able to capture.

When discussing our results that deviates from the conventional EMH, we want to address whether behavior among investors buying or selling stocks based on attention is explained by BF. We include two important parts of BF theory as potential explanations.

One theory involves herd behavior experienced in financial markets. Herd behavior is defined as an individual mimicking a larger group (Cipriani and Guarino 2009). This leads to people acting in ways they normally would not. The explanation of herd behavior is divided in two. 1) Individuals feel social pressure from a group. Thus, to be able to fit into a group, they follow others. 2) It is common for a person to think a large group of people cannot make mistakes. Even though individuals find something irrational, they follow the group consensus. They believe the group must know something else that the individual itself has overlooked (Bikhchandani and Sharma 2000).

This potentially indicates that if investors experience high Twitter volume regarding an event, they can be tempted to follow the crowd and invest accordingly, resulting in consecutive days of abnormal returns.

The second financial behavior experienced related to our study of inconsistency towards EMH is over- and underreaction. De Bondt and Thaler (1985) find that investors overreact to unexpected and dramatic news events. When experiencing overreaction in the market because of Twitter, this may indicate a positive or negative significant return on the event day, followed by an opposite reaction the days after.

2.5 Irrational attention-buying vs rational selling

In Barber and Odean's (2008) article, they confirm their hypothesis that individual investors are net buyers of attention-grabbing stocks. Attention-driven buying results from the difficulty individual investors experience searching through thousands of stocks they can potentially buy. They focus on three indirect observable measures probable to cause attention towards events; news, abnormal trading volume and extreme returns (Barber and Odean 2008, 787). According to assumptions in an efficient market, this is not how rational investors act.

Bahna (1995/1996) claims that negative earnings announcements attract more attention in the market, and seen in relation to Barber and Odean's (2008) results, we might find trading patterns for several days leading to findings similar to herd behavior. Barber and Odean (2008) states that when there is attention around a stock, the individual investors buy, and believe professional investors sells. This fits with how Shiller (2003) explains how efficient market theory asserts when irrational optimists buy stocks, smart money sells.

3 Hypothesis development

High volume of Twitter messages around an event increase attention towards the event. There will always be high frequency of Twitter messages during events, but what happens when abnormally high volume of tweets is observed? Will this be enough attention to make attention-buyers invest?

We formalize these wonderings into our hypotheses and raise the questions:
H_1 : Does abnormal Twitter attention, in form of volume, lead to inefficiencies in the market?
H_2 : Does abnormal Twitter attention, in form of sentiment, lead to inefficiencies in the market?

Based on the discussion portrayed in chapter 2.5, we expect our findings to reflect the pattern of individual irrational attention-driven purchases after events, and rational professional investors selling to them.

4 Methods

The methods chapter explain in detail how our data is retrieved, processed and analyzed. The first subchapter describes the process retrieving Twitter data. The following chapter explains our methods for classifying tweet sentiment with machine learning algorithms, and the last subchapter describes how we use event study method to measure the effect on market efficiency.

4.1 Twitter scraper

One way to retrieve Twitter messages is by using Twitter's own Application Programming Interface (API). This API service distinct between gathering tweets in real time and historically. In 2013 Twitter commercialized their API, mainly restricting retrieval of historical tweets, ranging back longer than one week, without paying a price per kilobyte of data retrieved. Since we want to gather historical tweets dated three years back, we needed another approach to gather our desired tweets. We use a web scraper called TwitterScraper developed by Taspinar[2] to achieve our retrieval. This scraper collects user id, tweet text and time stamps for all tweets it has been set up to scrape.

After implementing the scraper in Python we need to set up the retrieval specifications. We insert the specifications as a search string (i.e., 24AAPL%20since%3A2014-01-01%20until%3A2016-12-31 for retrieving tweets with the word $AAPL for the period 01.01.2014 to 31.12.2016). Here *24* represent a dollar sign, the next part of the string tells the scraper which characters to look for, and the last part of the string contains the period of which we want the search word to be retrieved from.

We create our database of tweets with a search phrase of companies' cash-tags. A cash-tag is the equivalent of the conventional hash-tag, noted with a dollar sign ($) in front, and the company's ticker name (e.g. AAPL for Apple) after. Given a search query with a cash-tag, the scraper retrieve tweets by *scrolling* (i.e. scrolling down the news feed on Twitter, and for every 20 tweets, the browser needs to refresh the page to show older tweets) through the content of the search query while storing tweets. Since the search string is limited to cash-tags for the companies, and not the company name, most tweets will be in a financial context. This makes the tweets more informative, and as a result, there are less positive or negative sentiment texts.

[2] https://github.com/taspinar/twitterscraper

4.2 Machine Learning Sentiment Analysis

For our text classification, we use three supervised machine learning algorithms to train a model that classifies text into either positive, negative or neutral sentiment. The algorithms used are; Naive Bayes (NB), Maximum Entropy (ME) and Support Vector Machines (SVM). We will in the following subchapters explain what supervised machine learning are, how we need to preprocess and standardize data to fit in the supervised algorithms, and how the three different algorithms work.

4.2.1 Supervised Machine Learning

The supervised learning method is performed with a supervisor that hands the machine labelled input to train with. The machine read the input, look at the feature, and then tries to find the most applicable pattern or cluster to an existing model. The features are characteristics we want the machine to look for when recognizing patterns or clusters. In our case this feature is words. After the machine has trained itself to recognize a pattern or grouping clusters by using these features, the machine builds a model with the trained features to predict an outcome or classify new information. This is fed into the model as input, and comes out as a classified or grouped results, called output (Murphy 2012). In our case information is a tweet regarding a company, and the output will either be positive, neutral or negative.

The general supervised learning model can be written mathematically to concretize it better. A feature vector, x, is input data. The function that gives the ground truth of the variable x is given by $y = f(x)$. Given a specific data, x, an output, y, is received. The output is, in the case of a few discrete values, the classifier. In the case of many values, or a natural order of values, a regressor. When building a ML algorithm, the key object is to minimize the cost of the model. This cost is defined by how much the classifier or regressor alternates from the truth. At the same time, we do not want an overly complex model. If this happens the model will match its training data, but will not be general enough to make good predictions or classifications. This is called overfitting and would make the models predictions poor (Murphy 2012).

4.2.2 Pre-processing

To use a training set with a supervised ML algorithm, pre-processing and standardizing information is necessary. If done correctly the results from the ML algorithms will be predicted with similar accuracy as the training set. Another reason for text processing is to exclude data that does not contribute towards the model's accuracy. In the area of Natural Language Processing (NLP), pre-processing is essential for the algorithms to be able to extract the right features of words describing sentiment factors.

The pre-processing can be done in different ways. We chose to follow the method Go, Bhayani and Huang (2009) use to classify sentiment for tweets in their study. The features in our algorithms will be words from a Twitter text. To separate these words as their own objects, we use the function Tokenize. This function separates all words into objects by recognizing white spaces as splitting signals, and by doing this, we can score each word with its own sentiment factor. Next, we remove all stop words. These words are short words like: *the, is, it, at, and, for, on, etc*. that do not add any sentiment to a sentence. For a human, these stop words help add context and flow, but not any sentiment. For our purpose of training a machine classifier we discard these words to give the words that truly matters more attention.

Usernames, symbols and links from Twitter messages are removed. Usernames do not bring sentiment information, and as the standard way of typing a username in a tweet is by adding an @ in front, we simply remove all words starting with @. Hashtags, words starting with (#), on the other hand could contain feelings (i.e., #Love). We can therefore not remove the word behind the hashtag, but we can remove the hashtag symbol itself, since the symbol does not contribute to sentiment. This is the same for all other symbols like: $, ", /, &, !, etc.

Characters in words that are written consecutively more than two times are removed and replaced with a double character instead. An example could be; "Microsoft's stocks are rising through the rooooof". In this case *rooooof* will be shortened to *roof*.

The last processing step is called stemming. This process shortens words to its root form. It is most easily explained by an example. If we have the word *running* in a text, the stemming function would reduce the word to *run*. The sentiment of the sentence would be the same, but the number of character is reduced.

After the pre-processing is done, the training set is ready to be fed into the algorithms. The three algorithms we use, all take words as features, which means we can use a standard pre-processing procedure for all our algorithms.

4.2.3 The Naive Bayes

The Naive Bayes method is one of the most used models in supervised learning. It is based on Bayes theorem, and is a probabilistic model (Go, Bhayani and Huang 2009). It is not computational expensive, which makes it possible for normal computers to do the training of the algorithm on big sets of training data without breaking down. The general model is written as:

$$P_{NB}(c|\mathrm{d}) \coloneqq \frac{(P(c)\sum_{i=1}^{m} P(f|\mathrm{c})^{n_i(d)})}{P(d)} \tag{1}$$

Where, in the case of sentiment classification training, *P(c|d)* is the probability of a positive, neutral or negative sentiment *c*, given a tweet *d*, from the training set. The f represents our words feature, and $n_i(d)$ represents the number of different word observations in a tweet *d*. The number of different features are denoted as *m*, but we only have one feature type, words, in this model. The Naive Bayes model assumes independency between its features (Mitchell 2016).

To optimize equation (1), the converging equation presented by Mitchel (2016) is used:

$$C_{NB} = argmaxP(C) * \Pi_i P(d_i|\mathrm{C}) \tag{2}$$

Here $P(d_i|\mathrm{C})$ is the conditional probability of the word i, belongs in the sentiment class *C*. When the algorithm is optimized based on the training set, each word in the training set contains a score, and we use these scores to predict sentiment to new information.

4.2.4 Maximum Entropy

The logic behind the Maximum Entropy (ME) model is to take a known event and maximize the uncertainty in this event (the entropy) as uniformly as possible while meeting a set of constraints known as evidence (Yin and Xi 2016). Maximizing the entropy, making the features that is uncertain as uniformly weighted as possible, elevates the certain features. This is done by adding higher weights to the certain factors given by the constraints (Nigam, Lafferty and McCallum 1999). When given enough training data, the features will be separated into categories that will become more certain (or uncertain) for each training set. The formula for the ME model is given as:

$$H(p) = -\sum p(a, b) \log p(a, b) \tag{3}$$

Where a is the feature in a context b. To optimize this equation the following formula is used:

$$p = argmax\ H(p) \tag{4}$$

Which is the most likely distribution that maximizes the equation (3).

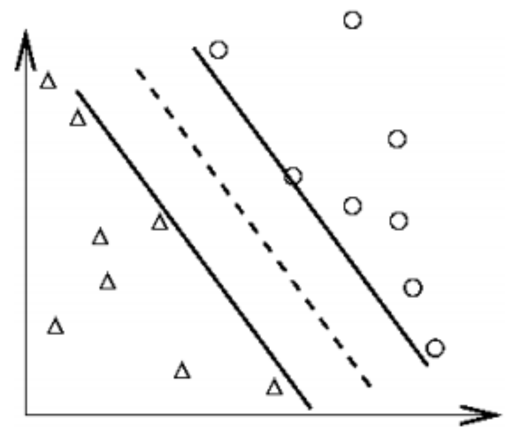
When using a ME model for text classification the feature, *a*, are words in sentence *b* which is already labeled as positive, neutral or negative. To best illustrate how the method works, we will provide an example.

The following sentence is classified as positive, and given to the ME algorithm to learn from; *I am happy.* The machine will then rate all words with 1/3 weight towards the positive sentiment constraint, because it does not know with any certainty which of the words in the sentence contribute to the positivity constraint. When the machine is given another positive sentiment sentence, this time containing the words: *She is happy,* the machine will know with higher certainty that the word happy reflects positivity in a sentence as it has been, up until now, presented more often in a positive sentence than other words. The word happy will therefore achieve a higher weight than the other two words, and be rated with a higher entropy score. Eventually the machine will have trained itself on enough labeled examples, our constraints, to be able to extract the correct sentiment from a sentence.

4.2.5 Support Vector Machines

The Support Vector Machines (SVM) algorithm is an algorithm made for classification and regression. The way a SVM algorithm works is that it denotes the presence of a feature and separate this feature data into either of two classes (Costache, Liénou and Datcu 2006).

It uses a hyperplane to separate the features, and seeks to maximize the distance from the hyperplane to the two sides of classes as shown in figure 4-1 retrieved from Hsu, Chang and Lin's (2016) article *A practicle guide to Support Vector Classification*.



Figur 4-1: SVM maximization of the margin to the hyperplane.

This hyperplane showed in figure 4-1 is given from the equation:

$$\langle \vec{w} * \vec{x} \rangle + b = \sum_i y_i \alpha_i \langle \vec{x}_i * \vec{x} \rangle + b = 0 \qquad (5)$$

$\vec{x}$ is given as n dimensional vector of input data, and y_i is the output values. $\vec{w}$ denotes the weight vector which defines the hyperplane, and α_i is the Lagrangian multipliers.

In our thesis, we use this method for text classification and the two sides are represented as the different sentiment classes. $\vec{x}$ represents the word feature from a tweet sentence, and y_i is the classification of sentiment. Since we have three sides; positive, neutral and negative, we need to introduce kernels. Kernels are another dimension that makes it possible to introduce another class. We need two kernels to solve our third class. This will introduce a 3-dimensional figure. The input data is either classified as positive or not, in the first kernel. If it is not positive, then the data is classified as either negative or not, in the second kernel. If the data is classified as not negative the data will be added to the neutral side.

4.2.6 Connecting the Algorithms

Upon putting the three algorithms together into one model that classifies the tweets, we use a voting setup to ultimately decide the sentiment. The voting method is calculated as an average between the three algorithms output. If two of the algorithms classify positive, and one algorithm classifies another sentiment score, the tweet will be classified as positive. If all three of the algorithms classify the tweet differently we have chosen the output to be neutral. When running our algorithms we get an accuracy of 73% correct classifications towards our testing data. From Pang and Lee's (2008) article our classification is approximately 13% better than humans when classifying sentiment.

4.3 Event studies

We use event study to measure the effect of Twitter attention in the market. The next subchapter will explain this study methodology, and how we use it to measure financial returns around an event.

By definition, an event study uses financial data to measure the impact of a specific event on the value of a firm (MacKinlay 1997). We define the event of interest and identify how long the period of examination around the event should be. The null hypothesis in an event study is that the event has no impact on the distribution of returns (MacKinlay 1997) and is directly attached to the underlying assumptions of EMH.

MacKinlay (1997) wrote the famous article *Event Studies in Economic and Finance* which we use as our main contribution when using this method. Over the past several decades, event studies have become an important part of finance, especially corporate finance, and used in combination with social media studies (Sprenger, et al. 2014a; Sprenger, et al. 2014b; Ranco, et al. 2005; Gabrovsek, et al. 2017).

An event can be earnings reports, new product release, mergers and acquisitions, issues of new debt or equity, and announcements of macroeconomic variables (MacKinlay 1997). We focus on quarterly earnings reports like several other research papers to be sure we have well documented basis for choosing our events.

The first model in an event study is the definition of abnormal returns a company can experience because of an event. The general model for an event study is given as:

$$AR_{i\tau} = R_{i\tau} - E(R_{i\tau}|X_\tau) \quad (6)$$

where $AR_{i\tau}$, $R_{i\tau}$ and $E(R_{i\tau}|X_\tau)$ are the abnormal, actual and normal returns for period τ. The X_τ is our conditioning information for the normal return model.

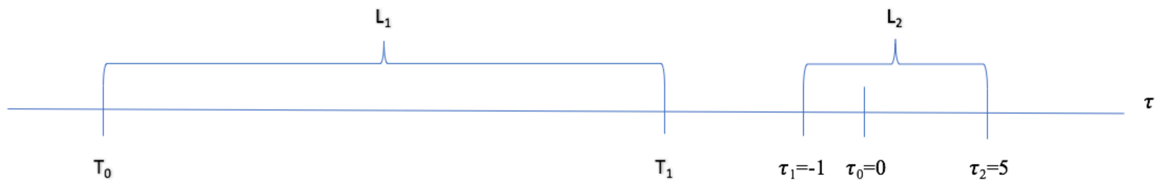
To estimate normal returns, we set a 120 days estimation window, three days prior to the quarterly earnings announcements to define normal returns during an event. The exclusion of three days prior to the event is done to not let the estimation window be influenced by the event window. For our estimation, we used the market model, defined as:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \quad (7)$$

Here, our market (m) is the NASDAQ, and the company i is one of the 10 companies represented on NASDAQ.

From the Gabrovsek, et al. (2017) article we learned that the earnings in the earnings announcement are often kept secret until the event date. In case of positive news, there might be some news might be reflected in the market the day before the event. In case of negative news, companies try to keep results hidden until the announcement, meaning negative news rarely are reacted upon in the market before the event date. To account for possible information leak the day before an event we set the window from $\tau = -1$. To observe whether firms exprience significant abnormal returns after the event, we set the end of the event window to five days after the event.

We identify time factor as τ and the event as $\tau = 0$. The estimation window L1 is 120 days denoted as T_0 to T_1, and T_1 is defined as $\tau = -3$. The event window L_2 is given as $\tau = -1$ to $\tau = 5$. We get a better overview of this looking at the figure 4-2.



Figur 4-2: Timeline for event study

The OLS estimators to calculate normal returns are given as:

$$\hat{\beta}_i = \frac{\sum_{\tau=T_0}^{T_1}(R_{i\tau} - \hat{\mu}_i)(R_{m\tau} - \hat{\mu}_m)}{\sum_{\tau=T_0}^{T_1}(R_{i\tau} - \hat{\mu}_i)^2} \tag{8}$$

where $R_{i\tau}$ and $R_{m\tau}$ is given as the firms and markets actual return, and the $\hat{\mu}_i$ and $\hat{\mu}_m$ are the firm and markets mean return during time τ and are calculated with:

$$\hat{\mu}_i = \frac{1}{L_1}\sum_{\tau=T_0}^{T_1} R_{i\tau} \tag{9}$$

And

$$\hat{\mu}_m = \frac{1}{L_1}\sum_{\tau=T_0}^{T_1} R_{m\tau} \tag{10}$$

The α parameter is given as:

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\beta}_i\hat{\mu}_m \tag{11}$$

and the variance of the model is given as:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{1}{L_1 - 2}\sum_{\tau=T_0}^{T_1}(R_{i\tau} - \hat{\alpha}_i - \hat{\beta}_i R_{m\tau})^2 \quad (12)$$

We estimate abnormal returns in events by measuring returns in the event period and subtract the estimations done above for each firm *i* at the different event dates τ:

$$AR_{i\tau} = R_{i\tau} - \hat{\alpha}_i - \hat{\beta}_i R_{m\tau} \quad (13)$$

The abnormal return under the null hypothesis, conditional on the event window market returns, will be jointly normally distributed with a zero conditional mean and conditional variance $\sigma^2(AR_{i\tau})$ (MacKinlay 1997) where:

$$\sigma^2(AR_{i\tau}) = \sigma_{\varepsilon_i}^2 + \frac{1}{L_1}\left[1 + \frac{(R_{m\tau} - \hat{\mu}_m)^2}{\hat{\sigma}_m^2}\right] \quad (14)$$

This formula consists of two components where the first component, $\sigma_{\varepsilon_i}^2$, can be found in formula (12), and the second component will cease to exist since our L_1 term is large enough to make the second component practically zero. From Pynnönen (2005) this happens when L1 increase to approximately 30 estimation days.

After identifying the abnormal returns in equation (13), we aggregate our results for each positive, negative and neutral event as an average abnormal return for each type of event by using the equation:

$$\overline{AR}_{\tau} = \frac{1}{N}\sum_{i=1}^{N} AR_{i\tau} \quad (15)$$

Where $AR_{i\tau}$ is defined as the abnormal return for each day τ for company *i*. The variance to this equation is given as:

$$var(\overline{AR}_\tau) = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{\varepsilon_i}^2 \tag{16}$$

When the equation (15) has been calculated for the three types of events, we can use these results to look at the average cumulative abnormal returns for the events by using:

$$\overline{CAR}(\tau_1\tau_2) = \sum_{\tau=\tau_1}^{\tau_2}\overline{AR}_\tau \tag{17}$$

To test the statistical significance of the aggregated average abnormal returns we need to use the following equation:

$$\theta_1 = \frac{\overline{AR}(\tau_1,\tau_2)}{var(\overline{AR}_\tau)} \sim N(0,1) \tag{18}$$

Since our test-statistics are given as a two-sided test from a normal distribution, we can indicate that the abnormal returns are significant with a 95% confidence if the absolute value of θ_1 > 1,96.

5 Data

In the data chapter, we present the data gathered and used in our study. The first part describes the companies analyzed, followed by a description of our financial data. The third subchapter explains Twitter data collected, and in the final subchapter we describe how the financial data are aligned with Twitter data.

5.1 Companies

Our study consists of ten technology companies listed on the Nasdaq stock exchange. These companies differ in market capitalization size and fame, but nonetheless relatively known firms. We distinct these organizations in two categories. 1) Well-known, famous companies; Apple, Facebook, Microsoft and Alphabet (Google), and 2) less famous firms; Intel, Nvidia, Broadcom, Activision Blizzard, Adobe Systems and Texas Instruments.

Table 5-1 present the aggregated total number of tweets and sentiment distribution for each firm. The description shows that *famous* companies are more frequently mentioned than other companies.

We argue this sorting based on subjective reasoning and from behavior of individual investors described by Barber and Odean (2008) The first group is known worldwide and investors are familiar with their products and services. The second section are also known, yet their business operations are not too familiar for investors. For instance, Texas Instruments do not only produce calculators, but are one of the leading semiconductor manufacturing companies.
It is not easy to understand all the technological specifications from their production line.

Intel, Nvidia and Activision Blizzard are infamous in the gaming society, and reviews of their products are important for their customers to know whether to buy their products. From an investor point of view, financial statements (i.e. earnings announcements) are more prominent in their assessment, than the actual products. For individual investors, news around these companies in social media might not be enough to grasp their attention because of lack of understanding the firm's operations.

This distinction is done to research whether when firms get more attention in social media and attract investors' interest, are there any consequences regarding market efficiency?

Table 5-1

Aggregated Twitter messages and sentiment distribution for each company.

Companies	Positive	Neutral	Negative	Total
Apple Inc.	71320	1290536	62180	1424036
Facebook Inc.	40486	690093	25826	756405
Alphabet Inc.	17256	405187	9462	431905
Microsoft Corp.	9980	267237	6398	283615
Intel Corp.	4403	163467	2633	170503
Nvidia Corp.	2520	84148	1773	88441
Broadcom Ltd.	749	53300	467	54516
Activision Blizzard	1533	49712	739	51984
Adobe Systems Inc.	898	45145	546	46589
Texas Instruments Inc.	346	30385	255	30986
Total	149491	3079210	110279	3338980

Table 5-1 gives a summary of the aggregated Twitter messages for each company, collected from Jan. 1, 2014 to Dec 31, 2016. The positive, neutral and negative columns describe the aggregated frequencies of tweets for the respective sentiment, classified by the sentiment algorithms described in chapter 4.

5.2 Financial data

We retrieve the financial data from Yahoo Finance[3], and consist of daily stock prices for the respective companies, and the Nasdaq Composite Index. The stock prices, respectively R_i and R_m, is calculated with the return formula consistent with MacKinlay's (1997) event study methodology:

$$R_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}} \tag{19}$$

where P_{it} represents the closing price of a company's stock on a given day, and P_{it-1} represents the closing price on the day before.

5.3 Twitter Data and Data Alignment

We have retrieved more than 3.3 million tweets, ranging from Jan 1, 2014 to Dec 31, 2016, which mentions the firms by their *cash-tag*. Tweets are restricted to English written messages as our sentiment classification machine is optimized towards the English language. The tweets

[3] https://finance.yahoo.com/

gathered are rated by a sentiment score; positive, neutral or negative, and aggregated to a measurement of daily frequency.

In Figure 5-1, we illustrate how earnings announcements are aligned with Twitter messages and stock returns. We find that all companies announce their quarterly earnings report after the stock exchange closes for the day. We define the event day as the following day of the announcement, from when the stock exchange opens at 09.30 EST, to 16.00 EST.

The Twitter data are gathered with the exact timestamp from when it was sent. To align it with daily financial stock returns, we must:

1) Change the time zone for Twitter data from UTC (Coordinated Universal Time) to EST (Eastern Standard Time). This is to reflect the information correctly in accordance to the stock market opening hours.
2) Aggregate message frequencies hourly.
3) Aggregate daily frequencies from hourly frequencies, starting at 16.00 present day to 16.00 the following day. This is done to reflect earnings announcement after the stock market closes. For instance, texts from Tuesday at 17.00 will be reflected in Wednesday's daily frequency up to 16.00, and so on.
4) Since Twitter messages can be sent every day, we transfer texts from Saturday and Sunday to Monday. In the case of holidays and other days the market is closed, we transfer the texts to the first available opening day.

Figure 5-1

Twitter and financial data alignment for the event window

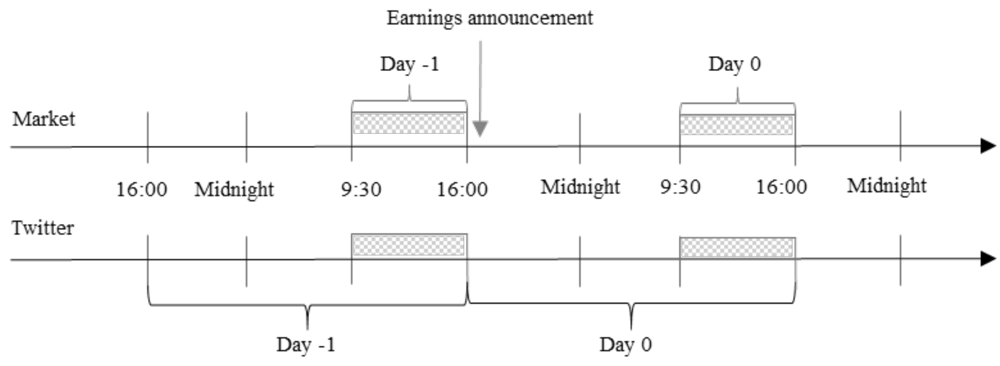


Figure 5-1 illustrates the alignment between financial and Twitter data around the event window. The *Market* time line represents conduction of the event study based on financial data. *Day -1* denotes the firm's after close earnings announcement date. *Day 0* denotes when the market can react to the respective announcement. The *Twitter* time line explains how tweets are aggregated daily in relation with opening and closing hours on the stock exchange, and how the time periods for *Day -1* and *Day 0* are characterized. The outlined boxes represent opening hours on Nasdaq Stock Exchange.

6 Empirical results

This section presents empirical results connecting Twitter and stock prices around earnings announcements for ten technology firms listed on Nasdaq stock exchange. Around events an increase in total number of tweets will be present whether it is surprising news. For instance, tweets will inform that Apple has just announced their earnings report which most investors are already aware of. To distinguish the attention that capture the investor's interests, we include the conditions abnormal tweet volume and sentiment, and exclude events not meeting an abnormality condition at 1% significance level. When presenting our results below, we focus on negative events. All neutral and positive events have been tested and found efficient, which the figures below illustrate. The same picture does not appear for negative events however.

We divide our results into two parts:

1) The comparing of aggregated abnormal returns for all companies. These findings are presented in three figures (i.e. figure 6-1 to 6-3) and table 6-1. Figure 6-1 include all negative events, figure 6-2 include only negative events meeting the criteria of significant abnormal tweet volume, and figure 6-3 include only negative events meeting the significant abnormal sentiment criteria.

2) In this part, we separate the companies in two categories based on our definition of famous and less famous companies as described in chapter 5.1. We present their results in the same manner as for the events in 1).

6.1 All companies

Figure 6-1 shows all aggregated positive, neutral and negative CARs for all events in our sample. From the graph, it looks like there are significant abnormal returns on the event day, but none in the consecutive days. As we mentioned in the beginning of this chapter this is correct for all positive and neutral events, but testing negative abnormal returns on day 1 for all events, we observe abnormal return significantly different from zero.

Figure 6-2 exhibits all aggregated results for events meeting the criteria of significant abnormal Twitter volume. We observe both positive and negative abnormal returns during the event window, but from table 6-1, we can assume that there are significant negative abnormal returns on day 1. The rest of the days are not significantly different from zero.

In figure 6-3 only events that meet the criteria of significant Twitter sentiment frequency are included. The positive and neutral abnormal returns are consistent with EMH as for our whole sample, which we also observe in the negative events. There are no consecutive abnormal returns observed after the event day, like we observed in the first two graphs.

These findings are supportive of our hypothesis that Twitter volume is inconsistent with EMH. Regarding the inclusion of events with abnormal sentiment frequency, we do not find significance to support our hypothesis. However, the sentiment model has the highest absolute abnormal return values, which can indicate that sentiment expressed on Twitter clarifies the impact of earnings announcements, which leads to an efficient market.

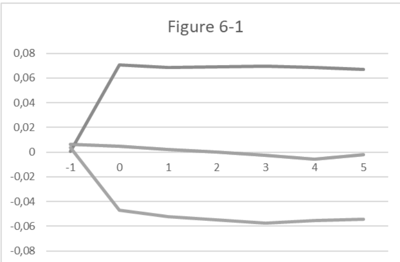


Figure 6-1 shows CAR for all the ten companies' events.

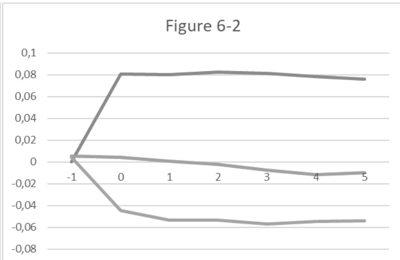


Figure 6-2 shows CAR for all the ten companies' events conditional on abnormal Twitter volume.

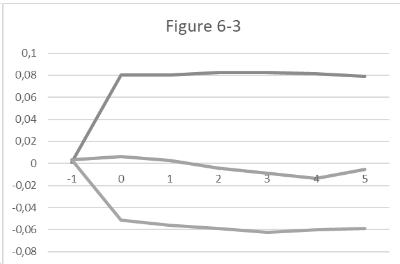


Figure 6-3 shows CAR for all the ten companies' events conditional on abnormal Twitter sentiment.

Table 6-1

Aggregated average abnormal returns and cumulative abnormal returns on negative earnings announcements for the 10 companies.

	Days						
	-1	0	1	2	3	4	5
AR	0,0036	-0,0506	-0,0055	-0,0024	-0,0025	0,0022	0,0008
	(1,67)	-(23,25)	-(2,55)	-(1,11)	-(1,16)	(1,00)	(0,35)
CAR	0,00362	-0,04695	-0,04695	-0,05490	-0,05742	-0,05524	-0,05447
AR Volume	0,0040	-0,0489	-0,0089	0,0003	-0,0037	0,0024	0,0004
	(1,66)	-(20,05)	-(3,66)	(0,13)	-(1,50)	(0,98)	(0,15)
CAR Volume	0,0040	-0,0448	-0,0538	-0,0534	-0,0571	-0,0547	-0,0543
AR sentiment	0,0031	-0,0548	-0,0046	-0,0027	-0,0037	0,0024	0,0015
	(1,30)	-(22,61)	-(1,90)	-(1,14)	-(1,54)	(1,01)	(0,60)
CAR Sentiment	0,0031	-0,0516	-0,0562	-0,0590	-0,0627	-0,0602	-0,0588

Table 6-1 presents the aggregated average abnormal returns and cumulative abnormal returns on negative earnings announcements for all ten firms. The first section shows the abnormal returns (AR) and cumulative abnormal returns (CAR) through the event window, given all negative events in the data sample. The next section includes abnormal and cumulative abnormal returns through the event window, conditional on abnormal tweet volume at 1% significance level. The last section includes abnormal and cumulative abnormal returns through the event window, conditional on abnormal sentiment frequency at 1% significance level. The test statistics shown in parentheses, have been calculated by dividing the aggregated AR by the average standard deviation from each event's estimation window.

6.2 The famous companies

In the figures represented below we have separated the companies as expressed in the start of this chapter. By looking at figure 6-4 to 6-6, we can immediately see that negative CAR results are different from the results seen at all firm events aggregated together.

In figure 6-4, showing all the four famous companies' events, we notice a steady downwards return slope for negative events, with a slightly upwards movement in the end. When testing the significance of the returns, shown in the first section of Table 6-2, we find all returns to be in consistency with the EMH, namely no significant abnormal returns different from zero in the days after the event.

In figure 6-5, where only events with significant abnormal Twitter volume is included, we register a crooked negative CAR graph, which shows an even steeper downwards movement than the line in 6-4. When the returns are tested in table 6-2, second section, we observe significant abnormal returns on the event day, day 1 and day 3. This contrasts with EMH, and support our hypothesis that Twitter attention through abnormal Twitter volume influence abnormal returns days after the event.

Figure 6-6, where only significant abnormal sentiment frequency is accounted for, we observe the same results for the negative events as when we looked at all the companies aggregated results. The graph in figure 6-6 has an extra downward bend at the end, but nothing significant, and again we see that Twitter sentiment is a factor not influencing the abnormal returns enough to create inconsistency against EMH.

The results from the famous companies give us ambiguous answers toward our hypotheses. As seen in figure 6-5, Twitter volume may have a significant effect on abnormal returns. As for our sentiment hypothesis, we are starting to realize this does not show the results first anticipated regarding abnormal returns days after an event.

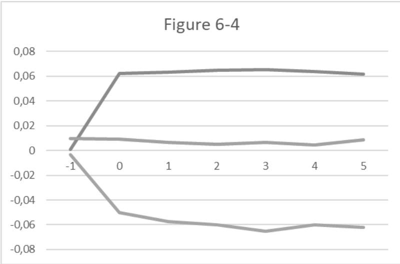


Figure 6-4 shows CAR for the famous companies' events.

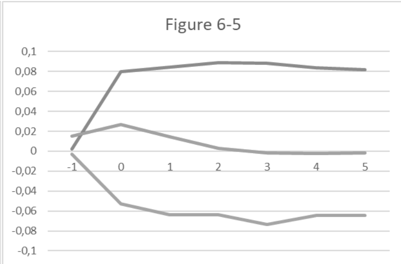


Figure 6-5 shows CAR for the famous companies' events conditional on abnormal Twitter volume.

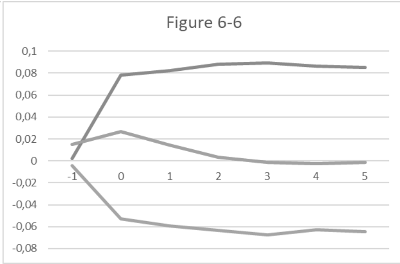


Figure 6-6 shows CAR for the famous companies' events conditional on abnormal Twitter sentiment.

Table 6-2

Aggregated average abnormal returns and cumulative abnormal returns on negative earnings announcements for the famous companies.

	Days						
	-1	0	1	2	3	4	5
AR	-0,0033	-0,0469	-0,0070	-0,0030	-0,0050	0,0050	-0,0018
	-(1,06)	-(15,12)	-(2,24)	-(0,96)	-(1,63)	(1,61)	-(0,57)
CAR	-0,00329	-0,05024	-0,05720	-0,06019	-0,06524	-0,06024	-0,06202
AR Volume	-0,0032	-0,0496	-0,0113	0,0000	-0,0093	0,0089	-0,0003
	-(0,84)	-(13,16)	-(2,99)	-(0,01)	-(2,46)	(2,37)	-(0,08)
CAR Volume	-0,0032	-0,0527	-0,0640	-0,0641	-0,0733	-0,0644	-0,0647
AR Sentiment	-0,0045	-0,0487	-0,0059	-0,0045	-0,0042	0,0051	-0,0016
	-(1,40)	-(15,19)	-(1,85)	-(1,40)	-(1,32)	(1,58)	-(0,50)
CAR Sentiment	-0,0045	-0,0531	-0,0591	-0,0635	-0,0678	-0,0627	-0,0643

Table 6-2 presents the aggregated average abnormal returns and cumulative abnormal returns on negative earnings announcements for the four famous firms. The first section shows the abnormal returns (AR) and cumulative abnormal returns (CAR) through the event window, given all negative events in the data sample. The next section includes abnormal and cumulative abnormal returns through the event window, conditional on abnormal tweet volume at 1% significance level. The last section includes abnormal and cumulative abnormal returns through the event window, conditional on abnormal sentiment frequency at 1% significance level. The test statistics shown in parentheses, have been calculated by dividing the aggregated AR by the average standard deviation from each event's estimation window.

6.3 The less famous companies

Our last results describe abnormal returns for the six less famous companies. All three figures 6-7 through 6-9 is seemingly the most efficient of all aggregated results in our thesis. When testing the significance of abnormal returns, we find no results deviate from EMH. As we can read from table 6-3, all statistically significant abnormal returns can only be observed on the event day.

Abnormal Twitter volume events have arguably the strongest absolute value on abnormal returns, but the results are still not significantly different from zero. Our assumptions that these firms are not attention-grabbing enough from attention through Twitter is strengthened as we reject our hypotheses in all scenarios. This might imply that investors' attention behavior may be found only in connection to large, well-known firms.

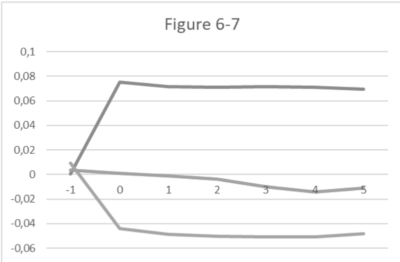


Figure 6-7 shows the CAR for the less famous companies' events.

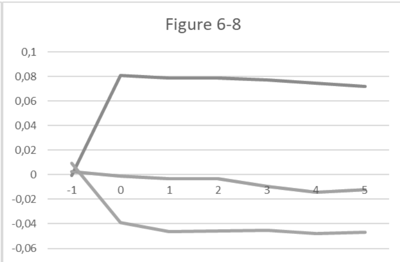


Figure 6-8 shows CAR for the less famous companies' events conditional on abnormal Twitter volume.

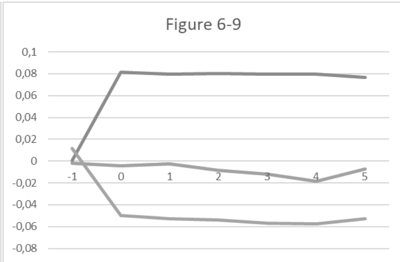


Figure 6-9 shows CAR for the less famous companies' events conditional on abnormal Twitter sentiment.

Table 6-3

Aggregated average abnormal returns and cumulative abnormal returns on negative earnings announcements for the less famous companies.

	Days						
	-1	0	1	2	3	4	5
AR	0,0095	-0,0536	-0,0043	-0,0019	-0,0004	-0,0002	0,0029
	(3,52)	-(19,95)	-(1,61)	-(0,71)	-(0,14)	-(0,08)	(1,09)
CAR	0,00947	-0,04417	-0,04850	-0,05042	-0,05080	-0,05101	-0,04809
AR Volume	0,0093	-0,0484	-0,0072	0,0006	0,0004	-0,0023	0,0008
	(3,05)	-(15,94)	-(2,38)	(0,19)	(0,14)	-(0,77)	(0,28)
CAR Volume	0,0093	-0,0391	-0,0463	-0,0457	-0,0453	-0,0477	-0,0468
AR Sentiment	0,0116	-0,0615	-0,0031	-0,0008	-0,0032	-0,0004	0,0048
	(3,17)	-(16,76)	-(0,85)	-(0,23)	-(0,86)	-(0,12)	(1,32)
CAR Sentiment	0,0116	-0,0499	-0,0530	-0,0539	-0,0570	-0,0575	-0,0526

Table 6-3 presents the aggregated average abnormal returns and cumulative abnormal returns on negative earnings announcements for all ten firms. The first section shows the abnormal returns (AR) and cumulative abnormal returns (CAR) through the event window, given all negative events in the data sample. The next section includes abnormal and cumulative abnormal returns through the event window, conditional on abnormal tweet volume at 1% significance level. The last section includes abnormal and cumulative abnormal returns through the event window, conditional on abnormal sentiment frequency at 1% significance level. The test statistics shown in parentheses, have been calculated by dividing the aggregated AR by the average standard deviation from each event's estimation window.

6.4 Summary of results

The first key result is the anticipated finding of several days of abnormal returns for famous companies when including only negative events that experience significant abnormal Twitter volume on the event day. This result supports our first hypothesis that investors act irrational on attention from Twitter. Negative abnormal return followed by more negative abnormal returns is referred to as underreaction (De Bondt and Thaler 1985). Should we believe Shiller's (2003) theory that professional investors take advantage of optimistic irrational buyers, there should exist strategies developed around these irrational investors to get rid of overvalued attention stocks after an event.

Bhana's (1995/1996) statement that attention is more reflected in negative events, is consistent with our results. Reasons as to why positive events are fully reflected when new information hits the market, and not negative results, might be because it is easier to draw conclusion of a company's future prospects when they present positive earnings. Gabrovsek, et al. (2017) discuss that positive results are possible leaked at least one day ahead of the announcement resulting in less volatile returns, which can influence the statistical significance of abnormal returns. This might be one reason for not finding abnormal returns the following days, as opposed to negative events where firms tend to keep their results back secret the earnings announcement.

.

The second key finding in our results is the fact that none of the events with sentiment abnormality criteria experienced abnormal returns after the event day. Gabrovsek, et al. (2017) might bring light to our findings, as they discover sentiment reflects stock movements on the day of the earnings statement, in accordance to EMH. Reasons may be that subjective sentiments, as our algorithms are based upon, are easier for individual investors to grasp, and their attention towards sentiment loaded tweets might ease the information processing to decide which stocks they ought to buy. Even though we reject our hypothesis regarding Twitter sentiment and market inefficiency, sentiment might have opposite effect. This may indicate, contrary to abnormal Twitter volume, paying attention to sentiments can reduce irrational, bad trades for individual investors.

Contributing to Barber and Odean's (2008) research, another proxy to identify individual investors' trading behavior can be found in whether a firm's operations are known to the investors. From our findings, the distinction between famous and less famous companies, and how abnormal tweet volume affect investors' attention towards them, might be an indicator as to how individual investors pick stocks.

6.5 Further implications

In our study, we have analyzed ten technology companies for a period of three years. To be able to find more certain results around attention, conducting a study analyzing several years of data, and evolving the study to include more companies, and apply two separate event studies on famous and less famous firms, and see whether results strengthens our findings. We have focused on the technology industry, but it would be interesting to expand the industry view on this topic to see if the same results that we find, can be replicated for instance in the clothing industry, or the financial industry.

One of our shortcomings is that our sentiment analysis is limited to our *normal* computational power. Machine learning algorithms demand the computational expensive handling of very complex systems and a lot of data. We have nevertheless been able to make a classifier that gives a sentiment prediction based on subjective feeling which scores almost as good as a human would be able to do, according to Pang and Lee (2008, 11).

7 Conclusion

In our thesis, we study whether Twitter attention influences market efficiency, and test both abnormal Twitter volume and abnormal Twitter sentiment as possible causes. We are able to show relations between abnormal Twitter volume and negative abnormal returns in the market, following a negative event. For positive events, we do not find any results indicating a relation. The same can be stated for abnormal Twitter sentiment around events, even negative ones.

We argue that some companies are more attention-grabbing than others, such as bigger organizations. We believe most investors, both individual and institutional, are following news tightly and have knowledge of their products and services, while investors do not have the same relationship towards other companies. Further, tweets reach out to investors in short time, and by detecting an abnormal number of tweets, the attention towards companies might impact investors' rationality.

The outcome from the less famous companies, strengthen our assumptions that known companies' attention influences the market efficiency. Earnings announcement news for the smaller organizations are immediately reflected in the stock prices. Each segment, whether Twitter is involved, has no following days with abnormal returns. In comparison, results from the attention-grabbing firms show differently. News are reflected in stock prices immediately in events when Twitter attention is not included, and even though stock prices decline the consecutive days, statistically, we cannot find additional days with abnormal return. Twitter volume, however, impacts the market otherwise. These findings show a connection between attention and certain firms, reflected in tweets, and more frequent days with abnormal returns.

As a result, we believe an indicator as to how individual investors pick stocks, is to distinguish whether a firm's operations are known to the investors, and how abnormal tweet volume affect traders' attention towards them.

8 Bibliography

Aharony, Joseph, and Itzhak Swary. 1980. "Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis.". *The Journal of Finance* 35 (1): 1-12.

Barber, Brad M., and Terrance Odean. 2008. "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors". *Review of Financial Studies* 21, (2): 785-818.

Bhana, Narendra. 1995/1996. "The share market reaction to earnings announcements—a test of the efficiency of the Johannesburg Stock Exchange.". *Investment Analysts Journal* 24 (42): 45-57.

Bikhchandani, Sushil, and Sunil Sharma. 2000. "Herd behavior in financial markets". *IMF Economic Review* 47, (3): 279-310.

Bloomfield, Robert, Robert Libby, and Mark W. Nelson. 2000. "Underreactions, overreactions and moderated confidence." *Journal of Financial Markets* 3 (2): 113-137, accessed May 20, 2017, doi: https://doi.org/10.1016/S1386-4181(00)00003-3

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market". *Journal of Computational science* 2, (1): 1-8.

Bondt, Werner FM., and Richard Thaler. 1985. "Does the stock market overreact?". *The Journal of Finance* 40, (3): 793-805.

Chowdhury, Gobinda G. 2003. "Natural language processing.". *Annual review of information science and technology* 37 (1): 51-89, accessed May 03, 2017, doi: https://doi.org/10.1002/aris.1440370103

Cipriani, Marco, and Antonio Guarino. 2009. "Herd behavior in financial markets: an experiment with financial market professionals.". *Journal of the European Economic Association* 7(1): 206-233, accessed May 01, 2017, doi: https://doi.org/10.1162/JEEA.2009.7.1.206

Costache, Mihai, Marie Liénou, and Mihai Datcu. 2006. "On Bayesian Inference, Maximum Entropy and Support Vector Machines Methods". In *Aip Conference Proceedings*, edited by Ali Mohammad-Djafari, Vol 872, (1): 43-51, accessed May 14, 2017, doi: http://dx.doi.org/10.1063/1.2423259

Fama, Eugene F. 1970. "Efficient capital markets: A review of theory and empirical work.". *The Journal of Finance* 25, (2): 383-417.

Gabrovsek, Peter, Darko Aleksovski, Igor Mozetič, and Miha Grčar. 2017. "Twitter sentiment around the Earnings Announcement events". *PloS one* 12, (2): e0173151.

Go, Alec, Richa Bhayani, and Lei Huang. 2009. "Twitter sentiment classification using distant supervision". *CS224N Project Report, Stanford* 1, (12).

Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. 2003."A practical guide to support vector classification". *National Taiwan University*: 1-16, accessed April 23, 2017, www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf

Kadiyala, Padmaja, and P. Raghavendra Rau. 2004. "Investor reaction to corporate event announcements: underreaction or overreaction?". *The Journal of Business* 77 (2): 357-386, accessed April 24, 2017, doi: https://doi.org/10.1086/381273

Lonie, Alasdair A., Gunasekarage Abeyratna, David M. Power and C. Donald Sinclair. 1996. "The stock market reaction to dividend announcements: A UK study of complex market signals." *Journal of Economic Studies* 23 (1): 32-52, accessed May 20, 2017, doi: http://dx.doi.org/10.1108/01443589610106534

MacKinlay, A. Craig. 1997. "Event studies in economics and finance". *Journal of Economic Literature* 35, (1): 13-39.

Mao, Huina, Scott Counts, and Johan Bollen. 2011. "Predicting financial markets: Comparing survey, news, twitter and search engine data". *arXiv:1112.1051*

Mao, Yuexin, Wei Wei, Bing Wang, and Benyuan Liu. 2012. "Correlating S&P 500 stocks with Twitter data". In *Proceedings of the first ACM international workshop on Hot Topics on Interdisciplinary Social Networks Research*, ACM (2012): 69-72, accessed May 18, 2017, doi: 10.1145/2392622.2392634.

Mlonzi, V. F., J. Kruger, and Meiya G. Nthoesane. 2011. "Share price reaction to earnings announcement on the JSE-ALtX: A test for market efficiency." *Southern African Business Review* 15 (3): 142-166.

Murphy, Kevin P. 2012. “*Machine learning: a probabilistic perspective”*. London: MIT press.

Nigam, Kamal., John Lafferty and Andrew McCallum. 1999. “Using maximum entropy for text classification”. In *IJCAI-99 Workshop on Machine Learning for Information Filtering* 1: 61–67, accessed April 24, 2017: http://www.kamalnigam.com/papers/maxent-ijcaiws99.pdf

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumbs up?: Sentiment classification using machine learning techniques.". In *Proceedings of the ACL-02* *conference on Empirical methods in natural language processing* 10: 79- 86. Association for Computational Linguistics: Accessed April 18, 2017, doi: 10.3115/1118693.1118704

Pang, Bo, and Lillian Lee. 2008. "Opinion mining and sentiment analysis". *Foundations and Trends® in Information Retrieval* 2, (1-2): 1-135, accessed April 15, 2017, doi: http://dx.doi.org/10.1561/1500000011

Pynnönen, Seppo. 2005. "On regression based event study.". *Acta Wasaensia* 143: 327-354.

Ranco, Gabriele, Darko Aleksovski, Guido Caldarelli, Miha Grčar and Igor Mozetič. 2015. "The effects of Twitter sentiment on stock price returns". *PloS one* 10, (9): Accessed May 12, 2017, doi: https://doi.org/10.1371/journal.pone.0138441

Rittenberg, Libby, and Timothy Tregarthen. 2008. "*Principles of Microeconomics*". New York: Flat World Knowledge, Inc.

Shiller, Robert J. 2003. "From efficient markets theory to behavioral finance.". *The Journal of Economic Perspectives* 17 (1): 83-104.

Sprenger, Timm O., Andranik Tumasjan, Philipp G. Sandner, and Isabell M. Welpe. 2014a. "Tweets and trades: The information content of stock microblogs". *European Financial Management* 20, (5): 926-957, accessed May 07, 2017, doi: https://doi.org/10.1111/j.1468-036X.2013.12007.x

Sprenger, Timm O., Philipp G. Sandner, Andranik Tumasjan, and Isabell M. Welpe. 2014b. "News or noise? Using Twitter to identify and understand company-specific news flow". *Journal of Business Finance & Accounting* 41, (7-8): 791-830, accessed May 03, 2017, doi: https://doi.org/10.1111/jbfa.12086

Yin, Chunyong, and Jinwen Xi. 2016. "Maximum entropy model for mobile text classification in cloud computing using improved information gain algorithm.". *Multimedia Tools and Applications* (2016): 1-17, accessed May 13, 2016, doi: https://doi.org/10.1007/s11042-016-3545-5

Zheludev, Ilya, Robert Smith, and Tomaso Aste. 2014. "When can social media lead financial markets?". *Scientific Reports* 4: 4213, accessed April 27, 2017, doi: https://doi.org/10.1038/srep04213