

Eötvös Loránd University
Faculty of Social Sciences
Applied Economics BA, III.

Stock market prediction using Twitter sentiment

Author: Kamenár Gyöngyvér
Supervisor: Stump Árpád

Table of Contents

I.	Introduction	4
II.	Literature review	6
II.1	Theoretical background	6
II.2	Twitter sentiment and stock market.....	8
III.	Data and sentiment analysis	10
III.1	Stock prices.....	10
III.2	Tweets	10
III.3	Sentiment analysis	11
IV.	Empirical analysis	14
IV.1	Correlation	15
IV.2	Unit root test	16
IV.3	Vector Autoregression	17
IV.4	Granger-causality.....	19
V.	Event study.....	21
V.1	Methodology.....	22
V.2	Event detection	24
V.2.1	Methodology	24
V.2.2	Amazon	26
V.2.3	Apple	27
V.2.4	Facebook	28
V.2.5	Microsoft	30
V.2.6	Tesla	31
V.3	Results	32
VI.	Conclusion.....	33
VII.	References	35

List of Tables

Table 1: Top 5 components of Nasdaq 100 on 6 th of April, 2021	10
Table 2: Number of tweets collected with the given query	11
Table 3: Correlation of return with sentiment score and tweet volume	15
Table 4: Results of the Dickey-Fuller test.....	17
Table 5:Results of Granger-causality	20
Table 6: Detected events for Amazon (good above 0.2 and bad below 0.1).....	27
Table 7: Detected events for Apple (good above 0.25 and bad below 0.1)	28
Table 8: Detected events for Facebook (good above 0.2 and bad below 0.1).....	29
Table 9:Detected events for Microsoft (good above 0.3 and bad below 0.1)	30
Table 10:Detected events for Tesla (good above 0.1 and bad below 0)	31
Table 11: Event matrix.....	32

List of Figures

Figure 1: Distribution of sentiment scores	13
Figure 2: Correlation between stock return and sentiment score	16
Figure 3: Event study timeline	22
Figure 4: Distribution of daily sentiment values.....	26
Figure 5:Daily volume of tweets containing \$AMZN in 2019	27
Figure 6: Daily volume of tweets containing \$AAPL in 2019.....	28
Figure 7: Daily volume of tweets containing \$FB in 2019	29
Figure 8: Daily volume of tweets containing \$MSFT in 2019.....	30
Figure 9: Daily volume of tweets containing \$TSLA in 2019	31
Figure 10: Average cumulative abnormal return for event types.....	33

I. Introduction

Can investors' sentiments predict stock returns? Is there any causal relationship between public sentiment and stock prices? Behavioral economics tells that emotions can directly affect individual decision-making, therefore investors' decisions too. A few decades ago it was hardly possible to observe thousands of people's moods and opinions around the globe daily but with the rise of social media in the 21st-century people happily post about their current status anytime or even several times a day.

Twitter is one of the most popular social media platforms worldwide, it has more than 350 million monthly active users as of October 2020 (Clement J., 2020) and approximately 500 million tweets are sent out each day (Ying Lin, 2020). Opposite to other platforms which are mainly used by youngsters, 63% of Twitter users are between the ages 35-65. The site is certainly the most popular in the US with more than 47 million accounts and 20% of US internet users access Twitter every month. Before anyone would think that social media is just used for entertaining purposes, one must see that even the world's most influential people consider it as one of the most important ways of informal communication, like former US presidents Donald J. Trump and Barack Obama, the CEO of Tesla Elon Musk, the founder of Microsoft Bill Gates, news sites like CNN, BBC, Reuters, The Economist and even an independent agency like NASA shares daily updates on its profile. As we can see a huge amount of information flows on Twitter and that can potentially influence investor decisions.

From an economical perspective, there is a widely known phenomenon, the efficient market hypothesis (EMH) developed by Fama (1970) which tells us that stock markets work efficiently and share prices immediately reflect all available information. According to EMH, stocks are always traded at their fair value and it is impossible to buy over- or undervalued stocks. Therefore, it is impossible to outperform the market either by technical or fundamental analysis. Investors can obtain higher returns only by riskier investments (Downey Lucas & Scott Gordon, 2020).

There are several criticisms and quantitative studies against EMH as well. Events like the market crash of October 1987, the Internet bubble of the late 1990s, or the Palm-3Com spinoff (Malkiel, 2003) provide some evidence against EMH. Theoretically, some arguments against EMH are information asymmetry between investors and transaction costs. However, from the point of this research, we should focus on market psychology and human irrationality, which provides the basis of behavioral finance.

Behavioral finance argues that humans are not always rational when they make a financial decision. We will argue that sentiments and the influence of others can affect investment decisions. Herd behavior and investing fads are often the results of such phenomena. We believe that Twitter is a great platform for observing others' sentiments, behavior, and expectations, thus we might be able to help predictions based on Twitter content.

The role of social media should not be underestimated. Just recently a social media extremely influenced the stock price of a company called GameStop. GameStop was an outdated, dying company with no prospects, thus hedge funds heavily shorted its stock (Grant, 2021). However, members of a forum called *wallstreetbets* on Reddit social media site figured out if they collectively drove up the stock price, hedge funds will have to buy back their short sellings to cut their losses which will drive up the price even further. This stock price skyrocketed from the closing of \$145.6 on 26th to \$469.42 on the 28th of January, 2021 (Misra and Grewal, 2021). This example perfectly illustrates the role of social media on financial markets and inspires us to study it deeper.

This study analyses the relationship between Twitter sentiments and daily stock returns in the case of 5 listed companies on Nasdaq, namely Amazon, Apple, Facebook, Microsoft, and Tesla. We are interested in whether Twitter sentiments can help predict stock return thus support investing decisions. Twitter sentiments of the tweets containing the stock symbol are measured with SentiStrength (Thelwall, n.d.) and a sentiment value is calculated for each day. We are interested in the following 4 hypotheses:

Hypothesis I: Daily stock return and Twitter sentiments are positively correlated.

Hypothesis II: Stock returns can help predict Twitter sentiments

Hypothesis III: Twitter sentiments can help predict stock returns

Hypothesis IV: Good Twitter events generate positive abnormal return while bad events
negative

We calculate the correlation coefficient for every company to test *Hypothesis I* and use the Granger-causality test for *Hypothesis II* and *Hypothesis III*. We use tweet volume and sentiment to detect and categorize influential events and execute the methodology described by MacKinlay (1997) to test *Hypothesis IV*.

We found positive correlation in all 5 cases but the highest values were in the case of Amazon (0.267) and Tesla (0.219). Regarding Granger-causality, we found, that Tesla sentiments

Granger-cause its returns and that Microsoft returns Granger-cause its sentiment value at 10 percent significance level. However, we did not find any significant relationship at the 5 percent level. In the event study, we detected and categorized 10 good, 14 neutral, and 18 bad events. We calculated the average cumulated abnormal returns for each event over the 41 days long event window and the results show positive abnormal return for the good event while negative for bad events.

The study is organized as follows: Section 2 describes the relevant literature while Section 3 introduces the data and sentiment analysis. Section 4 contains the empirical analysis and section 5 the event study. Finally, section 6 concludes the research.

II. Literature review

II.1 Theoretical background

The role of social media in financial markets might not be straightforward at first glance. However, there are several psychological phenomena connected to human irrationality which are catalyzed with the rise of online communication platforms and thus becoming more significant among investors. The basis of these originates from the interdisciplinary field of behavioral finance.

Behavioral finance argues that we can better understand some financial phenomena using models in which some participants are not fully rational (Thaler & Barberis, 2002). Investment decisions are also human decisions thus might not be fully rational all the time. Convincing evidence (Nofsinger, 2005) (Shu, 2010) suggests that investors' sentiment influence the choices they make on the market. A model behind investor sentiment was built by Barberis et al. (1998) and an extensive review by Daniel et al. (2002) presents the psychological biases and evidence of how these affect investor behavior. There are several identified phenomena supported by cognitive psychological evidence in behavioral finance, which suggest that our decisions are sometimes irrational. However, for this study, we should see those biases that arise with the use of Twitter. In our view, Twitter has two main impacts: it affects mood and allows the perception of other's expectations or opinions.

An interesting study on the effect of mood was done by Hirshleifer and Shumway (2003). They analyzed the relationship between sunshine and daily index returns in several cities using OLS, linear probability model and logit models. Based on psychological experiments sunshine has a positive effect on the mood of people and they investigated whether it influences investor

decisions. They found significant connections and they were able to improve the market portfolio with a weather-based trading strategy. However, after controlling for transaction cost the beneficial effect disappears. This inspired researchers who studied other influencers of mood, like sport events (Edmans et al., 2007) (Gerlach, 2011). All in all, the literature suggests that when people are happier and optimistic, they invest more and stock prices rise, while in the opposite case prices fall. Therefore, we believe capturing the collective mood makes it possible to better predict stock market movements (Zhang et al., 2011).

The other impact of Twitter is that it enables us to observe others' beliefs. What it has to do with the stock market? One of the first ideas comes from the famous metaphor of Beauty Contest in *The General Theory of Interest, Employment and Money* by Keynes (1936) which suggest that participants have to form beliefs about share prices based on other participants' beliefs of prices who also form beliefs based on others beliefs and so on (Allen et al., 2006), thus participants price assets based on higher-order reasoning and not based on fundamental value. This describes social dynamics in connection with the stock market which is investigated by R. J. Shiller et al. (1984) and Prechter and Parker (2007).

R. J. Shiller et al. (1984) states that stock price movements are affected by social dynamics from the observations of others in the market and human nature. He says that smart money leads the market but ordinary investors may overreact to news, cause swings and generate volatility. Based on social behavioral dynamics, Prechter and Parker (2007) rejects EMH and states that neoclassical economics does not offer any useful financial model. They explain that ambiguity about valuations by other participants gives rise to unconscious irrational herding that follows fluctuations in social mood. These social mood fluctuations in turn impel financial fluctuations.

Herding originates from information asymmetry and it is an important phenomenon to understand social dynamics. Brunnermeier (2001) argues that herding can be rational action if it offers payoff or informational externalities. In the case of the stock market and Twitter, both can be present so there is a motive for herding. The herd behavior is studied on financial markets, see Bikhchandani and Sharma's (2001) overview of empirical and theoretical research about this phenomenon on financial markets.

The roots of herding and fads can be originated from social psychology, namely from the concept of conformity (Bikhchandani et al., 1998). Aronson (2008) defines conformity as the change in one's opinion or behavior as a consequence of imagined or real pressure from a person or a group of people. The influences that people have based on the behavior or beliefs

of others called social influence. A special type of social influence is the informational social influence when we conform by the influence of others because we believe that they have better information. It is more common in ambiguous cases like valuations because agents think others are more correct.

In the case of investment decisions, especially in the short-term, this phenomenon can play an important role in stock price movements. It leads to herding behavior that may cause bubbles, trends, or investing fads as studied by (Bikhchandani et al., 1998). We believe that Twitter sentiments can signal these influences and can help to predict price movements on financial markets.

II.2 Twitter sentiment and stock market

With the rising popularity of social media in the past decade, researchers also started using it as a tool for measuring public sentiment. As detailed before, Twitter seems a great platform for signaling mood. However, the methodology to complete a sentiment analysis and build predictive models is often still different between scientists. We have a wide variety of tools in the intersection of computer science, data science, and econometrics. To somewhat understand the approaches, we must review the state-of-art stock market predictions using Twitter sentiment.

One of the most cited article on the field of sentiment analysis and stock market is written by Bollen et al. (2011) who studied whether the sentiment of daily Twitter feed can predict the value of the Dow Jones Industrial Average (DJIA) index in 2008. They analyzed daily large-scale Twitter feed with two mood tracking tools: Google-Profile of Mood States (GPOMS) that measures mood in 6 dimensions (calm, vital, alert, sure, kind, and happy), and OpinionFinder, which measures positive and negative mood. They use Granger-causality analysis and Self Organizing Fuzzy Neural Network (SOFNN) to test whether the created public mood time series can predict the changes of DJIA closing values. They found that stock market prediction models can be significantly improved by certain public mood dimensions, namely calm and happy measured by GPOMS, but not by others and neither by general positive mood measured by OpinionFinder. This study inspired Mittal and Goel (2012) who used nearly the same methodology on the same timeline but adding their own sentiment classification method and additional portfolio management consideration.

Mao et al. (2011) used six different data sources of investor sentiment ((1) investor intelligence survey data, (2) bullish percentage of daily sentiment index survey data, (3) tweet volumes of

financial search terms, (4) Twitter investor sentiment, (5) negative new sentiment and (6) Google search volume of financial queries) and compared their predicting power of 4 market indices: DJIA, Chicago Board Options Exchange Volatility Index (VIX), trading volume of DJIA, and gold prices from January 2008 to September 2011. They found that the two indicators of Twitter outperformed survey data sentiment and media news as well. This result confirms for our study that Twitter is a great platform for analyzing investor sentiment.

Gilbert and Karahalios (2010) analyzed over 20 million posts in 2008 on the site called *LiveJournal* (a social media and blogging service) and derived an anxiety index that measured the worry, fear, and anxiety of the site. They found significant Granger-causality and that the rise of the anxiety index predicts downward movement on S&P 500 index. It is important to note, that the analyzed posts are not connected to any financial terms or topics just generally discussing daily life, however the anxiety still impacted the market index based on this research.

The study of X. Zhang et al. (2011) measured the emotions of randomized sample of Twitter feed (one hundredth of all tweets per day) for 6 months in 2009. They found that expressions of fear, hope, and worry are negatively correlated with DJIA, Nasdaq, and S&P 500 while positively correlated with VIX.

Rao and Srivastava (2012) investigated the relationship between Twitter sentiments and financial measures like volatility, stock prices, and trading volume of DJIA, Nasdaq, and other 13 technological big-cap stocks. They found an extremely high correlation (up to 0.88 in the case of returns) and significant Granger-causality of short-term Twitter sentiments to the movement of indices.

Similar to our study, Sprenger et al. (2014) analyzed tweets with dollar-tagged symbol keywords. However, they used the Naïve Bayesian text classification method of S&P 100 companies to derive bullishness or bearishness from over 6 months in 2010. They found an association between message volume and trading volume, disagreement of tweets and volatility as well as tweet sentiment and stock returns. They also noticed that users with better than average investment advice are more likely to be retweeted and followed by more people.

The effects of Twitter sentiment on stock price returns by Ranco et al. (2015) used the same keyword methodology for tweet selection as Sprenger et al. (2014), Granholm and Gustafsson (2017), and our study. They analyzed 30 stock companies from the DJIA index over 15 months. The researcher investigated correlation, Granger-causality between stock return and the created Twitter sentiment time series from tweets containing the symbol of the given stock as a

keyword. They also carried out event studies for significant good, bad, and neutral events detected by tweet volume peaks. The results showed a relatively low correlation and Granger-causality but significant positive abnormal returns for good while negative for bad events.

A study that did not find any relationship between aggregate Twitter mood and stock market executed by Nofer and Hinz (2015) in Germany from 2011 to 2013. However, after taking into account the followers, they found significant predictive power of public mood. From their training dataset, they created a trading strategy that reached 36% return over 6 months period after correcting for transaction costs.

III. Data and sentiment analysis

III.1 Stock prices

For the analysis, we selected the top 5 weighted stocks of Nasdaq, namely Apple, Microsoft, Amazon, Tesla and Facebook. Their historical prices and earnings announcements are freely available on Investing.com (www.investing.com). It is important to note that these companies have already well-built brands and marketing, they are observed by millions all over the world and their performances are actively discussed in the news and social media. Nasdaq Composite index is also retrieved for use as a market model to calculate abnormal return during events.

Table 1: Top 5 components of Nasdaq 100 on 6th of April, 2021

#	Company	Symbol	Weight	Price (\$)
1	Apple Inc.	AAPL	10.892	126.34
2	Microsoft Corp.	MSFT	9.681	248.35
3	Amazon.com Inc.	AMZN	8.374	3228
4	Tesla Inc.	TSLA	4.23	692.3
5	Facebook Inc.	FB	3.829	306.35

Source: Own Editing

The daily stock closing price (P_t) data is used for each day to calculate the time series of daily returns R_t :

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

III.2 Tweets

We retrieved the tweets from the 1st of January 2019 to the 31st of December 2019 from Twitter (www.twitter.com) with a data mining method called *web scraping*. Tweets are referred to as

the posted content on the Twitter website. Similarly to other social media sites, tweets can be liked and shared and can contain links and keywords. Keywords in Twitter are indicated with a special character # called a hashtag in front of them to be easily found by anyone. To tag investment-related words, traders have adopted another form of this keyword system, which is a dollar sign \$ followed by the ticker symbol e.g. \$TSLA. Another feature of Twitter is that the content cannot be too large, its character limit currently is 280 but only 1 percent hit that limit and only 5 percent is longer than 190 characters (Perez, 2018). The average tweet length is 33 characters and the short content makes them easy to analyze from the perspective of sentiments.

On the Twitter feed, we are just interested in those, which are related to the mentioned stocks because these are the most likely to reflect investor sentiments. Therefore, we are using a query for every stock containing their symbol as a keyword: \$aapl, \$msft, \$amzn, \$tsla and \$fb. The text content, username, date, and the number of likes and retweets are obtained for each tweet that is listed by the mentioned queries. We also use a filter for the English language because we can analyze only English content. From the retrieved tweets, we remove duplicates and tweets without postdate because we cannot use them for daily sentiments. Some posts do not have text content because these are just links to articles, pictures, or videos so we also delete them from the analysis. Finally, the number of analyzed tweets for each stock is summarized in Table 2:

Table 2: Number of tweets collected with the given query

Query	Number of tweets
\$AMZN	98 440
\$AAPL	144 523
\$FB	83 697
\$MSFT	43 711
\$TSLA	175 885

Source: Own Editing

III.3 Sentiment analysis

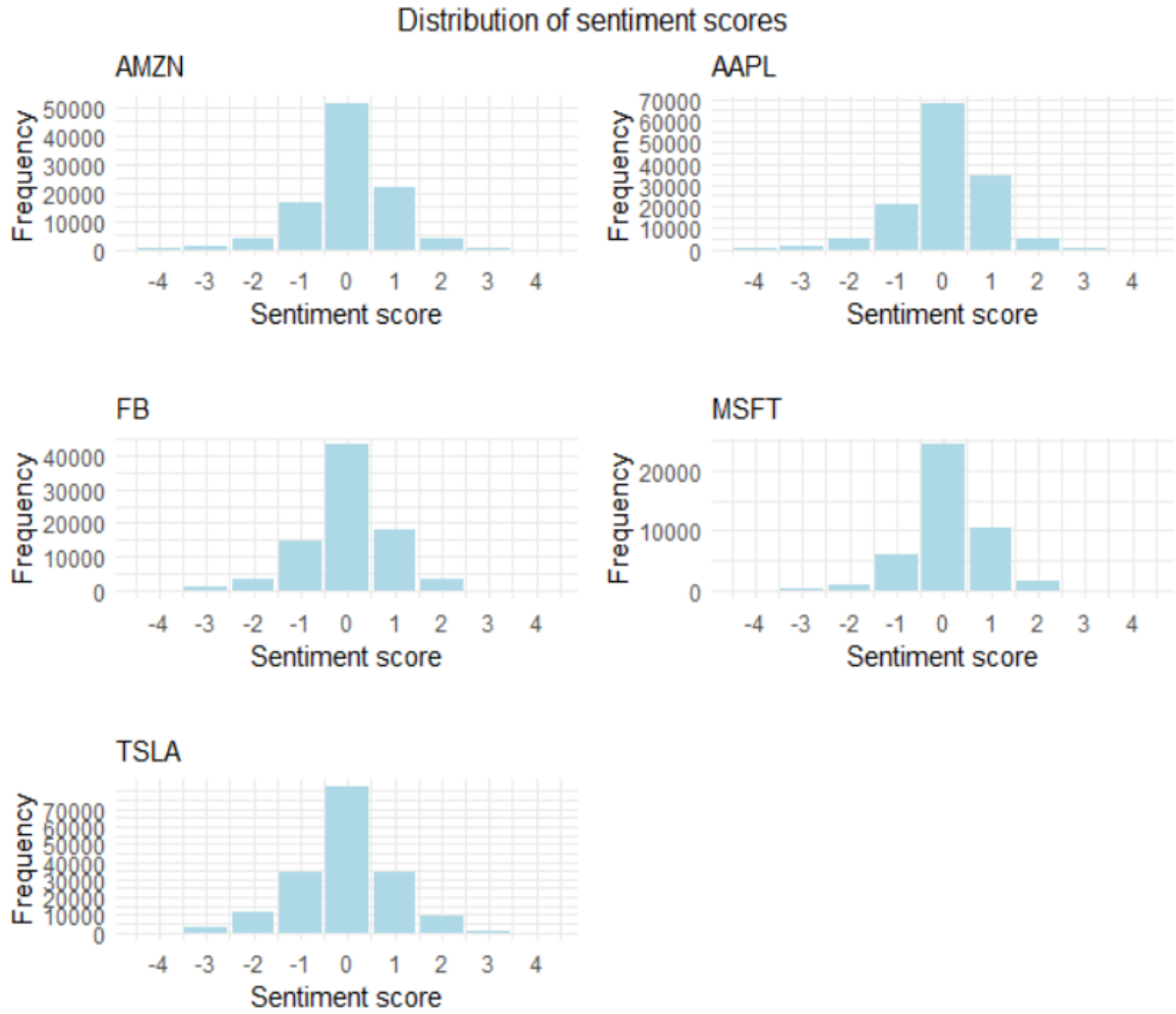
We used the SentiStrength software (Thelwall, n.d.) that is specially developed for social web sentiment analysis in short informal English text and used in academic circles e.g. Zheludev et al. (2014), Kucuktunc et al. (2012), Pfitzner et al. (2012) and Thelwall et al. (2011). SentiStrength has several advantages compared to other language processing tools concerning social media. It can handle emoticons which are often used as an expression of mood therefore it should be taken into account. Besides, SentiStrength can handle typos and multiple characters which are often intentionally included in an online post like writing "coooooo" instead of "cool".

The software also recognizes repeated punctuation and boosts the strength of the preceding sentiment by 1. Common idioms are also identified by SentiStrength and override the individual words' strengths. The software also recognizes negation and enhancement (Thelwall et al., 2012).

Evaluating sentiment analyzing tools is not a straightforward task because their accuracy mainly depends on the datatype and true human-level sentiments can be defined by humans. According to Thelwall et al. (2012), the program is not always better than machine learning approaches that utilize indirect indicators of sentiments. Islam and Zibran (2018) further developed SentiStrength with a software engineering (SE) domain dictionary and compared it with two other popular sentiment analyzing tools, namely the Natural Language Toolkit (NLTK) and Stanford NLP. They tested them on 5600 JIRA issue comments which were manually annotated too. The researchers found that the developed SentiStrength-SE substantially outperformed the other two domain-independent tools in identifying sentiment in software engineering texts.

By default, the program gives a positive and a negative sentiment value for each input text based on the fact (Berrios et al. 2015) that humans experience controversial mixed emotions. However, we used the scale option which is the sum of positive and negative sentiment values to get an individual sentiment score for each tweet. The possible range of this sentiment score is from -4 (most negative) to +4 (most positive). Figure 1 shows the distribution of tweets based on sentiment scores for each searched stock symbol.

Figure 1: Distribution of sentiment scores



Source: Own Editing

Based on the sentiment score we categorize them as positive when they have a positive sentiment score, negative when their sentiment score is negative, and neutral when the sentiment score is 0. Thereafter, we calculate the daily sentiment value (S) for each day (t) just like Ranco et al. (2015) did: the difference of quantity of positive tweets (tw_t^+) and quantity of negative tweets (tw_t^-) over the sum of them:

$$S_t = \frac{tw_t^+ - tw_t^-}{tw_t^+ + tw_t^-}$$

In Figure 1 we can see that most of the tweets' sentiment score is 0 for each stock because they do not express any positive or negative sentiment based on SentiStrength, thus categorized as neutral. However, they are not included in the formula because we suppose that these do not

determine the sentiment polarity of the given day. Losing many neutral tweets is the drawback of this methodology because we ignore a huge amount of content.

A possible beneficial idea could be to take into account the number of followers, retweets, or likes and then somehow re-categorize them based on the significance of their content. Sprenger et al. (2014) concluded that users with better than average investment advice have more followers and retweeted more often. Based on German data Nofer and Hinz (2015) also emphasized that it is necessary to take the number of followers into account to measure the contagion of mood. However, in our study, we do not apply any further methodology for sentiment analysis for two main reasons. Firstly, acquiring the followers and retweets requires a more complex method of web scraping thus huge computational power, which is not feasible for this study. Secondly, working out a methodology on how to get more information would make this study much more complicated without any guarantee that it can provide more accurate results. One of our main points is to keep the study feasible both methodologically and computationally in order to have the possibility that the procedure can be automated and provide an investment advisor system.

IV. Empirical analysis

In the previous section, we described the data source and methodology on how we got the two time series for each analyzed company, the stock return R_t and the daily sentiment value S_t for each trading day in 2019. In this section, we execute time series analysis to investigate the relation between daily stock return R_t and the Twitter sentiment value S_t . We calculate the correlation to get a general view about the relationship between stock return and Twitter sentiment. Then we perform unit root tests to determine whether the time series are stationary because only stationary time series can be included in the vector autoregression (VAR) model. Next, we build a vector autoregression model to select the optimal number of lags. Based on the VAR model, we test Granger-causality to decide whether one time series significantly Granger-causes the other and vice versa.

Based on the theoretical and intuitional background we expect high positive correlation between the daily stock return R_t and Twitter sentiment value S_t of the given companies. We do not have any expectations for the unit root test because in the case of non-stationary time series we can differentiate and test again. Regarding causality, we expect that stock return Granger-causes Twitter sentiment. Besides, based on the theoretical background we expect that Twitter sentiment also has some causal effect on stock returns. In other words, we expect them

to be endogenous with no clear Granger-causality direction. It is plausible that both returns and sentiments are influenced by news, so their error terms are not independent.

IV.1 Correlation

We calculated the correlation coefficient between daily return and sentiment polarity as well as between daily return and the volume of tweets. The Pearson correlation is calculated as the covariance over the product of their standard deviation (Wooldridge, 2016 p. 658.) described with the following formula:

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - (E(X))^2 * E(Y^2) - (E(Y))^2}}$$

The coefficient shows the linear relation between the variables and its value ranges from -1 to +1. As stated in Hypothesis I. we expect positive correlation between stock return and Twitter sentiment. We also check the correlation between return and the volume of tweets out of curiosity but we do not have any expectation for this on theoretical background. The results are summarized in Table 3:

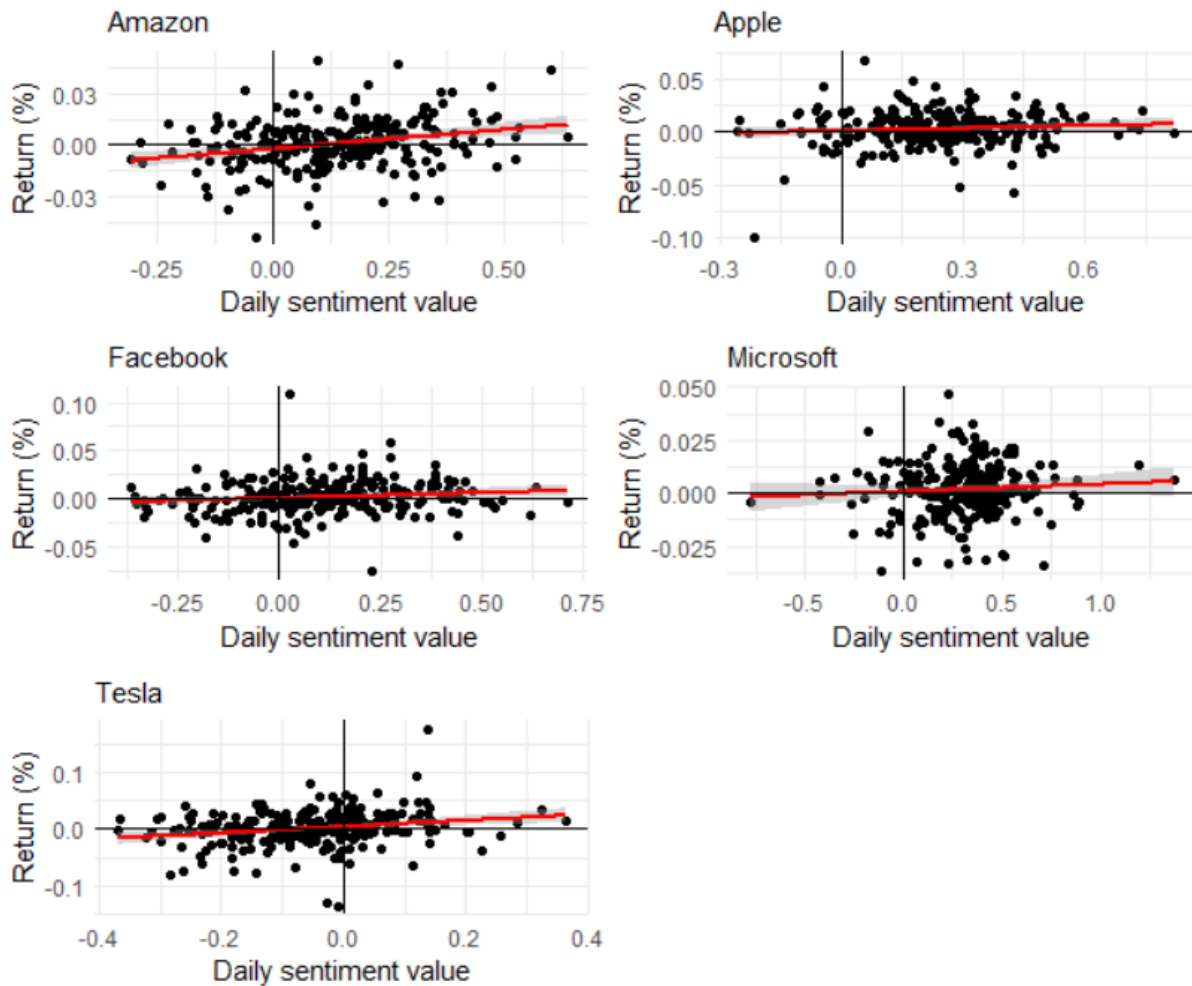
Table 3: Correlation of return with sentiment score and tweet volume

Stock	$\rho(R, \text{Sentiment score})$	$\rho(R, \text{Volume})$
Apple	0.088	-0.123
Amazon	0.267	0.085
Facebook	0.126	0.207
Microsoft	0.068	0.184
Tesla	0.219	-0.143

Source: Own Editing

As expected, there are positive correlation coefficients between stock return and sentiment score in all cases. We illustrated the scatter plot with the linear trend in Figure 2. The results are similar to Ranco's (2015) but in two cases (Microsoft and Apple) much lower than expected. There are already stronger correlations in the studied literature like in the article of Rao and Srivastava (2012). The correlation is the highest in the case of Amazon (0.267) and Tesla (0.219). However, the correlation between return and tweet volume is diverse. In the case of Apple and Tesla, the correlation is negative, indicating that when the volume is higher the return is lower. In the case of Facebook and Microsoft, the correlation between tweet volume and return is positive and surprisingly high. It suggests that when there are more tweets there are higher returns too.

Figure 2: Correlation between stock return and sentiment score



Source: Own editing with R

IV.2 Unit root test

Before building a vector autoregressive model and examining Granger-causality we have to make sure that we are using stationary time series. A time series process whose probability distributions (PD) are stable over time called stationary (Wooldridge, 2016 p. 345.). Symbolically, if we select random variables from the sequence and shift them k periods ahead, the joint PD must stay the same. Thus a stationer time series have stable variance and mean over time. A famous stationary time series in the white noise. The stationarity of the time series can be tested through the unit root with the Augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1981). The Augmented Dickey-Fuller test is the extended version of the simple Dickey-Fuller test because it allows the test of higher-order autoregressive processes i.e. more time lags of the series can be included. The general form also includes the possibility of α constant and β trend as we can see in the formula:

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \sum_{i=1}^k \theta_i \Delta y_{t-i} + \varepsilon_t$$

The null hypothesis of the test is $\delta = 0$ which means that the series has a unit root and called non-stationary. A well-known non-stationer series is the random walk process. We want to get a P-value lower than 0.05 to reject this hypothesis, thus concluding that $\delta < 0$ and the time series is stationary.

We tested the five stocks' daily return in percent and their daily sentiment score with the possibility of a constant (α). In Table 4 below we included the test statistics of the model with one lag and its P-value for every time series we want to use, then stated the conclusion based on the results. As we can see, all of the series are stationary so we can include them in a VAR model.

Table 4: Results of the Dickey-Fuller test

Time series	Dickey-Fuller statistic	test P-value	Conclusion
Amazon Return	-12.74	<0.01	Stationary
Amazon Sentiment	-6.34	<0.01	Stationary
Apple Return	-13.69	<0.01	Stationary
Apple Sentiment	-7.07	<0.01	Stationary
Facebook Return	-11.21	<0.01	Stationary
Facebook Sentiment	-6.49	<0.01	Stationary
Microsoft Return	-14.20	<0.01	Stationary
Microsoft Sentiment	-7.77	<0.01	Stationary
Tesla Return	-11.01	<0.01	Stationary
Tesla Sentiment	-5.95	<0.01	Stationary

Source: Own Editing

IV.3 Vector Autoregression

The autoregression estimates how a time series depends on its own past. The vector autoregressive model is a further developed version of autoregression because we can model several time series with their pasts, (Wooldridge, 2016 p. 590.) thus it is also useful for forecasting. The general equations for two variables y_t and z_t look like this:

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + \alpha_2 y_{t-2} + \gamma_2 z_{t-2} + \dots + u_{1t}$$

$$z_t = \mu_0 + \beta_1 y_{t-1} + \rho_1 z_{t-1} + \beta_2 y_{t-2} + \rho_2 z_{t-2} + \dots + u_{2t}$$

In our case, we estimate a vector autoregressive model with stock returns and Twitter sentiment. It means that the stock return is explained by the previous day(s) return(s) and the previous day(s) Twitter sentiment. Alternatively, Twitter sentiment value is explained by the sentiment value of the past day(s) and by the past day(s) stock return(s).

To build the optimal VAR model we have to select the optimal lag length or also called order. The lag length means, how many of the previous observations are included in the model. Considering that we have daily observations without any seasonality and our topic is the stock market, the intuition suggests that the number of lags should not be more than 1. It is because we assume that the market rapidly reacts to news or events and price movements are also swiftly reflected in public sentiment.

Nevertheless, we select the optimal order based on usual model selection criteria, namely F statistic and Bayesian Information Criterion (BIC) or also called Schwarz Criterion (Schwarz, 1978). While F statistic measures the joint significance of variables, BIC aims to maximize the likelihood function ($\mathcal{L}$) and also adds a penalty depending on the number of parameters (k) and observations (n) to avoid overfitting. BIC is defined with the following formula:

$$BIC = k \ln(n) - 2\ln(\mathcal{L})$$

Based on this selection process we found that in four out of five cases, Apple, Facebook, Microsoft, and Tesla, the optimal order is one. So to say, a given day's return is explained by just the previous one day's return and Twitter sentiment, and a given day's Twitter sentiment is explained just the previous one day's Twitter sentiment and return. Therefore, the VAR models are built the following way:

$$Return_t = \delta_0 + \alpha_1 Return_{t-1} + \gamma_1 Sentiment_{t-1} + u_{1t}$$

$$Sentiment_t = \mu_0 + \beta_1 Sentiment_{t-1} + \rho_1 Return_{t-1} + u_{2t}$$

In the case of Amazon return and stock, the selection process suggests that the optimal number of lags is two, thus the previous two day's observations ($t - 1$) and ($t - 2$) are also added:

$$Return_t = \delta_0 + \alpha_1 Return_{t-1} + \gamma_1 Sentiment_{t-1} + \alpha_2 Return_{t-2} + \gamma_2 Sentiment_{t-2} + u_{1t}$$

$$Sentiment_t = \mu_0 + \beta_1 Sentiment_{t-1} + \rho_1 Return_{t-1} + \beta_2 Sentiment_{t-2} + \rho_2 Return_{t-2} + u_{2t}$$

IV.4 Granger-causality

The method of Granger-causality examines the main thesis of this research stated in *Hypothesis II* and *Hypothesis III*. Namely, whether Twitter sentiments can help predict stock returns and vice versa. Granger causality (Granger, 1969) examines whether there is a causal relationship between two series by testing whether the lagged values of variable x can help predict y . The test is based on the comparison of two models. The first is when y is only explained by the lags of y and the second one is when y is explained by the lags of y and x as well. The null hypothesis is that the lagged value of x does not help to predict y .

For testing the causal effect of Twitter sentiment on stock return, we compare the model where the return is explained by just its own past values to a model where the return is explained by its own past values and sentiments as well. The question is, whether the model in which sentiments are added is better than the model in which just the return's past values are included. Inversely in the case when we test whether the stock return has a causal effect on Twitter sentiments. The model is built on the following formula:

$$y_t = \sum_{i=0}^k \alpha_i * y_{t-i} + \sum_{i=0}^k \beta_i * x_{t-i} + u_t$$

Then an F-test is used to measure whether the coefficients of lagged x are jointly significant. In determining the optimal order (lags) we rely on the previously selected orders of the VAR models. In the case of Amazon, the optimal order (lag length) is two, while in every other case it is one. The results of the tests are summarized in the following table:

Table 5: Results of Granger-causality

Null hypothesis	F statistic	P-value	Conclusion
Apple sentiment does not Granger-cause return	0.938	0.33	Accept
Apple return does not Granger-cause sentiment	1.005	0.32	Accept
Amazon sentiment does not Granger-cause return	0.175	0.84	Accept
Amazon return does not Granger-cause sentiment	1.387	0.24	Accept
Facebook sentiment does not Granger-cause return	0.291	0.59	Accept
Facebook return does not Granger-cause sentiment	0.556	0.46	Accept
Microsoft sentiment does not Granger-cause return	1.407	0.23	Accept
Microsoft return does not Granger-cause sentiment	3.006	0.08	Reject at 10%
Tesla sentiment does not Granger-cause return	3.234	0.07	Reject at 10%
Tesla return does not Granger-cause sentiment	2.248	0.13	Accept

Source: Own Editing

There are two forecasting relationships at 10 percent significance level but none of them are significant at 5 percent. Firstly, Microsoft return Granger-causes its Twitter sentiment score. Secondly, Tesla's sentiment score Granger-causes Tesla return. The reverse of this latter statement is also close to being significant at 10 percent level with a 0.13 P-value.

Contrary to our expectations, in most cases, we cannot say that stock returns can help predict Twitter sentiments and Twitter sentiments can help predict stock returns as stated in Hypothesis III and Hypothesis IV. The results are somewhat less promising than found by Ranco et al. (2015) with a similar methodology but support vector machine (SVM) based sentiment analysis technique. Out of 30 companies, they found 3 cases when Twitter sentiment Granger-causes stock return at 5% significance level. However, in 10 cases the stock return Granger-causes Twitter sentiment at 5% significance level based on their research.

We believe our results can be explained by daily aggregation and endogeneity. Our analysis examines daily Twitter content and returns, which shows some relationship in some cases but not really significant. If we could do this analysis on smaller intervals (e.g. hourly) there could be stronger relationship, like between past hour's sentiment affects this hour's return. Besides, there is endogeneity, the equations' error terms are not independent.

Overall, the forecasting relationship between stock return and the Twitter sentiment is the strongest in the case of Tesla. We know that Tesla is most popular amongst the analyzed 5 companies in terms of tweets (more than 175,000 tweets over one year). This may help to explain why it has the best forecasting relationship.

V. Event study

In the previous section, we examined whether there is a general relationship like the correlation between stock returns and Twitter sentiments connected to the given company. Furthermore, we tested the hypotheses that Twitter sentiments can help forecast stock returns and vice versa. With these methods, we analyzed the whole time series to draw general conclusions. However, we believe social media has a significant role during individual events too. By event we mean a day when there is unusually high media or social media activity about a company, signaling that lot of news burst out thus "*something interesting*" happened. We do not specify any further criteria on what is that "*something interesting*". We just rely on Twitter activity to show us the days when there was "*something interesting*". We will see that this approach is reliably detecting earnings announcements and influential company-related news.

In this section, we study the abnormal price returns during events. We detect the events based on tweet volume and categorize them as good, bad, or neutral based on the Twitter sentiment value of that day. Thereafter, we examine whether good events generate positive cumulated abnormal returns while bad events negative as stated in *Hypothesis IV*. This approach provides the possibility of an automated event detecting and categorizing process which can potentially support short-term investment decisions.

We use the event study framework described by MacKinlay (1997). This approach is targeted for event studies on the financial market which is exactly what we need to test *Hypothesis IV*. MacKinlay examined 600 earnings announcement events of 30 companies on DJIA and calculated abnormal return in case of good news, bad news, and no news. We want to test the same effect but on different events, companies, index, and timeline.

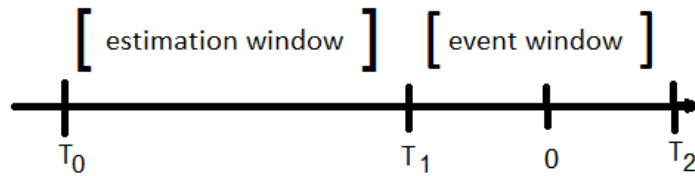
This methodology is used for analyzing events on the financial market in academic circles too. The event study of Ranco et al. (2015) is also based on this methodology. The specialty in our case is that we use tweet volume to detect important events and determine their category based on Twitter sentiment. We categorize 3 types of events: good, bad, and neutral. The hypothesis is that good events based on Twitter have a positive cumulative abnormal return during the event window, while bad events have a negative. For each company, we also select the highest

Twitter activity peak which is not connected to earnings announcements, and present what happened, how the public sentiment and market responded, and compare it with our analysis.

V.1 Methodology

The used methodology of an event study is based on two time intervals as illustrated in Figure 3. Firstly, the estimation window is used to estimate the required parameters ($\hat{\alpha}_i$ and $\hat{\beta}_i$) to calculate the expected returns $E[R_{i,t}]$ for each security (i) at time t . Secondly, the event window, where we investigate the impact of an event with abnormal and cumulative abnormal returns. We applied a 41-day-long event window and the day of the event is in the middle of the event window. Like (MacKinlay, 1997) did, we want to capture the abnormal returns 20 days prior to the event to 20 days after in order to clearly see its influence, and adding the event day itself we get the 41-day-long event window. It implies that the event window starts 20 days before the exact event and lasts for 20 days after.

Figure 3: Event study timeline



Source: Own Editing

The abnormal return $AR_{i,t}$ at time t is the actual return $R_{i,t}$ minus the normal return $E[R_{i,t}]$ of stock i .

$$AR_{i,t} = R_{i,t} - E[R_{i,t}]$$

The actual return is calculated from daily closing prices as described in section 3. The normal return is the expected return if the event did not happen.

We estimate the normal return $E[R_{i,t}]$ for each day with the market model approach. This statistical model relates the return $R_{i,t}$ of stock i to the market portfolio $R_{M,t}$ at time t . The linearity ensues from the joint normality of asset returns (MacKinlay, 1997). The parameters of the market model (α_i, β_i , and $\sigma_{\varepsilon_{i,t}}^2$ while $\varepsilon_{i,t}$ is the zero mean disturbance term) are estimated over the estimation window or in other words the past data. In our study, we used the Nasdaq Composite index as the market portfolio and ordinary least squares (OLS) procedure for estimating the market model parameters.

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + \varepsilon_{i,t}$$

$$E(\varepsilon_{i,t}) = 0 \text{ and } var(\varepsilon_{i,t}) = \sigma_{\varepsilon_{i,t}}^2$$

$$E[R_{i,t}] = \hat{\alpha}_i + \hat{\beta}_i R_{M,t}$$

We estimated the parameters of the market model ($\hat{\alpha}_i$ and $\hat{\beta}_i$) on the $[T_0, T_1]$ estimation window. The estimation window has to be right before the event window and not overlap in order to provide proper estimates which are not influenced by the event. The length of the estimation window should be large enough to assume that the additional variance due to sampling error in α_i and β_i is 0. We use 120 days as suggested by Campbell et al (1997). After calculating the expected return, we can calculate the abnormal return. The disturbance term ε of the market model on out of sample basis is the abnormal return $AR_{i,t}$.

$$AR_{i,t} = R_{i,t} - \hat{\alpha}_i - \hat{\beta}_i R_{M,t}$$

$$\sigma_{\varepsilon_{i,t}}^2 = \frac{1}{L-2} \sum_{t=T_0+1}^{T_1} (R_{i,t} - \hat{\alpha}_i - \hat{\beta}_i R_{M,t})^2$$

$$H_0: AR_{i,t} = N(0, \sigma^2(AR_{i,t}))$$

To conclude an event's impact, the abnormal returns must be aggregated through the event window. Besides, one observation is not very useful so we also aggregate across event observations. First, we aggregate through N events as the mean abnormal return of securities at time t in the event window. We define the cumulative abnormal return $CAR_i(t_1, t_2)$ over the event window as the sum of included abnormal returns from t_1 to t_2 . We determined the event window $(T_2 - T_1)$ to 41 days so the event window for an event at t is $[t-20, t+20]$. Cumulating the mean abnormal return of the observations on this interval we get the sample average cumulative abnormal return.

$$\overline{AR}_t = \frac{1}{N} \sum_{i=1}^N AR_{i,t}$$

$$var(\overline{AR}_t) = \frac{1}{N^2} \sum_{i=1}^N \sigma_{\varepsilon_i}^2$$

$$\overline{CAR}(t_1, t_2) = \sum_{t=t_1}^{t_2} \overline{AR}_{i,t}$$

$$var(\overline{CAR}(t_1, t_2)) = \sum_{t=t_1}^{t_2} var(\overline{AR}_t)$$

We can draw inferences about the cumulative abnormal return by testing the null hypothesis that the abnormal returns are zero. To test the hypothesis we have to know the variance of abnormal return but in practice $\sigma_{\varepsilon_i}^2$ is unknown. Therefore, we use the sample variance $\sigma_{\varepsilon_{i,t}}^2$ measured in the market model regression through the estimation window.

V.2 Event detection

Events in connection with listed companies usually mean huge public news like earnings announcements, acquisitions, product presentations, and so on. There is a huge tweet flood around these events, especially when we consider world-famous, big companies. Observing the number of posted tweets can help us detect events and daily Twitter sentiment values can determine what type of event it is. In this section, we describe the methodology used to recognize events and their polarity, based on Twitter.

V.2.1 Methodology

We detect events based on the time series of the daily volume of tweets (tw_t). When there is a peak compared to the previous and following day we suggest that there was an event. For detecting these peaks, we use a similar methodology to (Ranco et al., 2015) with an additional filter term to avoid overlapping events.

We use a moving window of $2 \times 5 + 1$ days centered at t_0 and let it slide along the time series. Within this moving window, we calculate the baseline activity tw_{base} as the median of the window. Then we specify the ratio $\varphi(t_0)$ for the central day t_0 as the relative difference with respect to the base activity:

$$\varphi(t_0) = \frac{tw_{t_0} - tw_{base}}{tw_{base}}$$

When this ratio exceeds the critical value we mark it as a peak. The critical value is not the same for every company, its range is from 0.5 to 1. It is necessary to determine the critical value one by one because of the different characteristics and distribution of the tweet volume time series. Otherwise, in some cases, we would detect too many and overlapping peaks while zero for other companies. Similarly to Ranco et al., (2015) the chosen critical value is determined on an experimental basis for every company. By default, we tried 1 as the critical value and in the case when it detected the appropriate number of events, we stayed with 1. We determined

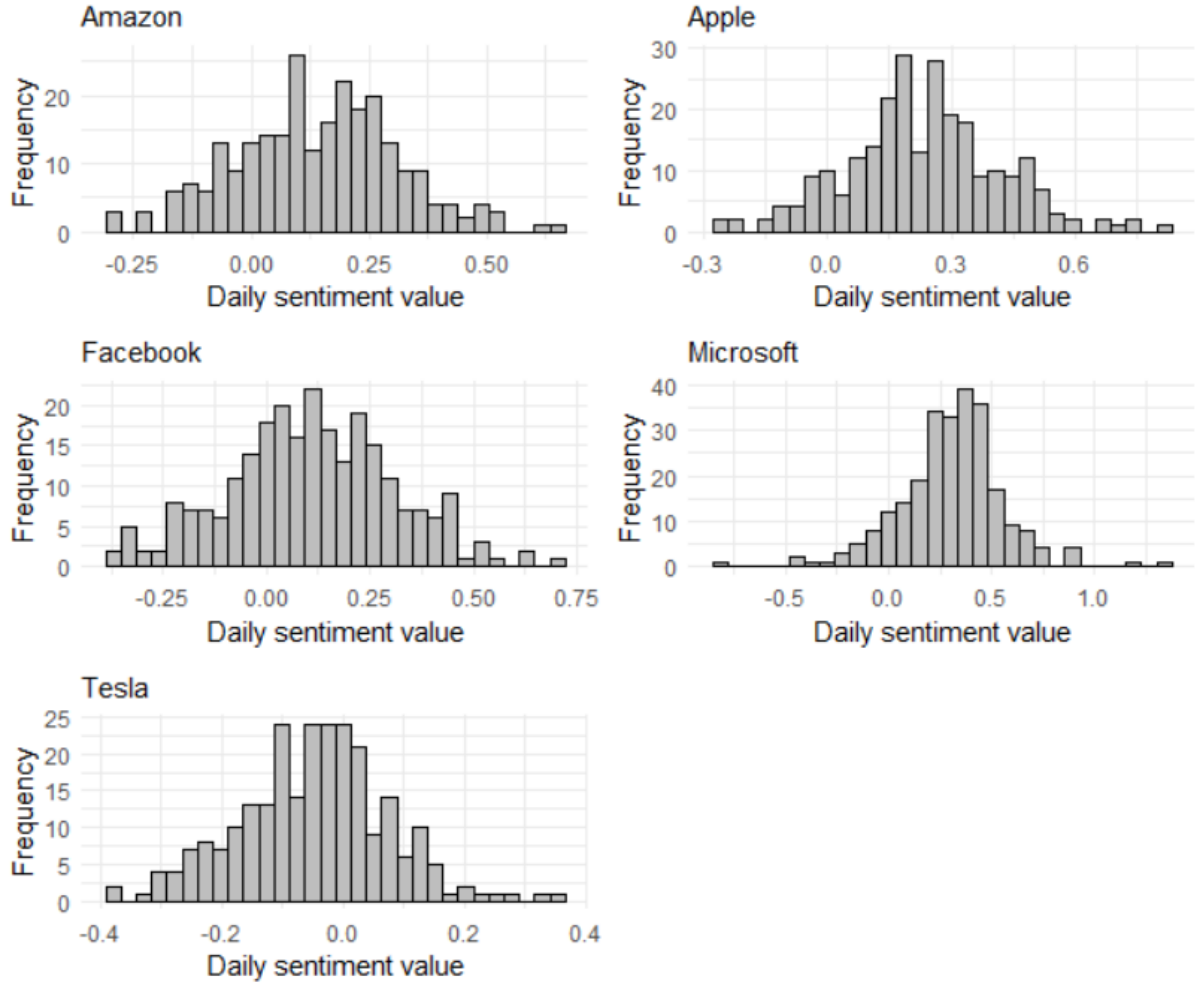
the range for the appropriate number of events between 5 and 15 considering that we have 252 trade days and the event windows are 41 days long. However, in some cases, the 1 critical value did not detect any or enough peaks so we started lowering it by 0.1 in each step. We reduced it until the number of events reached the appropriate range. The exact critical values for every company are detailed in the next subsections.

It sometimes happens that there are more peaks around an event. For example, the day before an event and the following day are also detected based on the critical value because these also have high tweet volume. We believe that these volume peaks are due to the same significant event and we want to detect the only one. It is also important that events should not overlap much, because we investigate abnormal returns from 20 days prior to the event to 20 days after. To avoid the overlap and co-occurrence of events we included an additional condition for marking a peak. This condition is that the given day (t_0) must have the highest tweet volume on $(-3; +3)$ days interval:

$$tw_{t_0} = \max([tw_{t_{-3}}, tw_{t_{+3}}])$$

Implementing this condition beside the appropriate critical value of $\varphi(t_0)$ we filtered out the tweet volume peaks due to the same event. Finally, the marked events' type (good, bad, neutral) is determined based on its daily sentiment value introduced in section 3. We elect a lower threshold below which the event is categorized as bad and a higher threshold above which the event is good. Between these threshold values, the event is categorized as neutral. Again, the threshold values are determined individually for every stock because their range and distribution are different as illustrated in Figure 4. In the next section, we list the events and determine their category.

Figure 4: Distribution of daily sentiment values



Source: Own editing with R

V.2.2 Amazon

The critical value of tweet volume ratio $\varphi(t_0)$ for Amazon is 0.5 which detected 9 events illustrated in Figure 5. The lower threshold value for categorizing bad events is 0.1 and the threshold for good events is 0.2. Based on these parameters we detected 4 bad events, 3 good ones, and 2 neutrals. Out of these 9 events, 4 is the date of quarterly earnings announcement (EA).

Amongst the events connected to Amazon and listed in Table 6, the highest tweet volume on the not EA dates was the 26th of December. We found that on this day Amazon announced its best holiday season in terms of sales (Goldman, 2019). The daily Twitter sentiment value based on our analysis is 0.6 which is well above the 0.2 threshold value, thus we categorized it as a good event. By intuition, we also believe it is good news. The stock price of Amazon rose that

day by more than 4 percent which is a record in 2019 and the cumulated abnormal return over the event window is also positive.

Figure 5: Daily volume of tweets containing \$AMZN in 2019

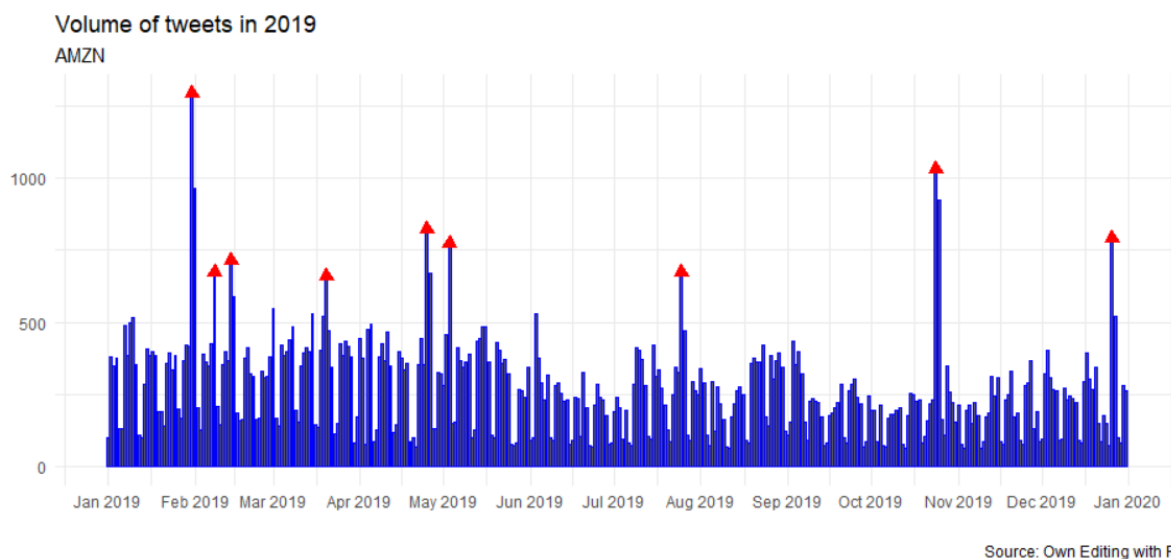


Table 6: Detected events for Amazon (good above 0.2 and bad below 0.1)

Date	EA	Twitter volume	Twitter sentiment value	Event type
2019-01-31	EA on 30th	1291	0.076	Bad
2019-02-08	-	672	-0.16	Bad
2019-02-14	-	715	0.28	Good
2019-03-20	-	660	0.14	Neutral
2019-04-25	EA	824	0.22	Good
2019-05-03	-	772	-0.05	Bad
2019-07-25	EA	673	0.11	Neutral
2019-10-24	EA	1029	-0.01	Bad
2019-12-26	-	791	0.60	Good

Source: Own Editing

V.2.3 Apple

The critical value for the tweet volume ratio $\varphi(t_0)$ is 1 in the case of Apple which recognized 7 events. Not surprisingly, 4 are at the date of the earnings announcement. The threshold sentiment value for good events is 0.25 while the threshold for bad events is 0.1. Therefore, we found 2 good, 2 bad, and 3 neutral events for Apple in 2019.

Surprisingly, there is an outstanding tweet volume value on the 3rd of January, 2019 which was not an EA day. As turned out, that day Apple lowered its revenue guidance for the first fiscal quarter by \$9 billion (Salinas, 2019). Obviously, it is bad news for the investors. Based on our Twitter sentiment analysis the daily sentiment value was -0.21. As mentioned, we categorize events as bad if their sentiment value is lower than 0.1, therefore it is also detected as a bad event. Naturally, the stock price also dramatically fell.

Figure 6: Daily volume of tweets containing \$AAPL in 2019

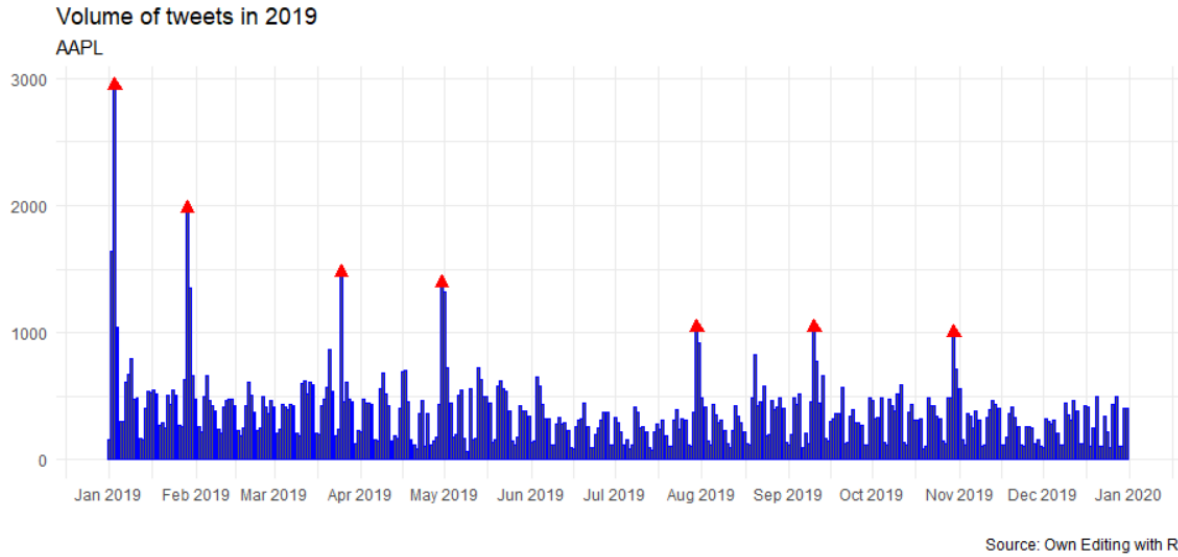


Table 7: Detected events for Apple (good above 0.25 and bad below 0.1)

Date	EA	Twitter volume	Twitter sentiment value	Event type
2019-01-03	-	2939	-0.21	Bad
2019-01-29	EA	1976	0.01	Bad
2019-03-25	-	1467	0.26	Good
2019-04-30	EA	1391	0.26	Good
2019-07-30	EA	1035	0.22	Neutral
2019-09-10	-	1044	0.12	Neutral
2019-10-30	EA	997	0.19	Neutral

Source: Own Editing

V.2.4 Facebook

The critical value of tweet volume ratio $\varphi(t_0)$ for Facebook is 0.6 which detected 8 events illustrated in figure 7. The lower threshold value for categorizing bad events is 0.1 and the threshold for good events is 0.2 as in the case of Amazon. Based on these parameters we

detected 6 bad events, 1 good and 1 neutral. Out of these 8 events, 4 is the date of the quarterly earnings announcement.

Figure 7: Daily volume of tweets containing \$FB in 2019

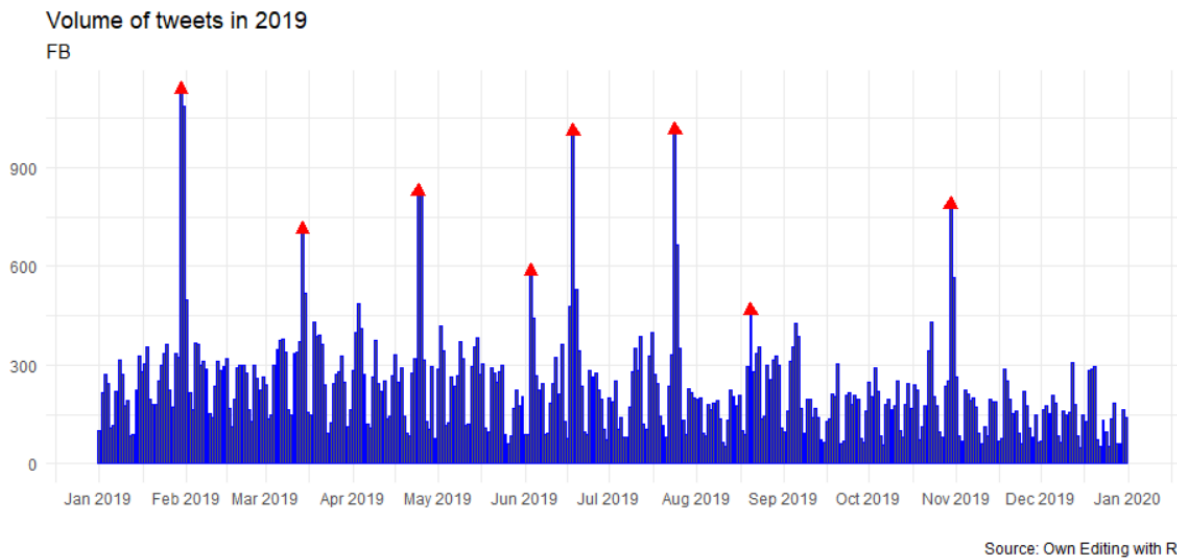


Table 8: Detected events for Facebook (good above 0.2 and bad below 0.1)

Date	EA	Twitter volume	Twitter sentiment value	Event type
2019-01-30	EA	1135	0.0	Bad
2019-03-14	-	713	-0.22	Bad
2019-04-24	EA	826	0.13	Neutral
2019-06-03	-	584	0.23	Good
2019-06-18	-	1009	0.06	Bad
2019-07-24	EA	1014	0.09	Bad
2019-08-20	-	465	-0.32	Bad
2019-10-30	EA	785	0.07	Bad

Source: Own Editing

After EA events, the highest tweet volume connected to Facebook was on the 18th of June, 2019 with more than 1000 tweets. That day Facebook announced that they plan a cryptocurrency-based global financial system (Isaac and Popper, 2019). We do not judge this announcement because we believe it cannot be not obviously categorized. A corporate financial investment decision like this has many benefits and drawbacks as well. However, we can detect the public mood on Twitter. The sentiment value on the 18th of June was 0.06. Based on the distribution of Figure 4 we take the events below 0.1 sentiment value as bad. The company's share price

fell that day and the cumulative abnormal return over the event window is also negative (-2.4%).

V.2.5 Microsoft

The critical value for the tweet volume ratio $\varphi(t_0)$ is 0.5 in the case of Microsoft which recognized 7 events. Not surprisingly, 4 are at the date of the earnings announcement. The threshold sentiment value for good events is 0.3 which is the highest among the 5 companies. The threshold for bad events is 0.1. Therefore, we found 3 good and 4 neutral events for Microsoft in 2019.

Figure 8: Daily volume of tweets containing \$MSFT in 2019

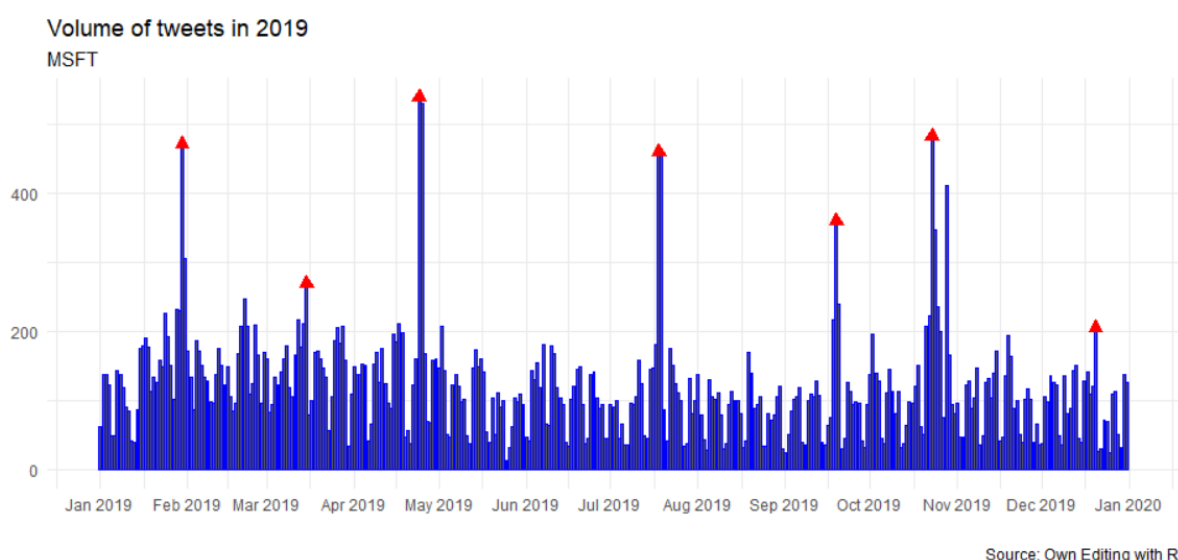


Table 9: Detected events for Microsoft (good above 0.3 and bad below 0.1)

Date	EA	Twitter volume	Twitter sentiment value	Event type
2019-01-30	EA	470	0.18	Neutral
2019-03-15	-	269	0.22	Neutral
2019-04-24	EA	538	0.43	Good
2019-07-18	EA	460	0.21	Neutral
2019-09-19	-	360	0.55	Good
2019-10-24	EA	482	0.26	Neutral
2019-12-20	-	205	0.44	Good

Source: Own Editing

Considering Microsoft events, the highest tweet volume amongst the not EAs was on the 19th of September. On the previous evening after the market close the company announced a quarterly dividend increase and a new share buyback program (Todd et al., 2019). Based on the

Twitter sentiment value (0.55) we categorized this event as good because it is above the 0.3 threshold value. Certainly, we also believe that this was a good event and the stock price also rose.

V.2.6 Tesla

The critical value of tweet volume ratio $\phi(t_0)$ for Tesla is 1 which detected 11 events illustrated in figure 9. The lower threshold value for categorizing bad events is 0 and the threshold for good events is 0.1. Based on these parameters we detected 6 bad events, 1 good, and 4 neutral.

Figure 9: Daily volume of tweets containing \$TSLA in 2019

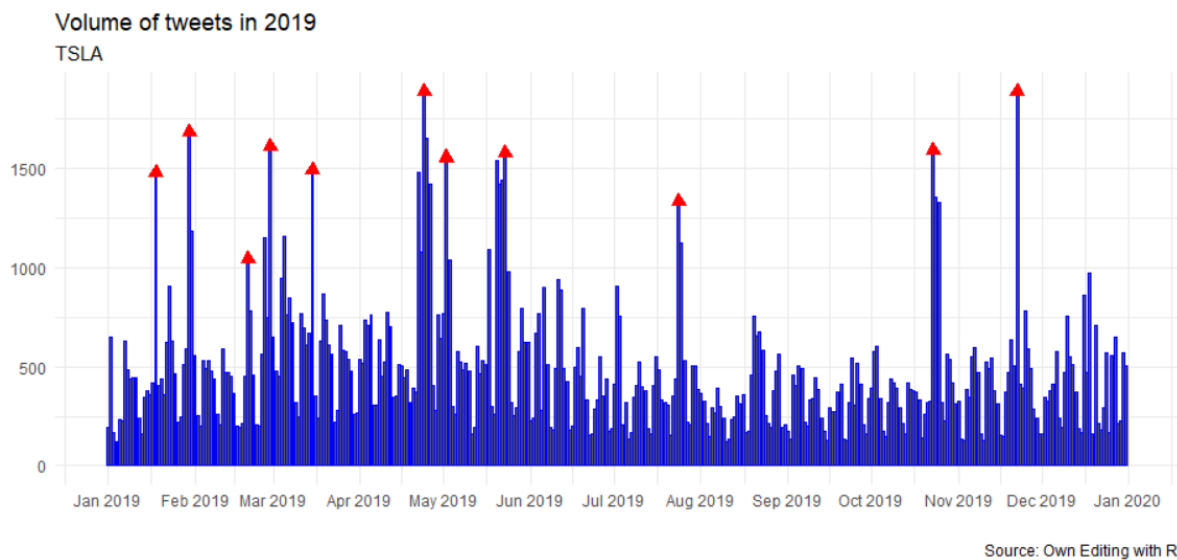


Table 10: Detected events for Tesla (good above 0.1 and bad below 0)

Date	EA	Twitter volume	Twitter sentiment value	Event type
2019-01-18	-	1482	-0.03	Bad
2019-01-30	EA on 29th	1681	0.03	Neutral
2019-02-20	-	1044	0.03	Neutral
2019-02-28	-	1608	0.08	Neutral
2019-03-15	-	1491	-0.01	Bad
2019-04-24	EA	1886	-0.10	Bad
2019-05-02	-	1555	-0.13	Bad
2019-05-23	-	1581	-0.15	Bad
2019-07-24	EA	1333	-0.02	Bad
2019-10-23	EA	1593	0.21	Good
2019-11-22	-	1886	0.01	Neutral

Source: Own Editing

The most tweet about Tesla was on the 22nd of November in 2019 which not earnings announcement day. That day Tesla presented its new vehicle called *cybertruck* and claimed that it is unbreakable (Ho, 2019). However, when the CEO Elon Musk threw a metal ball to the glass, to everyone's surprise, it broke. The judgment of this incident is ambiguous because many people think it was a media trick while others believe it was a huge failure. Based on the 0.01 Twitter sentiment score we evaluated it as neutral. The next day the stock price dropped 6 percent but the cumulated abnormal return over the event window is positive.

V.3 Results

As detailed before we detected 42 events during 2019 for the 5 companies based on tweet volume and categorized them based on their daily sentiment value. All in all, we signaled 10 events as good, 14 as neutral, and 18 as bad. The event types for each company are summarized in Table 11.

We determined the abnormal return and the cumulative abnormal return over the event window for every event as described in subsection 6.1. Then we calculated the average cumulative abnormal return (ACAR) for each event type. The results are illustrated in Figure 10.

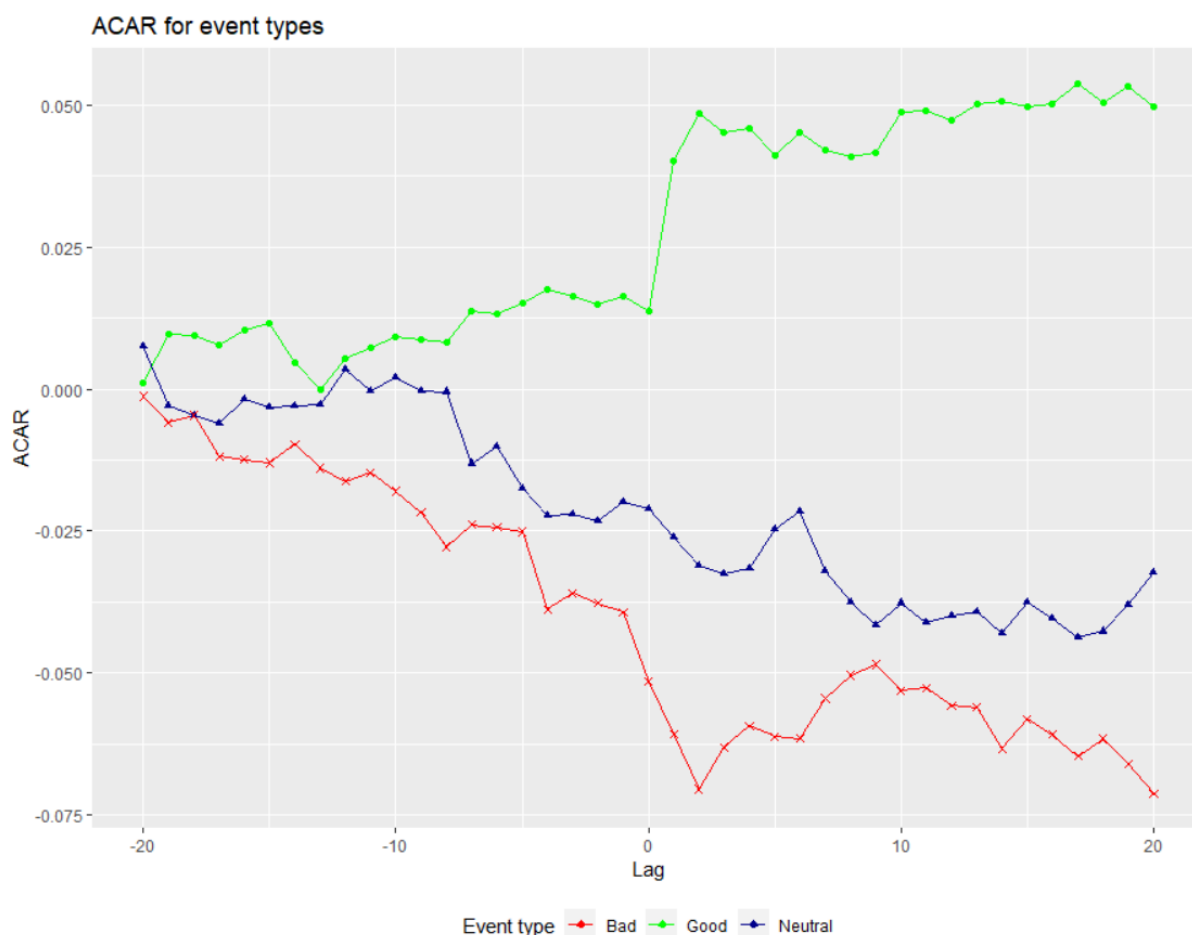
The conclusion is the same as expected in *Hypothesis IV*: on average good events generate positive CAR while bad events negative. We can see that the ACAR of neutral events is between the good and bad events' ACAR and slightly decreasing. At lag 1 point or in other words one day after the event we can see a huge increase in the good events' ACAR and a moderate decrease in bad events' ACAR. The neutral events' is also Interestingly, there are already differences in the even categories' ACAR well before the event occurred at the 0 lag point.

Table 11: Event matrix

	AMZN	AAPL	FB	MSFT	TSLA
Good	3	2	1	3	1
Neutral	2	3	1	4	4
Bad	4	2	6	0	6

Source: Own Editing

Figure 10: Average cumulative abnormal return for event types



Source: Own Editing with R

VI. Conclusion

The research question of this study is whether Twitter sentiments can help predict stock market movements and vice versa. Besides, we calculated the correlation between daily stock return and Twitter sentiment to see the linear relationship. Furthermore, we performed an event study to see whether Twitter volume and sentiments help detect events that influence the financial market and generate an abnormal return.

The theory behind the relationship between Twitter and the stock market is twofold. Firstly, we suppose that mood of investors can influence their financial decisions. Secondly, we assume that the observation of other's behavior can also influence agents' behavior, especially in such ambiguous cases like stock valuations. We consider Twitter as the greatest online platform for both signaling and observing mood, expectations, and opinions. Therefore, we believe that the appropriate analysis of Twitter content can help predict stock prices.

We analyzed the daily stock returns of the top 5 companies of the NASDAQ 100 index, namely Amazon, Apple, Facebook, Microsoft, and Tesla from the 1st of January to the 31st of December in 2019. To measure public mood, we retrieved all tweets containing the given stock's symbol with dollar-tag (\$AMZN, \$AAPL, \$FB, \$MSFT, and \$TSLA) in 2019. We used SentiStrength (Thelwall, n.d.) to measure the tweets' sentiment score and then categorized them as positive, negative, or neutral. For each day, we determined the daily sentiment value as the relative difference of the quantity of positive and negative tweets.

We found positive correlation between daily return and sentiment value for all five companies and the highest values were in the case of Amazon (0.267) and Tesla (0.219). We built VAR models and tested Granger-causality between stock return and daily sentiment. At 10 percent significance level we found, that Tesla sentiments Granger-cause its returns and that Microsoft returns Granger-cause its sentiment value. However, we did not find any significant relationship at 5 percent level. It means that the variables are endogenous, there is no clear Granger-causality direction between them.

In the event study, we detected 42 events (unusually high amount of tweets) for the observed five companies and categorized them based on the daily Twitter sentiment value. We categorized 10 events as good, 14 as neutral, and 18 as bad. We calculated abnormal return as the difference between actual return and expected return. The expected return is calculated based on the market model approach. We cumulated the abnormal returns for each event over the 41 days long event window and took the average of CAR for every event type. The results indicate positive abnormal return for the good events while negative for bad events. Interestingly, we can see the difference between good and bad events' ACAR several days before the actual event date.

In conclusion, we managed to support our hypotheses in some cases but not in others. We believe that this study provides a reliable methodology and feasible technical process in order to be automated, however, the results are not convincing enough yet to consider them as support in investment decisions. It does not mean that we reject the idea but it needs further improvements. A possible improvement could be made on sentiment analysis with finance-specific lexicon domain or advanced machine learning methods. Besides, the analysis could be carried out more frequently to provide several intraday values which might provide better forecasts.

VII. References

- Allen, F., Morris, S., & Shin, H. S. (2006). Beauty contests and iterated expectations in asset markets. *The Review of Financial Studies*, 19(3), 719–752.
- Aronson, E. (2008). *The Social Animal 10th ed.* Worth (10th Editi). Freeman.
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307–343.
- Berrios, R., Totterdell, P., & Kellett, S. (2015). Eliciting mixed emotions: a meta-analysis comparing models, types, and measures . In *Frontiers in Psychology* (Vol. 6, p. 428). <https://www.frontiersin.org/article/10.3389/fpsyg.2015.00428>
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3), 151–170.
- Bikhchandani, S., & Sharma, S. (2001). Herd Behavior in Financial Markets. *IMF Staff Papers*, 47, 1.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- Brunnermeier, M. K. (2001). *Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding*. Oxford University Press on Demand.
- Campbell, J. Y., Champbell, J. J., Campbell, J. W., Lo, A. W., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. princeton University press.
- Clement J. (2020, November 24). • *Most used social media 2020* | Statista. Statista. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Daniel, K., Hirshleifer, D., & Teoh, S. H. (2002). Investor psychology in capital markets: evidence and policy implications. *Journal of Monetary Economics*, 49(1), 139–209. [https://doi.org/https://doi.org/10.1016/S0304-3932\(01\)00091-5](https://doi.org/https://doi.org/10.1016/S0304-3932(01)00091-5)
- Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057–1072.
- Downey Lucas, & Scott Gordon. (2020, October 30). *Efficient Market Hypothesis (EMH)*

Definition.

Investopedia.

<https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp>

Edmans, A., Garcia, D., & Norli, Ø. (2007). Sports sentiment and stock returns. *The Journal of Finance*, 62(4), 1967–1998.

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

Gerlach, J. R. (2011). International sports and investor sentiment: do national team matches really affect stock market returns? *Applied Financial Economics*, 21(12), 863–880.

Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1).

Goldman, D. (2019, December 26). *Amazon has its best holiday season ever* . CNN. <https://edition.cnn.com/2019/12/26/business/amazon-holiday-sales/index.html>

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.

Granhölm, J., & Gustafsson, P. (2017). *The Quest for the Abnormal Return: A Study of Trading Strategies Based on Twitter Sentiment* [Umeå School of Business and Economics]. <https://www.diva-portal.org/smash/get/diva2:1115366/FULLTEXT01.pdf>

Grant, K. (2021, January 28). *GameStop: What is it and why is it trending?* . BBC . <https://www.bbc.com/news/newsbeat-55841719>

Hirshleifer, D., & Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance*, 58(3), 1009–1032. <https://doi.org/10.1111/1540-6261.00556>

Ho, V. (2019, November 23). *Elon Musk's net worth plunges \$768m in a day after cybertruck fiasco*. The Guardian. <https://www.theguardian.com/technology/2019/nov/22/elon-musk-net-worth-tesla-cybertruck>

Isaac, M., & Popper, N. (2019, June 18). Facebook Plans Global Financial System Based on Cryptocurrency. *The New York Times*. <https://www.nytimes.com/2019/06/18/technology/facebook-cryptocurrency-libra.html>

Islam, M., & Zibran, M. (2018). SentiStrength-SE: Exploiting Domain Specificity for Improved Sentiment Analysis in Software Engineering Text. *Journal of Systems and Software*, 145.

<https://doi.org/10.1016/j.jss.2018.08.030>

- Keynes, J. M. (1936). *The general theory of interest, employment and money* (p. 156). MacMillan.
- Kucuktunc, O., Cambazoglu, B. B., Weber, I., & Ferhatosmanoglu, H. (2012). A large-scale sentiment analysis for Yahoo! answers. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, 633–642.
- MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1), 13–39.
- Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1), 59–82.
- Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. *ArXiv Preprint ArXiv:1112.1051*.
- Misra, S., & Grewal, K. (2021, January 30). *The GameStop story — how a group of investors on Reddit gave Wall Street a wild week*. The Print. <https://theprint.in/theprint-essential/the-gamestop-story-how-a-group-of-investors-on-reddit-gave-wall-street-a-wild-week/595181/>
- Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Stanford University, CS229 (2011 Http://Cs229. Stanford. Edu/Proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. Pdf)*, 15.
- Nofer, M., & Hinz, O. (2015). Using twitter to predict the stock market. *Business & Information Systems Engineering*, 57(4), 229–242.
- Nofsinger, J. R. (2005). Social mood and financial economics. *The Journal of Behavioral Finance*, 6(3), 144–160.
- Perez, S. (2018, October 30). *Twitter’s doubling of character count from 140 to 280 had little impact on length of tweets | TechCrunch*. [Www.Techcrunch.Com. https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/](https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/)
- Pfitzner, R., Garas, A., & Schweitzer, F. (2012). Emotional divergence influences information spreading in twitter. *Proceedings of the International AAAI Conference on Web and Social*

Media, 6(1).

- Prechter, R. R., & Parker, W. D. (2007). The Financial/Economic Dichotomy in Social Behavioral Dynamics: The Socionomic Perspective. *Journal of Behavioral Finance*, 8(2), 84–108. <https://doi.org/10.1080/15427560701381028>
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS One*, 10(9), e0138441.
- Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*.
- Salinas, S. (2019, January 3). *Apple stock suffers biggest loss in 6 years after cutting guidance*. <https://www.cnbc.com/2019/01/03/apple-stock-falls-after-cutting-q1-guidance-on-weak-iphone-sales.html>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Shiller, R. J., Fischer, S., & Friedman, B. M. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 1984(2), 457–510.
- Shu, H.-C. (2010). Investor mood and financial markets. *Journal of Economic Behavior & Organization*, 76(2), 267–282.
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926–957.
- Thaler, R. H., & Barberis, N. (2002). *A survey of behavioral finance*. National Bureau of Economic Research.
- Thelwall, M. (n.d.). *SentiStrength Automatic Domain Independent Tool for Sentiment Analysis*. University of Wolverhampton. Retrieved October 15, 2020, from <http://sentistrength.wlv.ac.uk/>
- Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406–418.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social

web. *Journal of the American Society for Information Science and Technology*, 63(1), 163–173.

Todd, C. G., Bruce, B., Deeming, L., & Zobel, G. (2019). Short communication: Survival of replacement kids from birth to mating on commercial dairy goat farms in New Zealand. *Journal of Dairy Science*, 102(10), 9382–9388. <http://10.0.12.96/jds.2019-16264>

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Ying Lin. (2020, May 23). *10 Twitter Statistics Every Marketer Should Know in 2020 [Infographic]*. Wwww.Obrlo.Com. <https://www.oberlo.com/blog/twitter-statistics>

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.

Zheludev, I., Smith, R., & Aste, T. (2014). When can social media lead financial markets? *Scientific Reports*, 4(1), 1–12.

Software

Python Software Foundation. Python Language Reference, version 3.7. Available at <http://www.python.org>

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Thelwall, M. (n.d.). SentiStrength Automatic Domain Independent Tool for Sentiment Analysis. University of Wolverhampton. Retrieved October 15, 2020, from <http://sentistrength.wlv.ac.uk/>