UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER’S THESIS

INFORMATION DISSEMINATION VIA SOCIAL MEDIA: AN EXAMPLE OF BORIS JOHNSON'S TWEETS AND THE FINANCIAL MARKETS IN THE UK

Ljubljana, October 2021 MONIKA STOJKOVSKA

AUTHORSHIP STATEMENT

The undersigned Monika Stojkovska, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title Information dissemination via social media: an example of Boris Johnson's tweets and the financial markets in the UK, prepared under supervision of izr. Prof. Dr. Igor Lončarski

D E C L A R E

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge,non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, October 25th, 2021 Author's signature:

(Month in words / Day / Year, e. g. June 1st, 2012)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

sl. – Slovene

ADF – Augmented Dickey-Fuller

AIC – Akaike information criterion

API – Application Programming Interface

CAR – Cumulative abnormal return

DJIA – Dow Jones Industrial Average

EMH – Efficient Market Hypothesis

EMU – European Monetary Union

et al. – et alia

EUR – euro

ex. – example

FTSE – Financial Times Stock Exchange

GBP – United Kingdom Pound

GI – General Inquirer

GDP– Gross domestic product

i.i.d. – independent and identically distributed

KPSS – Kwiatkowski–Phillips–Schmidt–Shin

LM – Loughran Mcdonald

NLP – Natural language processing

NYSE – New York Stock Exchange

OLS – Ordinary least squares

RT – retweet

UK – United Kingdom

USA – United States of America

USD – U.S. Dollar

VAR – Value at risk

VIX – Chicago Board Options Exchange's CBOE volatility index

vs. – versus

INTRODUCTION

As human beings we have a hard time dealing with uncertainties. It is in our nature to look for patterns in order to predict the unknown. For thousands of years we have been perfecting models that predict the weather, earthquakes, sport event outcomes, election outcomes, market movements and so on. More of these predictions turn out to be wrong than right, but we never give up. Constantly looking for new, innovative ways to improve ourselves is also a part of what makes us humans.

After many years, stock market predictions remain to be an active area of research in the financial field. The Efficient Market Hypothesis (EMH), independently developed by Samuelson (1965) and Fama (1970), states that asset market prices are equal to fundamental prices, where the market prices reflect all available information. This implies that in the long run it is impossible to "beat the market", or in other words that stock predictions cannot outperform the random walk model. The hypothesis is widely accepted to hold true under the assumption that investors act rationally. Yet, recently a growing body of research has critically examined the EMH showing that stock prices do not follow a random walk and can therefore be predicted to some extent. Among the most successful attempts in predicting stock returns are Fama and Schwert (1977), Keim and Stambauch (1986), Campbell (1987), Fama and French (1988), Campbell and Shiller (1988), Lo and MacKinlay (1988). This brings into question the basic assumption of the EMH.

In addition to information, behavioral economists have shown that emotion (sentiment) has an important role in the decision making process and is a key factor in investor's herding behavior (Blasco, Corredor & Ferreruela, 2011; Jackson, Coulton & Dinh, 2015). Since then, a significant number of researchers have been focusing on analyzing the relationship between sentiment and the stock market. Among others, Wang, Keswani & Taylor (2004) theoretically show that the investor's sentiment is an innate factor that affects asset prices and therefore a measure of it might have predictive powers. In their paper De Long, Shleifer, Summers & Waldmann (1990) empirically show that sentiment is a part of the systematic risk that is priced. Later, Brown and Cliff (2004) discover a significant relationship between investor's sentiment and contemporaneous returns. Kumar and Lee (2006) look into the retail investor's sentiment and the return co-movements, while Baker and Wurgler (2006, 2007) examine the relationship between high and low levels of sentiment and returns. They discover that when the sentiment is low, returns are high and vice versa. Olaniyan, Stamate, Ouarbya & Logofatu (2015) find a significant relationship between positive sentiment and market volatility.

When it comes to analyzing the investor's sentiment, two main aspects need to be considered. First, it is of a great importance to acquire a relevant textual data source, from which the sentiment will be extracted. The recent technological revolution, accompanied with the widespread presence of computers and the internet, has created an unmatched source of information, forever changing the way we approach and analyze social and economical issues. The exponential growth of social media as a political and economical data news source has spawned research that regards its utility (Wu, Zheng & Olson, 2014). In this context, the growing popularity of Twitter as a micro-blogging website, where millions of users can share their opinions, follow and interact with each others, has attracted the attention of researchers from different fields. One of the biggest advantages of Twitter is its limited length of the characters (140) used to express the sentiment of the writer. Because of this, researchers have started to consider Twitter as a relevant source of textual data suitable for extracting indicators of public sentiment.

Among the first quantitative studies on Twitter sentiment and stock market prediction dates back to authors like Zhang, Fuehres & Gloor (2011), who measured the collective sentiment in form of hope and fear from large scale Twitter feed trying to predict stock market indicators such as Dow Jones, NASDAQ and S&P 500. They found that emotional tweet percentage is significantly negatively correlated with the stock indices, while positively correlated to the VIX index. In a similar fashion Bollen, Mao & Zeng (2010) investigate whether a collective sentiment measure extracted from large scale of Twitter feeds is correlated to the DJIA index over time. They discovered that the accuracy of DJIA predictions can be significantly improved by the inclusion of some of the public mood parameters. Vu, Chang, Ha & Collier (2012) look into the relationship between the public mood extracted from Twitter and NASDAQ tech companies. More recently, besides stock market indicators, researches have also used Twitter sentiment to predict other financial market movements such as: currency exchange rates, gold price and crude oil price (Zhang, Fuehres & Gloor, 2012).

After acquiring a relevant textual data source, the next step is to find a reliable and scalable way to assess investor's sentiment at a time scale appropriate for the market predictions. A lot of tools and techniques have been developed in the last few years, in order to achieve this (Das & Chen, 2007). The sentiment analysis techniques can be broadly defined as the analysis of textual data using natural language processing (NLP) tools in order to capture people's attitudes towards a certain topic. These methodologies can be quite diverse and vary from matching words in a text against positive and negative word lists to using mathematical algorithms to identify the sentiment of the texts.

The first approach is also known as the dictionary-based approach, where the analyzed text is considered to be a 'bag-of-words', because each word is analyzed separately without considering the order and thus the context of the text (Loughran & Mcdonald, 2016). The dictionary-based approach uses a mapping algorithm that first reads the text, separates it into words and then classifies them into dictionary categories that are previously defined (Li, 2010a). A wide variety of such pre-defined dictionaries exist going from general to finance specific ones. The second approach used for sentiment analysis is known as the machine learning approach. The machine learning approach uses statistical techniques to infer the content of texts and classify them according to statistical inference (Li, 2010a).

While this method has proven to be more accurate than the dictionary-based approach (Li, 2010b), it is also more complex to implement in practice.

In this master thesis, I will test a hypothesis based on the premise of behavioral finance, which says that the emotion of investors can affect their decision making process, thus, leading to a direct correlation between the public and market sentiment. In order to achieve this, the first part of the study will focus on a sentiment analysis performed on the publicly available tweets of Boris Johnson. Boris Johnson was elected as a Conservative leader and appointed prime minister of the UK, after the resignation of Theresa May in 2019. With this he became a prominent figure in the Brexit negotiation process, gaining great political and economical power. All of his actions became closely monitored by the decision makers. Considering Boris Johnson's great influence and the high frequency with which he uses Twitter as a social media outlet to share his thoughts and opinions with more than 3 million followers, I decided to extract the sentiment from his tweets in order to construct a measure of collective investor's sentiment. For this purpose, all of Boris Johnson's tweets for the time period of 8 months, beginning from 01 June 2019 to 01 February 2020, will be analyzed. Although he assumed office on 23 July 2020, tweets from the previous period will also be taken into account, since the decision makers might have paid bigger attention to them after he was nominated as one of the 10 candidates to replace Theresa May on 10 June 2020.

The second part of the thesis will focus on analyzing the relationship between the tweet sentiment and the financial markets in the UK. More specifically, I will be looking at the effects on the stock market and the exchange rate market. For the initial investigation of the relationship during the whole time period, I will apply the Pearson correlation and Granger correlation tests. Then, in order to analyze the relationship between the tweet sentiment and the UK stock market over a shorter period of time, I will adapt the well known "event study" from economics and finance (MacKinlay, 1997) to the analysis of the given Twitter data.

The purpose is to assess whether the sentiment derived from Boris Johnson's tweets can add any significant information to the predictability of the UK financial markets. More specifically I will study the relationship between the tweet sentiment and the UK stock market represented by the FTSE 100 Index and the exchange market represented by some exchange rates: GBP/EUR and GBP/USD. As a robustness check, I will also consider the effect of the tweet sentiment on the European stock market, represented by FTSE 100 Euro Index and the world market, represented by the FTSE All World Index, as well as the exchange rate EUR/USD.

The goal is to better understand the role of investor's sentiment in the movements of financial markets. If the relationship between the public sentiment and market sentiment turns out to be statistically significant, a new sentiment variable can be introduced to the existing predictive models, in order to improve their accuracy. Moreover, the goal of this thesis is to contribute to the existing literature analyzing the relationship between public sentiment extracted from Twitter data and movements of financial markets. While most of the empirical analysis done so far focuses on public sentiment measures constructed by analyzing a large scale of Twitter feeds, this thesis will focus on a single Twitter feed of an influential politician. The goal is to see if a different approach would lead to similar results. Additionally, finding significant results would go in line with the recent behavioral findings and might lead to further research that puts into question the classical theory.

1 LITERATURE REVIEW

Since the development of the Efficient market hypothesis, it has been believed that due to investor's rational behavior asset market prices fully reflect all available information and are therefore always at their fundamental value. In his paper Fama (1970) states that the term 'fully reflect' is very general and in order to be empirically testable it must be defined more precisely. The equilibrium expected return can be defined as:

$$E\left(p_{j,t+1|\varphi_t}\right) = \left[1 + E\left(r_{j,t+1|\varphi_t}\right)\right]p_{j,t} \tag{1}$$

where $p_{j,t}$ is the price of an asset j at time t, $r_{j,t}$ is the one period percentage return of the asset j given by $(p_{j,t+1} - p_{j,t})/p_{j,t}$, E is the expected value and φ_t the information set at time t (Fama, 1970).

The implication of Equation 1 is that when determining the equilibrium expected returns, the infomation set φ_t is fully utilized. In this sense the formation of the asset's price p_j 'fully reflects' the information set φ_t (Fama, 1970). Since the equilibrium expected returns reflect all available information it becomes impossible to "beat the market", or in other words predictions about the asset return cannot outperform the random walk model. The notion that no profit or return can be achieved in excess of the equilibrium expected return points to the conclusion that the development of market prices is a fair game and can therefore not be traded on.

The classical theory rules out the sentiment's predictive power by definition. Any movements away from the fundamental value that might occur as a result of emotions like euphoria or fear, would only be short lived and appear as a noise, if at all (Bormann, 2013). Since investors react rationally to all of the available information, the right expectations will always be formed on average. From this it follows that sentiment indicators cannot have any predictive power (Feldman, 2013).

Contrary to the widely accepted EMH, behavioral economists have shown that some part of the investor's decisions that is driven by emotion leads to a different result than the one predicted by the classical theory. Among the first researches who have disregarded the assumption of rationality and empirically shown that sentiment is a part of the systematic risk being priced are De Long, Shleifer, Summers and Waldmann (1990). According to the EMH assets should always be traded at their fundamental value, yet in their paper De Long, Shleifer, Summers & Waldmann develop a simple model in which prices can deviate from their fundamental value due to the unpredictable irrational behavior of investors. The model differentiates between two types of traders trading with either a risky or a risk free asset. The first type of traders are called arbitrage traders, who always act rationally according to fundamentals. In contrary to the rational invertors the noise traders trade irrationally. Their actions are affected by sentiment which is manifested in their beliefs of price development. Depending on their inner sentiment the noise traders can either be optimistic (bullish) or pessimistic (bearish) about the price development. As a result, they can push the price in either an upward or downward direction, away from its fundamental value. Due to the fact that the noise trader's sentiment is stochastic and therefore not predictable, it creates an additional risk in the asset price that prevents rational traders from aggressively betting against them and with that offsetting their effect. As a result, high fluctuations in the asset prices exit, moving them away from the fundamental value, even in the absence of fundamental risk. What is the most interesting is that the bearing of additional amount of risk that the noise traders themselves create enables them to earn a higher expected return than the sophisticated investors do, even though they are the ones who distort the prices. This result highlights the need for a closer inspection of the standard theory, which says that destabilizing speculation must be unprofitable and therefore noise traders do not persist on the markets.

From that point on, there has been a great refinement and development in the field of behavioral finance. The goal of behavioral economists is to find where the Efficient market hypothesis falls short and to discover other possible explanations for the stock price movements other than the random walk model. In this aspect, a significant number of researchers have been focusing on analyzing the relationship between investor's sentiment and the stock market.

Wüthrich, Permunetilleke, Leung, Cho, Zhang & Lam (1998) try to predict the stock market by using information from influential news articles published on the web. They analyze the connection to the closing values of major stock market indices in Europe, Asia and USA. Additionally, they suggest a trading strategy that exploits the textual information and show that such trading would outperform the strategy of stock fund managers.

Similarly, Chan (2003) examines monthly stock returns following public news by analyzing a large sample of newspaper headlines for randomly selected group of companies. He tests the hypothesis that the sentiment of investors with relation to the public news leads to abnormal returns. He finds that negative news seem to lead to underperformance, while positive news lead to a smaller drift the stock prices.

In their paper Baker & Wurgler (2006) investigate how investor's sentiment affects the cross section of stock returns. They find out that the effect of sentiment is bigger for those stocks whose values are highly subjective and therefore harder to arbitrage. Consistent with this, they prove that when the proxies for sentiment are low at the beginning of the period, subsequent returns are high for small stocks, young stock, stock with high volatility, low profitability and those stocks that show extreme growth. Contrary to this, when the beginning of period sentiment proxy is high, these categories earn low subsequent returns. Furthermore, when controlling for the three factor model of Fama and French (1993), they discover that the sentiment proxy has statistically significant predictive power, a finding that goes directly against the Efficient market hypothesis. Their study covers a time period of forty years, so the significant role of sentiment exceeds the short period horizon and is also valid on the longer run.

Similar results are obtained later in the study of Baker & Wurgler (2007). This time they reason that the classical financial models fail to explain big financial crises like the Great Crash of 1928 or the Dot.com bubble. For the purpose of finding out the real cause they develop a sentiment index that is able to capture the sentiment volatility around speculative major events like the one of the Dot.com crisis. Moreover, by relying upon sentiment and limits to arbitrage they try to explain which stocks are most likely to be affected by sentiment. The results obtained in this study are very similar to what they conclude in 2006. More precisely they discover that low capitalization, low profit, young, highly volatile and non- dividend stock of growth companies are more likely to be excessively sensitive to investor's sentiment.

The approach developed by Baker & Wurgler (2007) is later followed by Finter, Niessen-Ruenzi & Ruenzi (2012), who design a sentiment index for Germany. However, in their research they do not find a significant predictive power of the sentiment for future stock returns. They reason that their results might be due to the fact that in Germany the proportion of sophisticated traders compared to noise traders is bigger than the one in the USA. This finding highlights the importance of the relation between rational and noise traders for the significance of the sentiment as a market predictor.

Brown and Cliff (2005) construct a sentiment measure by using the media as a source, while focusing on the market newsletters. In their paper they analyze the relationship between the sentiment measure and stock market returns. The idea is to see whether there is an overvaluation of the stock market during optimistic periods, followed by periods characterized by low returns due to the reversal of the market prices to their fundamental value. They find that the sentiment measure is useful for predicting the market returns for a time period of 1 to 3 years. More specifically, the market proves to be overvalued during periods of optimism and undervalued during periods of pessimism. Furthermore, the sentiment measure has the ability to explain the deviations of market prices from their fundamental value. This goes in line with the model of De Long, Shleifer, Summers & Waldmann (1990), which assumes that noise investors who are either optimistic or pessimistic distort the marker prices driving them away from the fundamental value. The significance of their result passes the robustness check for the usual rational factors and different changes in the methodology.

Tetlock (2007) follows the logic of previous researchers that the media might have a notable impact on investor's behavior and thus influence the stock market returns in a significant way. More specifically, he looks into the relationship between the sentiment measure constructed from the contents of a popular Wall Street Journal and the stock market returns. His findings show that the high pessimism reflected in the media puts a downward pressure on the stock market prices followed by a period in which prices revert to their fundamental value. Moreover, he finds that extreme values of sentiment in the media predict high market trading volume. Such results go in line with the DeLong Shleifer, Summers & Waldmann (1990) model of noise and rational traders. The impact that the pessimism has on the market prices appears to be particularly large and slow to reverse for smaller stocks, which is consistent with the sentiment theory. On the other hand, the results show a weak connection of market sentiment with volatility.

Zhang & Skiena (2010) provide evidence that the news data is highly informative, by studying a large scale of newspapers for over four years. They exploit the significant results and design a trading strategy which proves to give favorable results with low volatility for the time period between 2005 and 2008. Similar results are obtained by Uhl, Pedersen and Malitius (2015), who combine company and macro specific news sentiment extracted from 100,000 news pieces per week in order to calculate momentum in news sentiment and with that design an approach for tactical asset allocation. Their results show that this method provides a valid outperformance in most of the years during the period between 2004 and 2014.

Mao, Counts and Bollen (2011) try to compare sentiment measures constructed by using different sources and evaluating their predictive power for several market indicators, including the DJIA index, gold prices, trading volumes, as well as volatility of the stock market measured by the VIX index. As a ground for their analysis, they look into several sources used for the construction of the sentiment measures, such as different online surveys, news headlines, data from search engines, as well as Twitter feed data. The news sentiment is measured by focusing on the tone of the words in the financial headlines, found in different newspapers. The Twitter sentiment is measured by taking the ratio of tweets with words that have bullish vs. those that have bearish tone, combined with the volume of 26 search queries. The same queries are used for calculating the average of the search volume from Google, in order to construct the sentiment measure for the Google search engine. VAR models and the Granger causality test are used for the purpose of determining if such sentiment measures can serve as predictors for the market indicators. Their findings point toward an insignificant relationship between the survey sentiment and the market indicators, contrary to the other sentiment sources, which prove to be useful for predicting the market indicators. All things considered, the Twitter sentiment has proven to outperform the rest of the sentiment measures.

Yuan (2012) looks into the effect of important market related events on the investor's behavior, their trading and subsequently on the market returns. More specifically, he examines events for some market indices, including the S&P 500 Index, NYSE Composite Index, NASDAQ Composite Index and the DJIA index, along with the front pages of the New York Times and the Los Angeles Times, reporting those events. His goal is to inspect the relationship of such news with the trading patterns of investors and eventually the market returns. The results show that when the market is high, important events reported in the newspapers cause the investors to dramatically sell their stock holdings. Such aggressive selling leads to a disruption in the market prices, which are negatively impacted causing a reduction in market returns by 19 basis points on days coming after the important events.

Garcia (2013) focuses on studying the effect of sentiment on asset prices for a longer period of time that goes from 1905 to 2005. His sentiment measure is based upon the fraction of positive and negative words in the columns of financial news, published in the New York Times. The main finding is that the sentiment extracted from news is useful for predicting the stock returns at daily frequency, a relationship that is strongly emphasized during recessions. More precisely, if one standard deviation change in the sentiment measure leads to a 12 basis points change in the DJIA index in times of recession, the effect in times of expansion is only 3.5 basis points. Moreover, he discovers that the effect is stronger during the weekends probably because investors have more time to read and consequently react to the news. Nevertheless, the impact of news only last for approximately four days, after which asset prices revert to their fundamental value.

Similarly to the study of Mao, Counts and Bollen (2011), Yu, Duan and Cao (2013) attempt to compare different sources of sentiment. First they examine the effect of the conventional media while focusing on the most influential newspapers, television companies and financial magazines. Then, they look into the social media sources such as forums, blogs and micro blogs (ex. Twitter). They try to understand the correlation between a sentiment measure defined as the difference between the number of positive and negative words in the news with the performance of the stock market on the short horizon. While they do find a significant relationship between the sentiment extracted from the conventional media and the stock markets, the blog and Twitter sentiment prove to have an event stronger impact.

Da, Engelberg, Joseph and Gao (2015) construct a measure of collective sentiment by analyzing a daily internet search volume from millions of people. By aggregating the volume while concentrating on queries that are related to concerns such as recession, unemployment and bankruptcy they construct a so called “fears” index that serves as a new measure of investor sentiment. They find that the fears index is useful in predicting the short term return reversals, short term increases in volatility and fund flow from equity to bond funds. This results are consistent with the theories of behavioral finance.

Ferguson, Philip, Lam & Guo (2015) find a significant relationship between the positive/negative tone and volume of company specific news in the UK media and stock returns for the period from 1981 to 2010. The volume seems to have a bigger effect than the tone. Additionally, the predictive power of tone is more pronounced among lower visibility firms.

All of the mentioned authors above seem to agree that there is a significant impact of news on the investor's sentiment and with that consequently on their trading behavior. This type of behavior causes changes on the markets, not predicted by the classical theory. However, as time changes there is also a shift in the perspective as to which news source is the most relevant one triggering the investor's sentiment. Among others, Mao, Counts and Bollen (2011) and Yu, Duan and Cao (2013) have shown that when comparing different news sources, micro blogs such as Twitter have proven to be the best predictors of market movements. Since then researchers have considered Twitter as a relevant source for extracting the investor's sentiment.

Some of the first quantitative studies on Twitter sentiment and stock market prediction include Bollen, Mao & Zeng (2010) who are interested in whether societies can experience collective mood states that impact their decision making process and consequently with that if they can have any predictive power of the market indicators. In order to get an answer to these questions they investigate whether a collective sentiment measure extracted from large scale of Twitter feeds is correlated to the DJIA index over time. The sentiment is measured by mood tracking tools and represented in 6 dimensions: calm, alert, sure, vital, kind and happy. Granger causality test and a machine learning model are used in order to test their hypothesis. In the end, they discover that the accuracy of DJIA predictions can be significantly improved by the inclusion of some of the public mood parameters.

In a similar fashion, Zhang, Fuehres & Gloor (2011) focus on Twitter feed data for a period of six months in order to create a randomized sample as a portion of the full volume of the collected tweets. They measure the sentiment as a collective hope or fear and analyze the relationship with different market indicators. More precisely, they focus on the Dow Jones Index, NASDAQ Index and S&P 500 Index. The results suggest that on one hand there is a significant negative correlation between the emotional tweet percentage and the stock indices, while the correlation with the VIX is significant and positive. They conclude that the emotions conveyed through Twitter can be a useful predictor of how the stock market is going to move the next day.

Papaioannou, Russo, Papaioannou & Siettos (2013) also use the Twitter social networking platform in order to extract early indicators of market trends. More specifically, they concentrate on modeling and predicting the EUR/USD exchange rate on an intraday scale with a high frequency. Their results show that the information extracted from the Twitter platform in certain cases improves the forecasting efficiency when regarding the short intraday period.

Smailović, Grčar, Lavrač and Žnidaršič (2013) look into the relationship between the public opinion extracted from Twitter feed data and the movements in stock closing prices. Sentiment is measured in terms of polarity which can either be positive or negative. They find that the polarity is useful for predicting the stock price movement few days in advance. Hence, using this information improves the effectiveness of the models.

Pagolu, Reddy, Panda & Majhi (2016) follow the logic that nowadays the public sentiment is perfectly captured by social media, especially by the famous microblogging platform Twitter. They are interested in the connection of public sentiment about a certain company expressed in the tweets and the stock prices of that particular stock. They find that positive news reflected in the tweets encourages investors to buy the stocks of a company and consequently lead to an increase in the stock's price. The same holds true for negative news, which result in a fall in the stock price. Similar results are also shown in the research of Xu & Cohen (2018).

If we look at the literature so far, it can be observed that when analyzing the effect of public sentiment on investor's behaviour and consequently on the market indicators, researchers first concentrated on conventional news sources. Then, as times changed, their focus shifted on news broadcasted via social media. This is well expected if we take into account the role and impact of social media in our everyday lives. In these studies the main focus is put on the sentiment of noise traders, as they have the biggest potential to move prices away from their fundamental value, as De Long (1990) predicted. While, most of the studies point toward a significant role of sentiment in predicting the movements in financial markets, the design of a general predictive model that would exploit such findings faces few challenges, since there are many different news sources, sentiment extraction techniques and time spans that can be considered. Nevertheless, due to the huge amount of evidence that points towards such a significant relationship, it becomes essential to further analyze, work towards discovering the best techniques and in that way optimizing the effectiveness of the existing predictive models.

2 DATA

The analysis in this master thesis focuses on investigating the relationship between the Twitter and market data. Details of both are given in the following section.

2.1 Twitter Data

The first data source for the purpose of the analysis is Twitter, or more specifically Boris Johnson’s Twitter feed. By using Twitter API and the Python’s library ‘Tweepy’ all of

Boris Johnson's tweets for the time period of 8 months, beginning from 01 June 2019 to 01 February 2020, are extracted. This amounts to a total of 1.439 tweets.

Table 1: Example of Boris Johnson's tweets

Date	Tweet
01.02.2020	b'Thank you to @SteveBarclay, @LordCallanan, @JamesDuddridge and to everyone at DExEU for all of your hard work and d\xe2\x80\xa6 https://t.co/ztdrsJfeGc'
01.02.2020	b'RT @Conservatives: \xe2\x80\x9cWhen I look at the potential of this country waiting to be unleashed, I know that we can turn this opportunity into a s\xe2\x80\xa6'
01.02.2020	b'RT @BorisJohnson: Tonight we are leaving the European Union. https://t.co/zZBsrf4BLe'
01.02.2020	b"As we embark on a new chapter, let's come together and build a brighter future for the next generation. https://t.co/0VHVPfv5sp"
31.01.2020	b'Tonight we have left the EU - an extraordinary turning point in the life of this country. Let us come together now\xe2\x80\xa6 https://t.co/AHuez5IULK'
31.01.2020	b'RT @10DowningStreet: WATCH LIVE: We are counting down to leaving the EU ahead of our departure at 11pm https://t.co/xgV5BrtLQS'
31.01.2020	b'Tonight we are leaving the European Union. https://t.co/zZBsrf4BLe'
31.01.2020	b'RT @10DowningStreet: PM @BorisJohnson chaired the last Cabinet meeting before we leave the EU in Sunderland. \n\nHe visited local businesses\xe2\x80\xa6'
31.01.2020	b'This government will unite and level up our country.\n\nAnd as we build a new relationship with the EU, I urge everyo\xe2\x80\xa6 https://t.co/89qveMWeJJ'
31.01.2020	b'Tonight at 10pm I\xe2\x80\x99ll be giving an address to the nation, which you can watch on my Facebook page: https://t.co/ido2hy9GGX'
31.01.2020	b"RT @Conservatives: \xf0\x9f\x92\xac What: The Prime Minister's Address to the Nation.\n\xe2\x8f\xb0 When: 10pm tonight.\n\xf0\x9f\x93\xb2 Where: @BorisJohnson's Facebook page.\n\n#Brex\xe2\x80\xa6"
31.01.2020	b'RT @10DowningStreet: Today is the day the UK is leaving the EU. https://t.co/Zuhzl0l5sT'
30.01.2020	b"RT @10DowningStreet: Today PM @BorisJohnson hosted a Kids' Question Time at Downing Street. Watch what happened here. https://t.co/qvqzaq64\xe2\x80\xa6"
30.01.2020	b'Today I invited children from across our country to Downing Street to ask me their questions about the future of th\xe2\x80\xa6 https://t.co/99OFt9XX36'
29.01.2020	b'Youth knife crime is a massive issue for this country. We need to tackle it. #PeoplesPMQs https://t.co/R6JYU0OIcw'
29.01.2020	b'What I will be doing when we leave the EU at 11pm this Friday. #PeoplesPMQs https://t.co/KrQ4C2f6Oe'
29.01.2020	b'Thank you for your questions this evening for the #PeoplesPMQs. \n\nTune in for my address this Friday at 10pm. https://t.co/FHyNQJ4AUN'
29.01.2020	b"RT @Conservatives: 2 days.\n\xf0\x9f\x91\x87\nUntil #BrexitDay \n\xf0\x9f\x91\x87\nWe're getting it done. https://t.co/0bf7Uidnrc"
28.01.2020	b'I\xe2\x80\x99m delighted to hear @Sainsburys is leading the way for UK business by pledging to go carbon neutral, backed by si\xe2\x80\xa6 https://t.co/IOxJ2lhBxL'
28.01.2020	b'Fantastic to visit @kingsmathschool yesterday and meet with these bright students and their teachers. \n\nWe\xe2\x80\x99re makin\xe2\x80\xa6 https://t.co/C8HtLtMT7A'

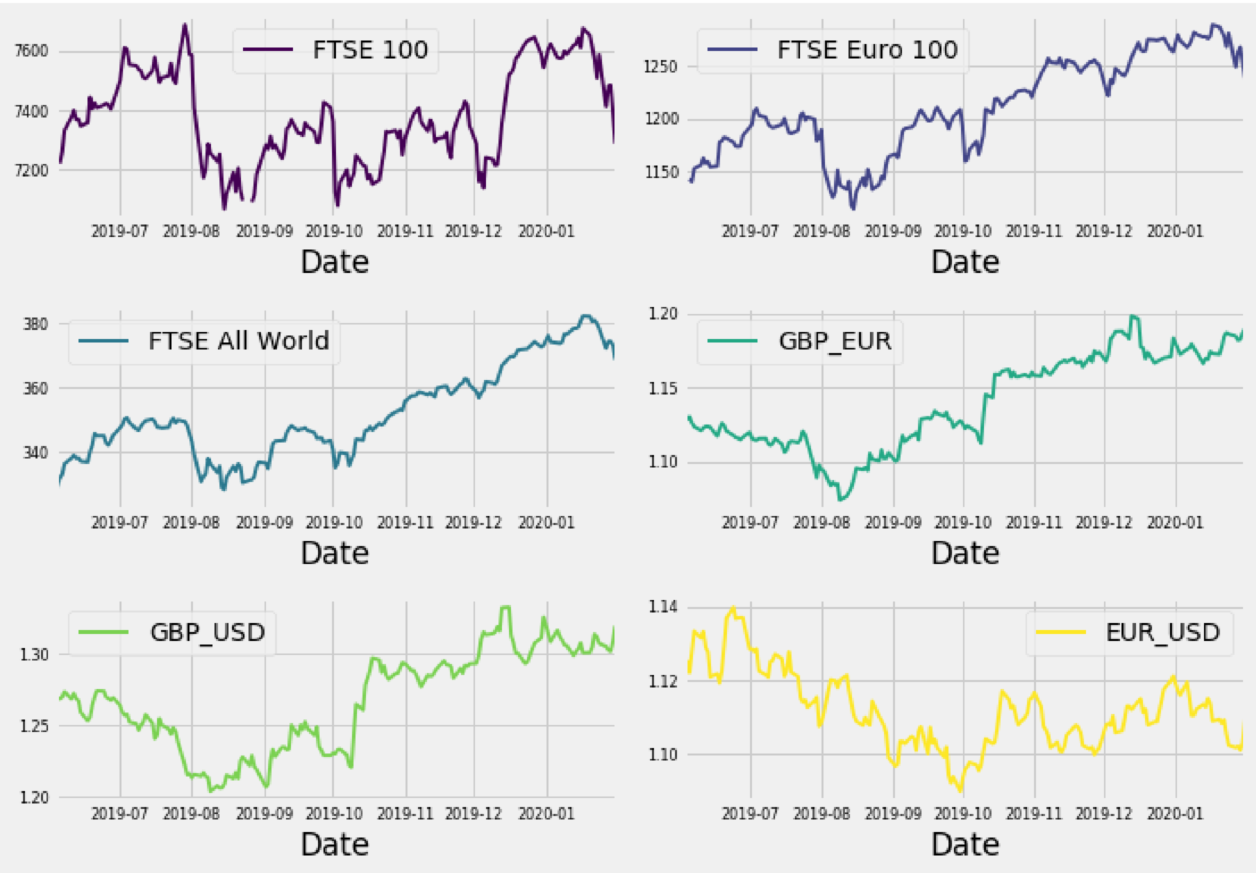
Source: Own work.

Table 1 shows an example of the last 20 tweets from the dataset.

2.2 Market Data

The analysis is conducted on 3 major stock indices: FTSE 100 Index, FTSE Euro 100 Index and FTSE All World Index, that serve as proxies for the UK, European and the global stock market respectively. The Financial Times Stock Exchange 100 Index (FTSE 100), also known as "Footsie", is a share index that represents the performance of the 100 highest capitalization companies listed on the London Stock Exchange. With a market capitalization of nearly £ 2 trillion (May, 2020), it is considered to be the performance benchmark for most investors and is most widely used as a UK stock market indicator. The FTSE Euro 100 Index represents the performance of the 100 highest capitalization blue chip companies in Europe countries that are part of the European Monetary Union (EMU) and is used as an indicator for the stock market in Europe, while the FTSE All World Index tracks the performance of approximately 3.900 large and mid-capitalization companies in nearly 50 countries around the world, covering both developed and emerging markets. The FTSE All World Index covers more than 95% of the global investible market capitalization, which makes it a good indicator for the world stock market.

Figure 1: Market index and exchange rate prices



Source: Own work.

Apart from the stock indices the data also includes 3 exchange rates: GBP/EUR, GBP/USD and EUR/USD. The daily closing prices are downloaded from investing.com and cover the same time period as the Twitter data. This leads to a total of 171 daily prices for the FTSE 100 index, 172 daily prices for the FTSE Euro 100 index and 175 daily prices for the FTSE All World Index and all of the exchange rates. The daily index and exchange rate prices, for the period between 01 June 2019 and 01 February 2020 are shown on Figure 1.

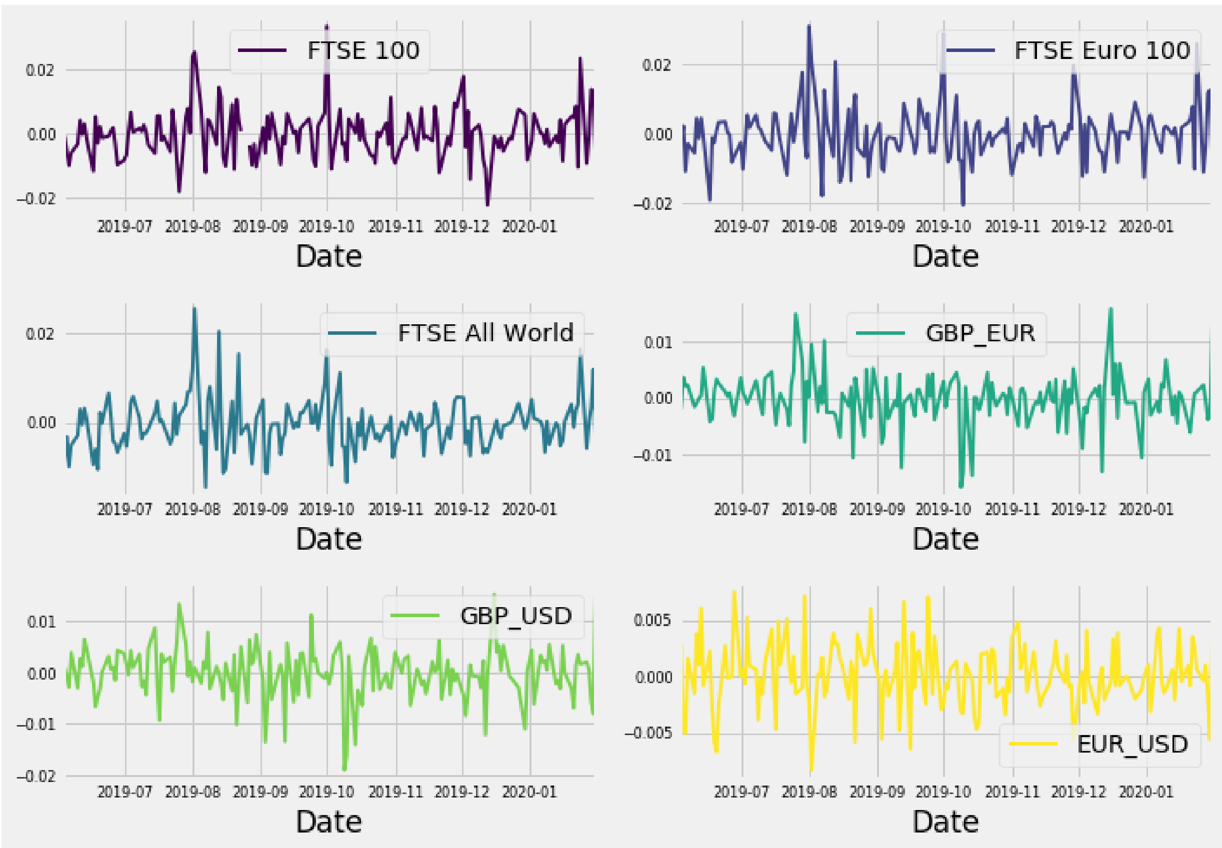
For each index and exchange rate the time series of daily returns is calculated as follows:

$$R_d = \frac{P_d - P_{d-1}}{P_{d-1}} \tag{2}$$

where P_d is closing price at day d.

The daily index and exchange rate returns, for the same period, are shown on Figure 2.

Figure 2: Market index and exchange rate



Source: Own work.

3 METHODOLOGY

This section firstly describes the methods most commonly used for sentiment analysis, while focusing on the dictionary-based approach. Then, it introduces the Pearson correlation and Granger causality analysis, methods used for determining the relationship between the variables of interest over the whole time period. Finally, it describes the event study methodology in debt, by firstly looking into the event detection process, followed by the estimation of normal returns, along with the calculation of cumulative abnormal returns and their statistical validation.

3.1 Sentiment Analysis

The most commonly used methods for sentiment textual analysis are the dictionary-based approach and the machine learning method (Kearney and Liu, 2014).

In natural language processing, the dictionary-based approach is also known as the 'bag-of-words' method, where the analyzed text is considered to be the bag of words, while the order, and thus context, of the words is completely ignored (Loughran & Mcdonald, 2016). This is because the dictionary-based approach uses a mapping algorithm that first reads the text, separates it into words and then classifies them into dictionary categories that are previously defined (Li, 2010a). A wide variety of pre-defined dictionaries exist. One of the most frequently used general English language linguistic built-in dictionary is the General Inquirer (GI), designed by Philip Stone, a specialist in social psychology. Most of the word lists in the GI come from the Harvard IV dictionaries. While the use of such dictionaries can be convenient for some analysis, there might be a problem when they are applied to a research from the finance specific field. The issue is that what might be considered to be positive or negative in general terms might not necessarily be viewed as such in financial terms. For example, words like ‘tax’, ‘debt’, or 'liability' are placed on the negative word lists of the general dictionaries, but are not neccessarily considered as negative in the financial context (Loughran & Mcdonald, 2016). Furthermore, Loughran & Mcdonald (2011) have shown that almost 73,8% of the negative word counts in the GI/Harvard lists are typically not negative when observed from the financial point of view. In order to overcome the problem of the difference between what is included and considered as positive/negative in the general vs. the finance field, dictionaries specific to the finance domain have been built by researchers for the purpose of generating more accurate and efficient sentiment scores. The LM lists, developed by Loughran and Mcdonald, are most commonly used in recent studies, including the ones of Garcia (2013), Jegadeesh and Wu (2013). When comparing different dictionaries, researchers show that such context specific dictionaries are more powerful than the general ones used in previous research.

One of the biggest advantages of the dictionary-based approach is its simplicity. Due to the availability of well established programs like GI, the dictionary-based approach is most commonly used in the literature and gives a more straightforward base for replicating the analysis of other researches. Another thing is that by using a dictionary the researches subjectivity is avoided. Since the computer program tabulates the frequency count of words, this method can be applied to large samples. However, general dictionaries are not suitable for textual analysis in the financial context. This problem, as previously discussed, can be largely avoided by using finance-specific word lists (such as the LM lists). Because of the fact that each word in the text gets analyzed separately ("bag-of-words"), the structure along with any linear ordering of words within the context is ignored, which can consequently lead to a smaller accuracy rate for this method.

The second approach used for sentiment analysis is the machine learning approach. The machine learning approach uses statistical techniques to infer the content of texts and classify them according to statistical inference (Li, 2010a). In order to do so, first the text is divided into two parts: a 'training set' and a 'testing set'. Each word, phrase, or sentence in the 'training set' is manually labeled as positive, negative or any other possible dimension of sentiment. A sentiment analysis algorithm is then trained on the labeled 'training set'. The goal is for the algorithm to learn the sentiment classification rules on the training set and then later apply those rules on the rest of the corpus.

The main advantage of the machine learning approach is given by the higher accuracy rate. It has usually proven to be more accurate than the dictionary-based approach (Li, 2010b). In the research of Huang et al. the classification accuracy achieved using the Naïve Bayes machine learning approach is 80.9% in the in sample validation and 76.9% in the out-of-sample validation. This is substantially higher than what they achieved using the dictionary-based approach based on the general GI dictionary, which was 48.4%. However, the implementation of the machine learning method is more time consuming and costly, since the text in the 'training set' must be manually classified. Furthermore there is a strict criterion for selecting the people who read the 'training set', which adds more complexity to the method.

Due to the complex nature of the machine learning method, for the purpose of this master thesis, I will use the dictionary-based approach or more specifically the LM dictionary approach.

3.1.1 LM Dictionary Approach

The procedure for the tweet sentiment classification, using the LM dictionary approach follows the following order:

- First, the raw tweets need to be preprocessed. More specifically, the tweets are cleaned out of any signs, numbers, links or hashtags that do not add any meaning to the text. Tagged users (e.g. '@SteveBarclay') are also removed, in order to make sure that the tweets are independent of other users included in the discussion. Then, the tweets are tokenized, meaning that they are converted into a list of lower case words. The next step is to remove any stop words (such as 'a', 'an', 'the'), pre-defined in the NLTK library in Python, along with the 'RT' mark in front of the re-tweets. The idea is create a more meaningful textual data source from which the sentiment is going to be extracted. An example of the process for the last 5 tweets in the dataset is shown in Table 2.

- The next step is to is to count the number of positive and negative words in all of the tweets. For this purpose, first the positive and negative LM dictionaries need to be loaded in the program. The LM negative dictionary consists of 2355 words and the LM positive dictionary consists of 354 words. Then, a function that makes the program check if the words are in the dictionary needs to be created and if this is the case they will be added it to a list and assigned a value of their own frequency. The end result is a list of all positive and negative words for every tweet, along with the frequencies with which they appear. The frequencies then are summed up for each tweet, in order to calculate the total number of positive and negative words per tweet.

- The tweets are then labeled as: positive, negative or neutral. If the number of positive words in a tweet is bigger than the number of negative words, the tweet is labeled as positive, if the opposite is true the tweet is labeled as negative and if the number of positive and negative words is the same the tweet is labeled as neutral. The end result is a total of 905 neutral tweets (63%), 334 positive tweets (23%) and 200 negative tweets (14%).

- Finally, since the sentiment of tweets on a daily basis is of the interest for the analysis, a measure of the daily polarity is calculated as follows:

$$P_d = \frac{tw_d^+ - tw_d^-}{tw_d^+ + tw_d^-} \tag{2}$$

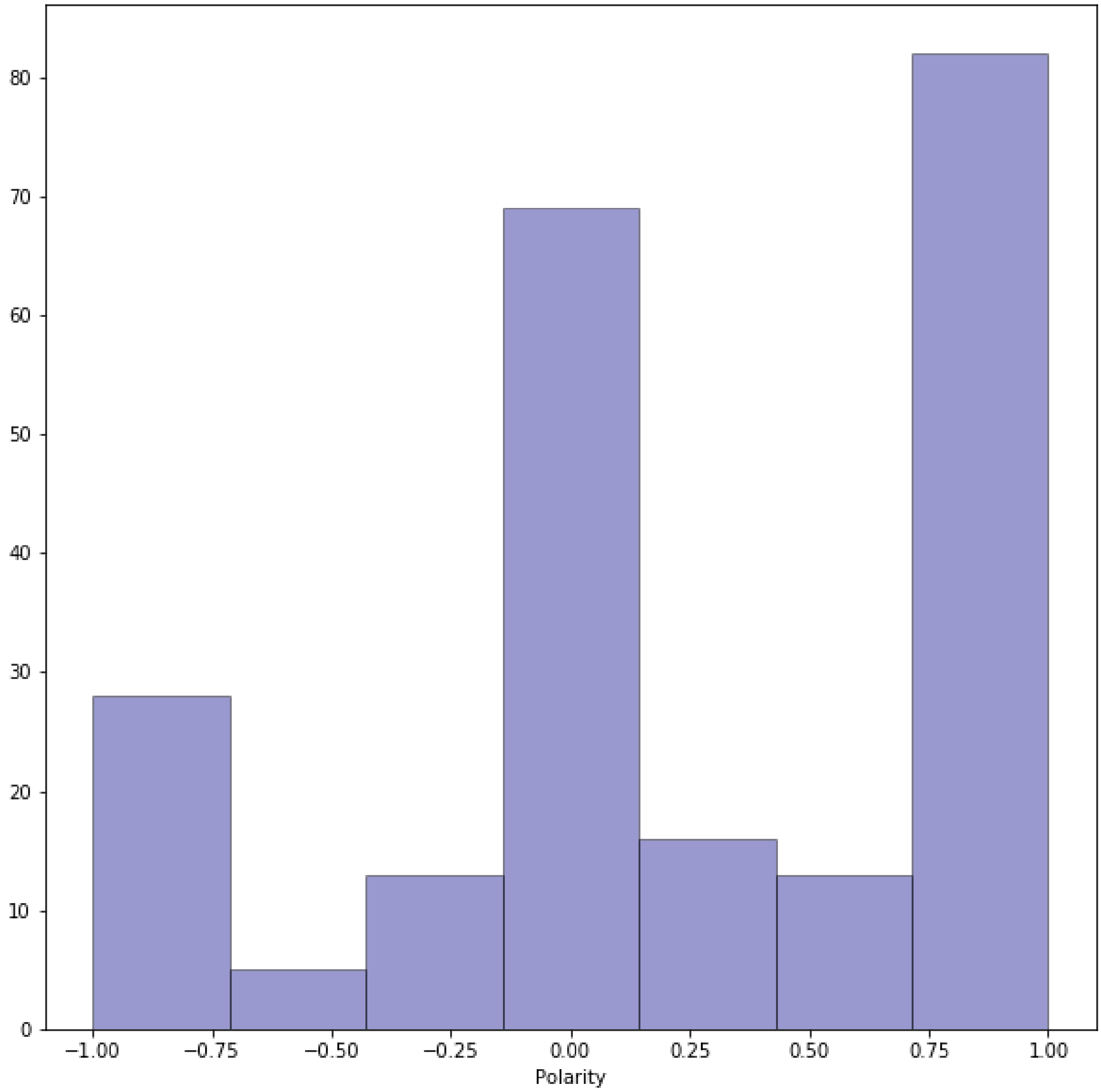
where tw_d^+ is the number of positive tweets in a day, tw_d^- is the number of negative tweets in a day and $tw_d^+ + tw_d^-$ the number of non-neutral tweets. Tweet volume is defined as the number of total tweets per day (Zhang and Skiena, 2010).

Table 2: Example of preprocessing the Boris Johnson's tweets

Date	Raw Tweet	Without hex	Without links	Without taggs	Tweet tokenized	Tweet clean
01.02.2020	b'Thank you to @SteveBarclay, @LordCallanan, @JamesDuddridge and to everyone at DExEU for all of...	b'Thank you to @SteveBarclay, @LordCallanan, @JamesDuddridge and to everyone at DExEU for all of...	b'Thank you to @SteveBarclay, @LordCallanan, @JamesDuddridge and to everyone at DExEU for all of...	b Thank you to and to everyone at DExEU for all of your hard work and d	[b, thank, you, to, and, to, everyone, at, dexeu, for, all, of, your, hard, work, and, d]	[thank, everyone, dexeu, hard, work]
01.02.2020	b'RT@Conservatives: \xe2\x80\x9cWhen I look at the potential of this country waiting to be unle...	b'RT @Conservatives: When I look at the potential of this country waiting to be unleashed, I kno...	b'RT@Conservatives: When I look at the potential of this country waiting to be unleashed, I kno...	b RT When I look at the potential of this country waiting to be unleashed I know that we can..	[b, rt, when, i, look, at, the, potential, of, this, country, waiting, to, be, unleashed, i, kno...	[look, potential, country, waiting, unleashed, know, turn, opportunity]
01.02.2020	b'RT @BorisJohnson: Tonight we are leaving the European Union. https://t.co/zZBsrf4BLe'	b'RT @BorisJohnson: Tonight we are leaving the European Union. https://t.co/zZBsrf4BLe'	b'RT @BorisJohnson: Tonight we are leaving the European Union. , '	b RT Tonight we are leaving the European Union	[b, rt, tonight, we, are, leaving, the, european, union]	[tonight, leaving, european, union]
01.02.2020	b"As we embark on a new chapter, let's come together and build a brighter future for the next ge...	b"As we embark on a new chapter, let's come together and build a brighter future for the next ge...	b"As we embark on a new chapter, let's come together and build a brighter future for the next ge...	b As we embark on a new chapter let s come together and build a brighter future for the next gen...	[b, as, we, embark, on, a, new, chapter, let, s, come, together, and, build, a, brighter, future...	[embark, new, chapter, let, come, together, build, brighter, future, next, generation]
31.01.2020	b'Tonight we have left the EU - an extraordinary turning point in the life of this country. Let ...	b'Tonight we have left the EU - an extraordinary turning point in the life of this country. Let ...	b'Tonight we have left the EU - an extraordinary turning point in the life of this country. Let ...	b Tonight we have left the EU an extraordinary turning point in the life of this country Let us ...	[b, tonight, we, have, left, the, eu, an, extraordinary, turning, point, in, the, life, of, this...	[tonight, left, eu, extraordinary, turning, point, life, country, let, us, come, together]

Source: Own work.

Figure 3: Distribution of tweet polarity

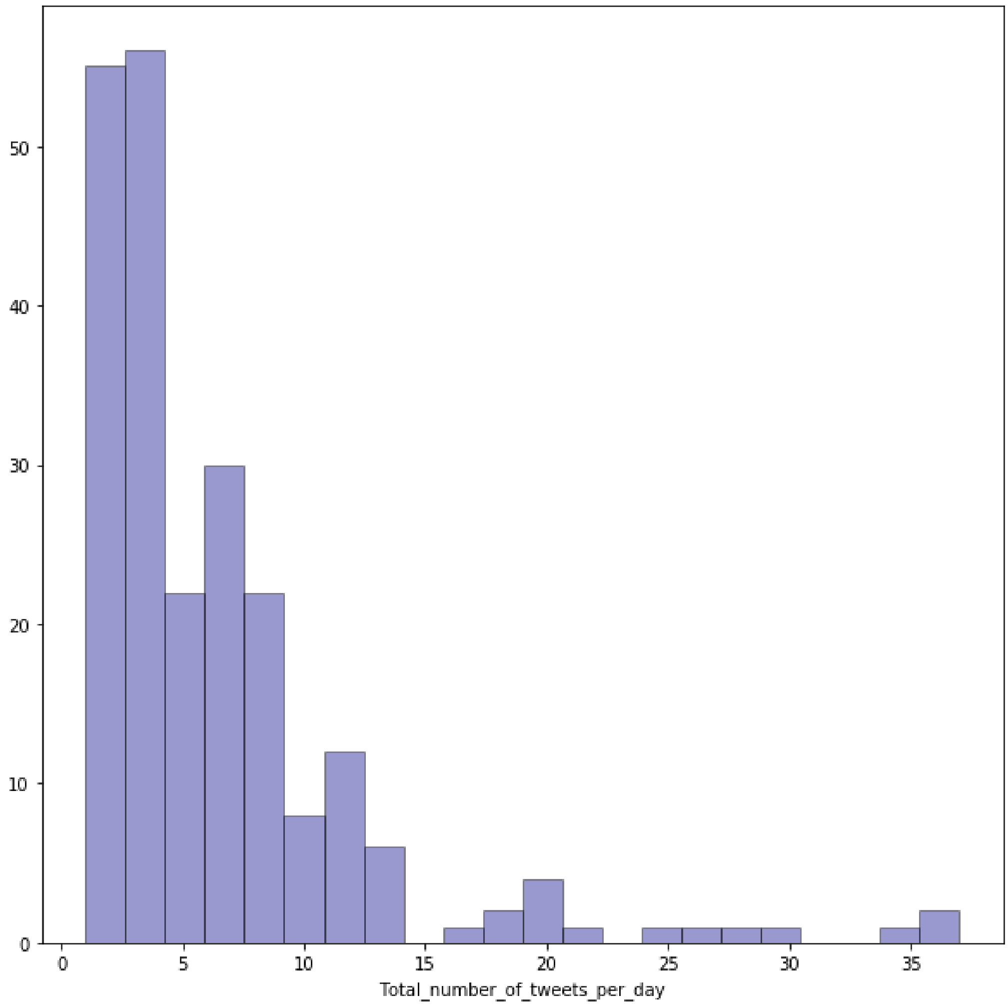


Source: Own work.

The distribution of the tweet polarity is shown on Figure 3. Because of the fact that the plot represents a density function, all of the shown values are without any temporal ordering. It can be observed that for most of the days the tweets have a positive sentiment, with high polarity score of 1 or neutral sentiment with polarity score of 0.

The distribution of the tweet volume is shown on Figure 4. The figure shows that on most days the number of tweets goes from 1 to 10, while there are very few cases where the tweet number goes higher than 15 per day.

Figure 4: Distribution of tweet volume



Source: Own work.

3.2 Pearson Correlation

One of the methods used to investigate the relationship between the tweet polarity time series X_t and the market return time series Y_t is the Pearson correlation. More specifically, the Pearson correlation is used to measure the strength of the linear relationship between the two time series over the whole time period. The correlation coefficient $\rho(X_t, Y_t)$, which quantifies the linear contemporaneous dependence, is defined by Karl Pearson (1920) as:

$$\rho(X_t, Y_t) = \frac{(X_t Y_t) - (X_t)(Y_t)}{\sqrt{((X_t{}^2) - (X_t)^2)((Y_t{}^2) - (Y_t)^2)}} \tag{3}$$

The correlation coefficient can go in an range from -1 to 1. A positive correlation coefficient would mean that for every positive increase in one variable there is also a positive increase in the other variable. A negative correlation coefficient, on the other hand, would mean that for every positive increase in one variable there is also a negative decrease in the other variable. The bigger the absolute value of the correlation coefficient, the stronger the relationship between the variables. A correlation coefficient of 1 points towards a perfect positive relationship between the variables of interest, whereas a correlation coefficient of -1 points towards a perfect negative relationship between the variables.

In order to calculate the correlation coefficient and continue with the further analysis, the two time series need to be of same length. The polarity time series contains data for all of the dates on which Boris Johnson tweeted. For the analyzed period of 8 months, there are only 7 days on which he did not tweet, for which the polarity is set to be 0. The market returns time series, however, miss data for the weekends and holidays. For the purpose of not losing any valuable information about the tweet polarity on the days for which the markets were closed, I decided to average them with the polarity score of the next trading day. In this way, tweet polarity in some cases reflects the average sentiment of two or more days.

Table 3: Pearson correlation between average tweet polarity and market returns

Variables	**Pearson Correlation Coefficient**	**P-value**
FTSE 100 & Tweet Polarity	-0.035379	0.6459
FTSE Euro 100 & Tweet Polarity	-0.010477	0.8914
FTSE All World & Tweet Polarity	-0.046774	0.5387
GBP/EUR & Tweet Polarity	-0.00736	0.9229
GBP/USD & Tweet Polarity	-0.049507	0.5152
EUR/USD & Tweet Polarity	-0.075844	0.3184

Source: Own work.

The Pearson correlation coefficients are shown in Table 3. It can be observed that for all market return time series, the correlation with the polarity time series is very small and negative. This could be due to the fact that the whole time period is considered, as opposed to only the polarity/volume tweet peaks. Similar results are also observed in previous studies, such as those of Mao, Counts and Bollen (2011) and Ranco, Aleksovski, Caldarelli, Grčar and Mozetič (2015).

As a robustness check the analysis is repeated without averaging the tweet polarity. This means that in the second approach the tweet polarity for the non-trading days is completely disregarded. Results are shown in Table 4. As it can be observed the Pearson correlation coefficients in Table 3 and Table 4 do not differ in any significant way. Again the relationship between the market time series and the polarity time series for the whole time period is very small and in most cases negative.

Table 4: Pearson correlation between tweet polarity and market returns

Variables	**Pearson Correlation Coefficient**	**P-value**
FTSE 100 & Tweet Polarity	-0.041349	0.5912
FTSE Euro 100 & Tweet Polarity	0.00297	0.9691
FTSE All World & Tweet Polarity	-0.031199	0.6819
GBP/EUR & Tweet Polarity	0.000294	0.9969
GBP/USD & Tweet Polarity	-0.038186	0.6158
EUR/USD & Tweet Polarity	-0.066717	0.3803

Source: Own work.

3.3 Granger Causality

Apart from the correlation analysis, in order to further test the tweet sentiment's possible utility as a market indicator, a Granger causality test is also applied. The Granger causality test, first introduced by Granger (1969), is used to determine whether the tweet sentiment time series X_t is useful for predicting the market return time series Y_t. It needs to be noted, however, that Granger causality is not by definition same as true causality. Rather than testing if the polarity times series X_t causes the market time series Y_t in the traditional sense, the Granger causality tests if X_t forecasts Y_t (Piškorec, Antulov-Fantulin, Novak, Mozetič, Grnčar, Vodenska & Šmuč, 2014). The test is performed by regressing Y_t on its own past values, as well as those that include past values of X_t. In mathematical terms, this can be expressed as:

$$Y_t = \sum_{i=1}^{m} \beta_i Y_{t-1} + \sum_{i=1}^{m} \delta_i X_{t-1} + \varepsilon_t \qquad (4)$$

An F-test is used to test the null hypothesis set as:

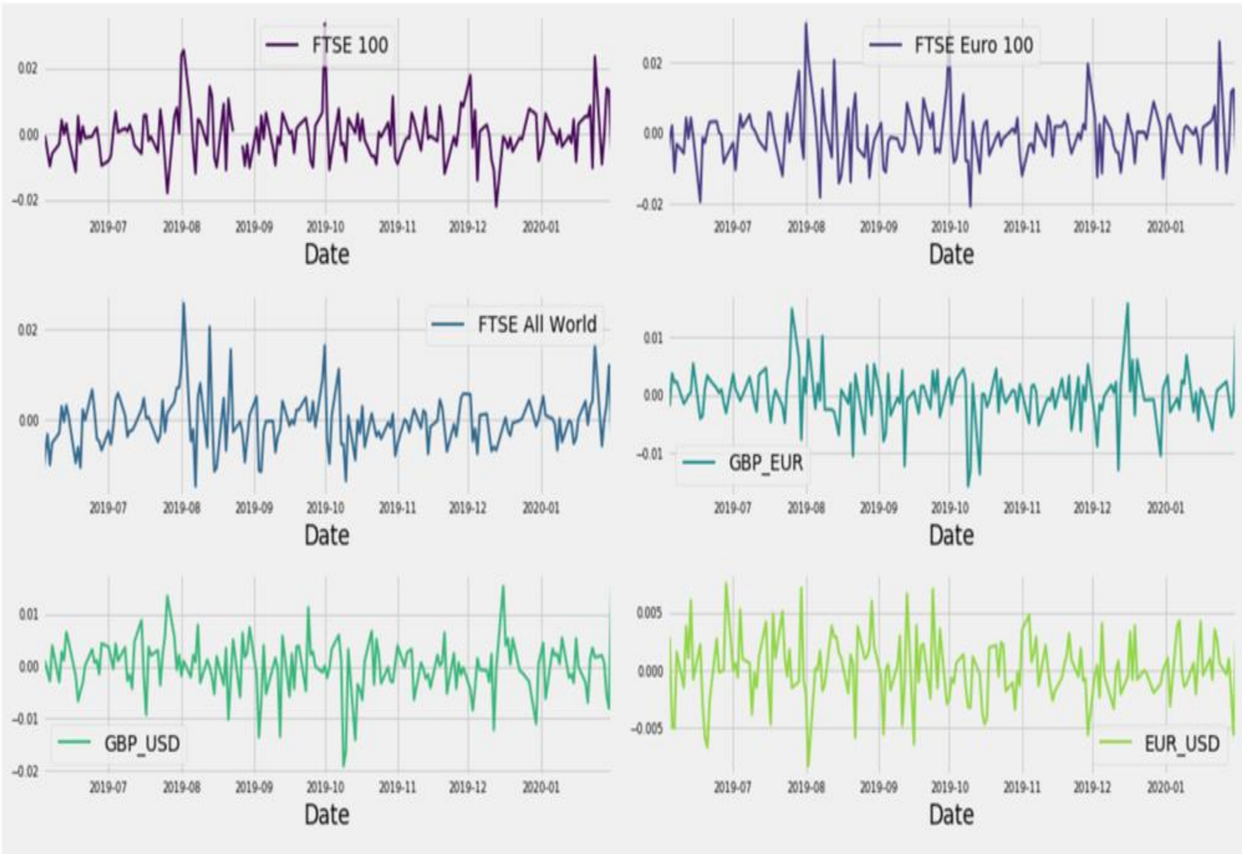
$$H_0: X_t \text{ does not Granger cause } Y_t$$

The idea is to see if the predictive model gets better by the inclusion of the history values of X_t, rather than just those of Y_t. This analysis should provide a deeper insight into the connection between the tweet polarity time series and the market return time series for the entire time period. The Granger causality test requires stationarity of the time series, so the first step is to test for stationarity and consequently make the necessary transformations if the time series prove to be non-stationary.

3.3.1 ADF Test

Stationarity of a given time series implies that its statistical properties do not change over time. More precisely, among other properties, the mean and variance stay constant over time. This is essential requirement for performing a lot of statistical tests, including the Granger causality test.

Figure 5: Market returns time series



Source: Own work.

Plots of the market return time series used in the analysis are given on Figure 5. However, just by observing, it cannot be determined if the displayed time series are stationary or not.

In statistics the Augmented Dickey-Fuller (ADF) test, developed by David Dickey and Wayne Fuller in 1979, is one of the most commonly used unit root tests designed for determining if a time series is stationary. ADF tests the null hypothesis that a unit root is present in an autoregressive model against the alternative hypothesis that the time series is stationary (Mushtaq, 2011). In mathematical terms the model is given by the following expression:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{5}$$

where y_t is the observed time series, α is a constant, β is the time trend coefficient, γ is the coefficient for the first lag of the series, Δy_{t-1} is the first difference of the series at time $t-1$ and p is the order of the autoregressive process (Dickey and Fuller, 1979).

There are many methods that can be used for determining the optimal lag length (p) for the model. One of them is by examining the Akaike information criterion (AIC), formulated by the Japanese statistician Hirotugu Akaike in 1974. This step is important for the analysis since estimating too many parameters might lead to overfitting the model. Overfitting is a problem since it reduces the model's generalizability outside the original dataset. AIC tackles this issue by dealing with the trade off between simplicity and goodness of fit. It does this by rewarding the goodness of fit assessed by the model likelihood function, while including a penalty term for adding additional parameters to the model. The AIC is given by:

$$AIC = 2k + 2\ln(\hat{L}) \tag{6}$$

where k is the number of estimated parameters and $\hat{L}$ is the maximum value of the likelihood function for the model (Akaike, 1974). The smaller the value for AIC, the better the model fit is.

After the right order of the model is determined in this way, the ADF test is carried out for the purpose of testing the null hypothesis set as:

$$H_0 \colon \gamma = 0$$

Under the null hypothesis the time series has a unit root, or in other words it is not stationary. In contrast, when the process has no unit root the time series is stationary. The logic behind this is that since a stationary time series exhibits reversion to the mean, the lag value must provide some relevant information useful for predicting the change in the time series, hence γ must in that case be statistically significantly different from zero.

The test statistics is given by:

$$DF_T = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \tag{7}$$

If the value from the test statistics is less than the critical value for the ADF test, the null hypothesis can be rejected. This would mean that there is no unit root present, or in other words that the time series is stationary.

Table 5: Results of the ADF test for stationarity of the time series

Time Series	**Test Statistics**	**Critical Value 5%**	**Number of lags chosen**
FTSE 100	-8.2734	-2.879	2
FTSE Euro 100	-10.469	-2.879	1
FTSE All World	-3.5996	-2.879	6
GBP/EUR	-13.6684	-2.878	0
GBP/USD	-12.5749	-2.878	0
EUR/USD	-12.5707	-2.878	0

Source: Own work.

The results of the ADF test for all of the time series are presented in Table 5. They include the test statistics value, the critical value for a significance level of 5% (alpha = 0.05), as well as the optimal number of lags chosen by the AIC criterion for each of the time series. It can be observed that the test statistics value is less than the critical value for all instances, meaning that there is sufficient evidence to reject the null hypothesis. This leads to the conclusion that all of the time series are stationary, hence no transformation of the data is needed. However, before continuing with the further analysis another commonly used unit root test will be applied as a robustness check for the results obtained by the ADF test.

3.3.2 KPSS Test

In statistics the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is used for testing if a time series is stationary around a deterministic trend or non-stationary due to the presence of a unit root (Kwiatkowski, Phillips, Schmidt & Shin, 1992). Contrary to the ADF test, the time series is stationary under the null hypothesis. Another important difference with the ADF test is that the KPSS test may not necessarily reject the null hypothesis even if the given time series follows a deterministic trend, meaning that it is steadily increasing or decreasing through time.

The KPSS test is based on the following linear regression model:

$$x_t = r_t + \beta_t + \varepsilon_t \tag{8}$$

where x_t is the observed time series, β_t is the deterministic trend, ε_t is a stationary error and r_t is a random walk given by:

$$r_t = r_{t-1} + u_t \tag{9}$$

where u_t is independently identically distributed with a 0 mean and σ_u^2 varience. The initial value r_0 is considered to be a fixed intercept. Under the null hypothesis $\sigma_u^2 = 0$ (Kwiatkowski, Phillips, Schmidt & Shin, 1992).

Table 6: Results of the KPSS test for stationarity of the time series

Time Series	**Test Statistics**
FTSE 100	0.0770
FTSE Euro 100	0.0832
FTSE All World	0.0834
GBP/EUR	0.1663
GBP/USD	0.1675
EUR/USD	0.0819

Source: Own work.

Table 6 shows the results for the KPSS test. For a significance level of 5% (alpha = 0.05) the critical value is 0.463. If the value of the test-statistics is bigger than the critical value, the null hypothesis can be rejected and hence concluded that the time series is not stationary. Since the test statistic values for all of the times series presented in Table 6 are below 0.463, there is not enough evidence to reject the null hypothesis. This goes in line with the conclusion obtained from the ADF test results that all of the time series are stationary.

3.3.3 Statistical Validation

After making sure that all of the time series are stationary, the right order of the model (m) is determined by using the AIC criterion. Then an F-test is used to test the null hypothesis that the polarity time series $\boldsymbol{X_t}$ does not Granger cause each of the market returns time series $\boldsymbol{Y_t}$.

Table 7: Results of the Granger causality test with average polarity

$\boldsymbol{X_t}$	$\boldsymbol{Y_t}$	**P-value**
Tweet Polarity	FTSE 100	0.0929
Tweet Polarity	FTSE Euro 100	0.0206
Tweet Polarity	FTSE All World	0.0985
Tweet Polarity	GBP/EUR	0.5695
Tweet Polarity	GBP/USD	0.2836
Tweet Polarity	EUR/USD	0.0134

Source: Own work.

The results are shown in Table 7. If the p-values got as a result from the test are less than the significance level of 0.05, the null hypothesis can be rejected meaning that time series $\boldsymbol{X_t}$ Granger causes time series $\boldsymbol{Y_t}$. Since the p-value is less than the significance level in the case of the FTSE Euro 100 Index and the EUR/USD time series, the null hypothesis can be rejected and therefore concluded that the tweet polarity time series is useful in predicting the FTSE Euro 100 and the EUR/USD time series. For the rest of the time series, the null hypothesis cannot be rejected.

Just like with the Pearson correlation analysis, the Granger causality process is repeated when tweet polarity for the non-trading days is not taken into consideration. Results are presented in Table 8. Again the p-value is less than the significance level in the case of the FTSE Euro 100 and the EUR/USD time series. The null hypothesis is rejected in those cases and for the analyzed period of 8 months it can be said that the tweet polarity time series is useful in predicting the FTSE Euro 100 and the EUR/USD time series.

Table 8: Results of the Granger causality test

X_t	Y_t	P-value
Tweet Polarity	FTSE 100	0.0612
Tweet Polarity	FTSE Euro 100	0.0131
Tweet Polarity	FTSE All World	0.0696
Tweet Polarity	GBP/EUR	0.7292
Tweet Polarity	GBP/USD	0.2967
Tweet Polarity	EUR/USD	0.0063

Source: Own work.

Both the Pearson correlation and Granger causality analysis focus on inspecting the relationship between the tweet polarity time series and the market return time series for the whole time period of 8 months. While, this might be a good starting point that gives us some insight about the significance of the analyzed relationship, it is not very informative about what happens on the shorter horizon. The analysis does not consider tweet polarity or tweet volume peaks, but rather looks into the average effect over the whole time period.

The second method used in this study relates to the relationship between the tweet polarity and the index market returns over a shorter period of time. More specifically, the event study focuses on the abnormal market returns observed during some external events. In this case the external events of interest are the Boris Johnson's tweets. The next subsection focuses on describing the event study procedure.

3.4 Event Study

For a long time, measuring the effect of a specific economic event on the value of some stock or index has been among the most common tasks of economists. The event study methodology is a statistical method frequently used by a lot of financial experts for achieving exactly that purpose. For example, Kothari and Warner (2004) find that the number of published event studies exceeds 500, a number that only continued to grow until today. By inspecting the financial market data over a very short period of time, an event study measures the effect of a given event. The aim is to check if unexpected or abnormal returns for the researched market data are present on the event date. If this is indeed the case and the analyzed event provides new information that disrupts the market price in a significant way, a measurement of such effect might lead to the design of better predictive models.

Event studies have been used for a long period of time. One of its earliest forms dates back to James Dolley (1933), who analyze the price effects of stock splits, by examining the nominal price changes at the time of the split. In his study he uses a sample of 95 splits, for the 10 year period between 1921 and 1931, and finds that the stock price increased in 57 of the cases.

Starting from the early 1930s until the late 1060s the event study methodology gradually became more sophisticated. Most of the improvements come from the removal of general stock market price movements and separation of the confounding events. In the late 1960s studies similar to the ones of Ray Ball and Philip Brown (1968) and Fama, Fisher, Jensen, & Roll (1969) have introduced an event study methodology very similar to the one used today. In their paper Fama, Fisher, Jensen & Roll analyse how common stock prices adjust to new information about stock splits by concentrating on the unexpected part only, while Ball and Brown consider the information content of earnings.

In the years that followed further modifications to the event study methodology have been introduced. These modifications deal with the problems arising from violations of the assumptions upon which the statistical models used in earlier research are built, by making certain adjustments in the methodology's design. Two of the most influential papers from that period, which deal with the importance of possible complications and the needed modifications, are the ones of Stephen Brown and Jerold Warner (1980) and Stephen Brown and Mark Weinstein (1985). The first paper published in 1980 concentrates on the implementation issues of monthly data, while the second one published in 1985 focuses on issues regarding daily data.

Nowadays, the event study methodology has found many applications. From being used in economics, finance, accounting, management to information technology, law, political science and many other fields. Usually in economics and finance, an event study is conducted to analyse the effects of dividend payment announcements, stock splits, earning announcements, acquisitions or announcements of macroeconomic indicators, such as GDP rate, inflation and employment (MacKinlay, 1997).

In this master thesis the event study methodology is used to analyze the effect of Boris Johnson’s tweets or more precisely the sentiment extracted from those tweets. This is done by examining the stock market returns in the UK represented by the FTSE 100 Index, as well as the European stock market returns represented by the FTSE Euro 100 Index, over a very short period of time. The short horizon is justified if markets are efficient. Under the EMH, security prices reflect all new information, since rational investors immediately react to them. The goal is to see if the sentiment extracted from Boris Johnson’s tweets provides any significant information to the stock market place. If that is the case, some correlation between the returns of the indices and the tweet polarity is expected.

Although different event studies might be organized in a different way, they generally follow a certain structure that includes the following steps:

1. Definition of the events.
2. Estimation of the normal returns for the period before the events, using different statistical or economic models (ex. the market model).
3. Calculation of the normal returns for the event dates, by using the estimates obtained from the market model for the previous period.
4. Calculation of the abnormal returns for the events, by taking the difference between actual returns and estimated normal returns.
5. Assessing the statistical significance of the abnormal returns on the event date.

The next subsection focuses on the description of the event study methodology. Following the usual practice, the analysis starts with a detection of the events of interest and identification of the period over which the index prices involved in the analysis are examined. This period is called the event window and it is customary to make it longer than the specific date of the event. By doing so, examination of the short period surrounding the event is also taken into account. Typically this period is extended to few days before and after the event date. Evaluation of the event's effect on the marker index requires a measure of the abnormal returns for the event window. The abnormal return is defined as the difference between the actual return of the index over the event window and the normal return of the index over the event window. So, the next step in the analysis is to define a model that will be used for the estimation of normal returns. Typically there are two most common choices for such models: the constant mean model and the market model. Both of them are statistical models and as such rely upon assumptions about the distribution of index returns. As the name implies, the constant mean model is built around the assumption that the mean return of the index is constant through time. On the other hand, the market model assumes a stable linear relationship between the index and market return. Since the normal return is defined as the expected return that is not dependent on the event taking place, estimates of the normal returns need to be determined for the period before the event window. For an example, an event study that uses daily index data and the market model can go back up to 400 days prior to the event in order to obtain the market model parameters. This period is called an estimation window and is defined along with the event window in the subsection below. After the parameter estimates for the normal performance model are obtained, the abnormal returns over the event window are easily calculated. Next follows the aggregation of individual index abnormal returns along with the design of the testing framework for the significance of the cumulated abnormal returns. Finally, the subsection is concluded with a presentation of the empirical results obtained from the process.

3.4.2 Detection of Events

The focus of this event study is to analyze the effect of new information that comes in a form of a polarity score, which captures the tone (sentiment) of the given information. So, the events that are of interest here are the Boris Johnson's tweets or more precisely the sentiment extracted from those tweets.

Table 9: Events

Event Date	P_d	Event Classification
31.01.2020	-1	negative
06.12.2019	0.75	positive
04.12.2019	1	positive
03.12.2019	1	positive
02.12.2019	-1	negative
01.12.2019	-1	negative
29.11.2019	-1	negative
28.11.2019	-0.6	negative
27.11.2019	1	positive
21.11.2019	1	positive
19.11.2019	0.867	positive
17.11.2019	0.6	positive
15.11.2019	0.667	positive
14.11.2019	1	positive
07.11.2019	1	positive
06.11.2019	1	positive
31.10.2019	1	positive
18.10.2019	1	positive
16.10.2019	1	positive
15.10.2019	1	positive
11.10.2019	0.6	positive
10.10.2019	1	positive
30.09.2019	1	positive
29.09.2019	1	positive
19.09.2019	1	positive
06.09.2019	1	positive
26.08.2019	1	positive
19.08.2019	-0.6	negative
13.08.2019	-1	negative
12.08.2019	-1	negative
31.07.2019	1	positive
29.06.2019	1	positive
19.06.2019	0.6	positive
03.06.2019	1	positive

Source: Own work.

However, it does not make sense to consider all of the days for the analyzed period, going from 01 June 2019 to 01 February 2020, as event dates. Only days with a significant tweet polarity, tweet volume or a combination of the two should be considered as important for the analysis. This recognizes the fact that bigger polarity score might be more important than a smaller one. The idea is that a stronger message received by the market participants is more likely to cause a reaction of a larger magnitude and with that a significant change in the market returns. Let's first try to examine only the tweet polarity as a factor for determining if a day should be regarded as an event or not. As it can be observed, the distribution of tweet polarity alone (shown on Figure 3) is not very useful, since most of the tweets have either a polarity of 0 or an extreme one of 1/-1. This would mean that by looking at the tweet polarity alone, almost 80% of the data would be classified as events. So in addition to the tweet polarity, the tweet volume (shown on Figure 4) also needs to be considered as a factor for detecting the event dates. As a result, the rules for event detection are presented as follows. A day is classified as an event if:

- The $P_d > 0.5$ or $P_d < -0.5$ and
- The number of tweets (N) > 7

The total number of such days amounts to 34 dates or 15% of the tweet data. The events are further classified into two groups: positive or negative events, depending on the daily tweet polarity. This is done in the following way:

- If $P_d > 0$ the event is classified as a positive event
- If $P_d < 0$ the event is classified as a negative event

Table 9 above shows the event dates, along with their classification based on the polarity score. It can be observed that most of the event dates are labeled as positive as opposed to those that are labeled as negative. Total number of positive events amounts to 26 dates or 76% of the data, while the total number of negative events amounts to 8 dates or 24% of the data.

3.4.3 Defining the Event Window and Estimation Window

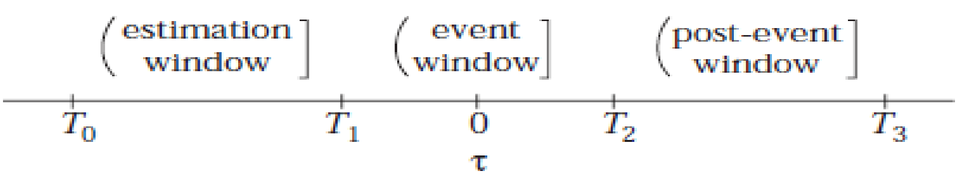
The aim of the event study is to check if unexpected or abnormal returns for the FTSE 100 Index and the FTSE Euro 100 Index are present on the event dates previously defined. However, some information might be anticipated by the market participants even before the event day. In a similar sense, it might take a while before the market participants react to the given information. In order to account for this, index returns before and after the event date are also considered. This period is called an event window and it is usually one or two days before and after the event date. An event window of 1 day prior and 1 day after the event date, as well as an event window of 2 days prior and 2 days after the event date is of interest for this analysis. Considering longer event windows (ex. +/- 5 days) will only increase the possibility of other events taking place at that time, and with that the likelihood of them influencing the change of the index prices.

The normal returns needed for the calculation of abnormal returns are estimated for the period before the event date. That period is called an estimation window and its duration can go from 100 to 400 days before the event date. Longer estimation window means more observations and with that more precise estimates. In this event study, the estimation window is chosen to be 250 days prior the event date. As it is usually the case, the event window does not overlap with the estimation window, so that the estimated normal return parameters are not influenced by the returns around the event. This is done in order to avoid both the normal and abnormal returns to capture the events impact. The whole methodology is built around the assumption that the event impact is captured by the abnormal returns, so it becomes clear why allowing for overlapping event and estimation window would be problematic.

Usually in event studies it is also a common practice to exclude previous events from the estimation window, so they do not in any way influence the estimation of the normal returns. However, since in this case some of the event dates are right next to each other, leading to an overlap between the event windows, excluding previous events might mean a loss of important data. In order to check if there is a significant difference in the results between the two approaches the analysis is repeated twice. First with and then without removing previous event dates from the estimation window. Results are then compared.

Before going any further with the analysis, a certain notation regarding the timeline of the event study needs to be defined first. Denoting $\tau = 0$ as the event date, $\tau = T_1 + 1$ represents the event window and $\tau = T_0 + 1$ to $\tau = T_1$ represents the estimation window. The length of the estimation window will then be defined as $L_1 = T_1 + T_0$ and the length of event window as $L_2 = T_2 - T_1$. Due to the event window length being larger than the day of the event announcement, abnormal returns around the event day will also be facilitated. Let the post event window be from $\tau = T_2 + 1$ to $\tau = T_3$ (MacKinley, 1997). Following this notation, the time line is constructed and illustrated on Figure 6.

Figure 6: Time line for the event study



Source:MacKinlay (1997).

So, as mentioned above, in order to calculate the abnormal returns for the event window, first normal returns need to be estimated for the estimation window. The following subsection gives a short overview of the most commonly used methods for the estimation of normal returns, while focusing on the model chosen for the purpose of this analysis.

3.4.4 Estimation of Normal Returns

Various models used for the estimation of normal returns exist. They can be largely classified into two groups: statistical and economic models. The statistical models rely upon statistical assumptions about the financial market returns. On the other hand, the economic models are based upon assumptions about investor's behavior. It needs to be noted, however, that while the economic models do not depend solely on statistical assumptions, they also need to be included when using one of this methods in practice (MacKinley, 1997).

There are two main statistical models most commonly used in practice: the constant mean return model and the market model. Both of them impose an assumption that the market returns are jointly multivariate normally, independently and identically distributed (i.i.d.). Although this assumption in its nature is very restrictive, in reality is has proven to be empirically reasonable, so it does not often lead to biases causing problems for the analysis (MacKinley, 1997).

3.4.4.1 Constant Mean Return Model

Out of the two statistical models the constant mean return is the simpler one. It is given by the following equation:

$$R_{it} = \mu_i + \vartheta_{it} \tag{10}$$

where μ_i is the mean return for asset i, R_{it} is the return for asset i at time t and ϑ_{it} is the disturbance term for asset i at time t, that has zero mean and variance $\sigma_{\vartheta_{it}}$.

Regardless of its simplicity, some researchers have shown that the constant mean model often produces similar results to the ones of more complex models (Brown & Warner, 1980; Brown & Weinstein, 1985).

3.4.4.2 Market Model

The other frequently used statistical model for the purpose of estimating the normal returns is the market model. As such it relies upon assumptions about the distribution of market returns. Its linear specification comes from the assumption that the market returns are jointly multivariate normal, independently and identically distributed (MacKinley, 1997).

The market model connects the return of an asset (index) to the return of the market portfolio. For each asset i, the normal return R_{it} is given by:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \tag{11}$$

where

$$E(\varepsilon_{it}) = 0, \qquad var(\varepsilon_{it}) = \sigma^2_{\varepsilon i}$$

and R_{mt} is the return of the market portfolio, α_i, β_i and $\sigma^2_{\varepsilon i}$ are the parameters of the model and ε_{it} is the disturbance term with zero mean (MacKinley, 1997).

Under the general assumptions and conditions the ordinary least squares (OLS) is a consistent and efficient estimator for the parameters of the market model. For the íth market index in event time the OLS estimators of the market model parameters, for the observations in the estimation window, are calculated as follows:

$$\hat{\beta}_i = \frac{\sum_{\tau = T0+1}^{T_1}(R_{i\tau} - \hat{\mu}_i)(R_{m\tau} - \hat{\mu}_m)}{\sum_{\tau = T0+1}^{T_1}(R_{m\tau} - \hat{\mu}_m{}^2)} \tag{12}$$

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\beta}_i \hat{\mu}_m \tag{13}$$

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{L_1 - 2} \sum_{\tau = T0+1}^{T_1} (R_{i\tau} - \hat{\alpha}_i - \hat{\beta}_i R_{m\tau}) \tag{14}$$

Where

$$\hat{\mu}_i = \frac{1}{L_1} \sum_{\tau = T0+1}^{T_1} R_{i\tau} \tag{15}$$

$$\hat{\mu}_m = \frac{1}{L_1} \sum_{\tau = T0+1}^{T_1} R_{m\tau} \tag{16}$$

and $R_{i\tau}$ represents the return of index í in the event period τ and $R_{m\tau}$ are market returns in the event period τ (MacKinley, 1997). The market return is given by the FTSE All World Index.

It is important to point out that when normal returns are estimated with the market model, a fragment of the return related to the market's return variance is explained, which consequently leads to a reduction of the variance in the abnormal return. If this reduction is sufficient enough, it increases the ability of the model to detect the event effect. Here lies the potential improvement of the market model over the constant mean return model.

Given its potential advantage, for the purpose of this analysis the marker model will be used for estimating the normal returns. Following Equation 11, the market index returns are regressed on the returns of the global market portfolio represented by FTSE All World Index. Estimates for the regression parameters are obtained in the process. After this is done, normal expected returns of the market indices can be easily calculated for the event window. The next subsection focuses on describing the process for the calculation of abnormal returns.

3.4.5 Abnormal Returns

The abnormal returns in the event window are given by the difference between the actual index returns and the expected normal index returns, whose parameters are estimated with the market model previously described. Therefore, the daily abnormal return for the index i is calculated as:

$$\widehat{AR}_{it} = R_{it} - \hat{\alpha} - \widehat{\beta_i} * R_{mt} \tag{17}$$

Under the null hypothesis, the distribution of the sample abnormal return of a given observation in the event window is normal with mean zero and variance σ^2 (MacKinley, 1997).

$$\widehat{AR}_{it} \sim N(0, \sigma^2(\widehat{AR}_{it}))$$

3.4.6 Statistical Validation

In the following subsegment, the aggregation of abnormal returns and testing for statistical significance is considered. In order to investigate the effect of the events, the abnormal return observations need to be aggregated. The aggregation is done through time by summing up the abnormal returns over all days in the event window. Let's denote $\widehat{\mathrm{CAR}}_{i(t1,t2)}$ as the cumulative abnormal returns, given by:

$$\widehat{CAR}_{i(t1,t2)} = \sum_{t=t1}^{t2} \widehat{AR}_{it} \tag{18}$$

The distribution of CAR under the null is also normal with mean zero and variance σ^2 (MacKinley, 1997).

$$\widehat{CAR}_{(t1,t2)} \sim N(0, {\sigma_i}^2{}_{(t1,t2)})$$

The null hypothesis is set as:

$$H_0\text{: AR} = 0$$

In other words, under the null events have no significant impact on the stock returns, or it can be said that the abnormal returns are not statistically significantly different from zero. In order to test the null hypothesis, the t-test is applied. It is important to note that the t-test is a parametric test and as such it assumes that abnormal returns are normally distributed. The test statistics is given by:

$$\theta_1 = \frac{\widehat{CAR}_{(t1,t2)}}{var(\widehat{CAR}_{(t1,t2)})^{1/2}} \tag{19}$$

where

$$var(\widehat{CAR}_{(t1,t2)}) = \sum_{t=t1}^{t2} var(\widehat{AR}_{it}) \tag{20}$$

In practice $\sigma^2_{\varepsilon_i}$ is not known, so an estimator must be used in order to calculate the variance of abnormal returns. The variance is given by:

$$var\left(\widehat{AR}_{it}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2_{\varepsilon_i} \tag{21}$$

where N is the number of events (MacKinley, 1997).

Some researches, however, point out to the limitation that comes from the assumption about the normal distribution of returns. Fama (1976), for example, has shown that daily returns depart from normality more often than monthly returns. The evidence he obtained from his research suggests that relative to a normal distribution, the distribution of daily returns has fatter tails. Moreover, according to Brown and Warner (1985) the same holds true for daily excess returns. This alone brings into question the validity of the results obtained by the parametric t-test.

Unlike the parametric tests, the nonparametric tests do not impose any assumptions about the distribution of abnormal returns (Cowan, 1992). In that regard, it can be said that the nonparametric tests are less restrictive. Daniel (1990) has come to the conclusion that nonparametric tests are less likely to be improperly used, because minimal assumptions about the data are required. Moreover, the computations are simple and can be performed rather quickly. Therefore, researchers with limited knowledge in mathematics and statistics might find them easier to understand. Similarly, Gibbons (1976) argues that the easiest way to abuse any statistical technique is to not fully understand the assumptions neccesary for the validity of the tests. That is why the nonparametric tests might come of use in cases where such assumptions are brought into question.

However, although nonparametric tests have the reputation of having simple calculations, Daniel (1990) has shown that sometimes the arithmetic that goes in the process is tedious and requires high computing power, especially when the sample size is large. Similarly, Siegel (1956) has stated that nonparametric statistical tests throw away information, if all of the assumptions required by the parametric tests are met in the data. Hence, for the purpose of not being wasteful of important data, the nonparametric tests are not typically used in isolation, but jointly with the parametric tests. They can serve as a robustness check for the conclusions based upon the results from parametric tests.

One of the most commonly used nonparametric tests in event studies is the rank test. The results obtained from a simulation experiment in the study of Corrado (1989) and Corrado and Zivney (1992) suggest that the sign test delivers reliable results in event studies. Moreover, they have additionally proven that their version of the sign test is more reliable than the standardly used parametric t-test. The basic idea of the sign test is that under the null hypothesis of no significant market reaction to the events, it is equally possible that the abnormal returns will be positive or negative at a given time.

Another commonly used nonparametric test in event studies is the Wilcoxon signed-rank test, proposed by Frank Wilcoxon in 1945. Just like the other nonparametric tests, the Wilcoxon signed-rank test is applied for data where the robustness check against non-normal distribution is desirable. Aside from the sign, it also takes into account the magnitude of the abnormal returns. In other words, it acknowledges the fact that a bigger abnormal return might be more important than a smaller one (D'elia, 2005). In order to perform the Wilcoxon signed-rank test, first each of the abnormal returns needs to be transformed in its absolute value. Then, the obtained values are ranked going from the lowest to the highest. Finally the sign removed from the abnormal returns, by the transformation in absolute value, is again assigned to each rank. The idea of the Wilcoxon signed-rank test is that under the null hypothesis, the sum of the ranked abnormal returns above and below the medium should be similar.

The statistic is defined as:

$$T^* = \sum_{i=1}^{N} r_i^* \tag{22}$$

where N is the number of observations and r_i^* is the positive or negative rank of the absolute value of abnormal returns on a day of the event window (Luoma, 2011).

In this analysis, for the reasons previously discussed the nonparametric Wilcoxon signed-rank test will be used jointly with the parametric t-test and serve as a robustness check for the results.

Results from the event study analysis are presented in Table 10 and Table 11. As previously described a day is classified as an event if the daily tweet polarity is above 0.5 or below -0.5 and the number of tweets on that day is bigger than 7. This gives a total number of 34 event observations (N). The events are further classified into two groups: positive and negative events, depending on the tweet polarity on the given day. If the polarity is above zero, the event is categorized as a positive event and if the polarity is below zero the event is categorized as a negative event. The number of positive events is 26, while the number of negative events is 8. The presented results are organized in a way so that the average effect of all event observations is first displayed, followed by the effect of positive and negative events separately.

Returns of two different market indices are of interest for this analysis: FTSE 100 index and FTSE Euro 100 index. The FTSE 100 index represents the UK stock market, while the FTSE Euro 100 index represents the European stock market. Results for both of them are presented separately in the tables. The market model is used for estimation of the normal returns coefficients over the estimation window that goes 250 days prior to the event date. FTSE All World index is used as a proxy for the market portfolio. Table 10 shows the results of the event study without any modifications to the estimation windows, while Table 11 shows the results when previous event windows are taken from the estimation window.

Abnormal returns for the event windows are calculated as the difference between actual and estimated index returns. Results for the analysis with two different event window lengths are presented in both of the tables. First an event window of 1 day prior and 1 day after the event date (-1, 1) is considered, followed by an event window of 2 days prior and 2 days after the event date (-2, 2). The daily abnormal returns in each event window are then cumulated to get the cumulative abnormal return (CAR), which is later averaged across all 34 events to show the average cumulative abnormal return (mean).

Table 10: Results without taking previous event windows

	FTSE 100			FTSE Euro 100		
	(N=34)			(N=34)		
	Mean	**T-test**	**Signed-rank test**	**Mean**	**T-test**	**Signed-rank test**
CAR (-1,1)						
All events	0.0053	3.9827*	(0.0011)*	0.0012	1.3539	(0.4070)
		(0.0004)			(0.1849)	
Positive events	0.0049	3.03505*	(0.0092)*	0.0015	1.4918	(0.3282)
		(0.0055)			(0.1483)	
Negative events	0.0063	3.1199*	(0.0499)*	0.0004	0.2012	(1.0)
		(0.01684)			(0.8461)	
CAR (-2,2)						
All events	0.0086	5.2542*	(<8.12e-05)*	0.0029	2.3820	(0.0593)
		(<8.71e-06)			(0.0531)	
Positive events	0.0076	3.6987*	(0.0026)*	0.0026	1.8212	(0.1243)
		(0.0010)			(0.08055)	
Negative events	0.0117	7.8345*	(0.0117)*	0.0036	1.6960	(0.1234)
		(0.0001)			(0.1336)	

Source: Own work.

Table 11: Results with taking previous event windows

	FTSE 100			FTSE Euro 100		
	(N=34)			(N=34)		
	Mean	**T-test**	**Signed-rank test**	**Mean**	**T-test**	**Signed-rank test**
CAR (-1,1)						
All events	0.0061	4.7064*	(0.0002)*	0.0015	1.5859	(0.3172)
		(<4.37e-05)			(0.1223)	
Positive events	0.0055	3.5331*	(0.0030)*	0.0019	1.9224	(0.1013)
		(0.0016)			(0.0660)	
Negative events	0.0064	3.1748*	(0.0928)	0.0005	0.2220	(1.0)
		(0.0156)			(0.8306)	
CAR (-2,2)						
All events	0.0109	6.8563*	(<5.20e-06)*	0.0035	2.9883*	(0.0111)*
		(<7.94e-08)			(0.0052)	
Positive events	0.0091	4.6412*	(0.0004)*	0.0029	2.1292*	(0.0655)
		(<9.44e-05)			(0.0432)	
Negative events	0.0118	7.8371*	(0.0117)*	0.0036	1.6777	(0.1614)
		(0.0001)			(0.1372)	

Source: Own work.

Parametric t-test and non-parametric Wilcoxon signed-rank test are both conducted in order to test the null hypothesis, under which abnormal returns are not statistically significantly different from zero. T-statistics of the t-test along with p-values of both the t-test and the Wilcoxon signed-rank test are presented in the tables. Parentheses are used to distinguish the p-values from the t-statistics of the tests. The p-value is interpreted in the context of a significance level, which in this case is 5% ($\alpha = 0.05$). If the p-value is above the significance level, there is not enough evidence to reject the null hypothesis. This would mean that the events have no significant impact on the index returns. The symbol "*" indicates significant values of the tests.

Table 10 shows statistically significant results for the FTSE 100 index for both of the event window lengths. At a significance level of 5%, both the parametric t-test and the nonparametric Wilcoxon signed-rank test yield p-values less than 0.05. This means that there is enough evidence to reject the null hypothesis and therefore conclude that the events have a significant impact on the FTSE 100 index returns. This holds true when inspecting the effect of all events together, as well as when inspecting the effect of positive and negative events separately. For the FTSE Euro 100 index, however, the p-values from both tests fail to reject the null hypothesis, which leads to the conclusion that the events do not have a significant impact on the FTSE Euro 100 index returns.

Table 11 shows very similar results. It can be again observed that for the FTSE 100 index p-values for both the parametric t-test and the nonparametric Wilcoxon signed-rank test point toward significant results for both event window lengths. The only significant t-test result that is not supported by the nonparametric test is the one for negative events when the event window goes from one day prior the event date to one day after the event date. For the FTSE Euro 100 index, however, results seem to slightly differ for the longer event window length. Here, a significant result is obtained for all of the events together, as well as for the positive events, meaning that abnormal returns statistically significantly different from zero were detected around those events. However, the significant result for the positive events obtained by the t-test is not confirmed by the nonparametric Wilcoxon signed-rank test.

As it can be seen from the above, results presented in Table 10 and Table 11 are very similar to each other, meaning that taking out previous events from the estimation window did not lead to significant changes in the results. This might be due to the fact that a longer period is considered for the estimation window, so that previous events do not affect the normal return estimates as much.

4 RESULTS DISCUSSION

Two main approaches were used in this master thesis for the purpose of testing the relationship between the tweet sentiment and the financial markets in the UK. More specifically, I was analyzing the effect of Boris Johnson’s tweets on the stock market, represented by the FTSE 100 Index and the exchange rate market represented by two exchange rates: GBP/EUR and GBP/USD. As a robustness check, I also considered the effect of the tweet sentiment on the European stock market, represented by FTSE 100 Euro Index and the world market, represented by the FTSE All World Index, as well as the exchange rate EUR/USD. The first approach related to relationship for the whole time period of 8 months, while the second concentrated on a small time frame surrounding the events which might have had the biggest impact on the market participants, due to their high frequency and strength.

In order to look at the effect over the whole time period analyzed, I applied the Pearson correlation and Granger causality tests. More specifically, the Pearson correlation was used to measure the strength of the linear relationship between the analyzed variables, while the Granger causality test was used to determine whether the tweet sentiment was useful for predicting the market data. The results obtained from the correlation analysis pointed towards a very small and negative correlation between the tweet polarity and market returns time series. The Granger causality analysis showed that the tweet polarity time series is useful for predicting the FTSE Euro 100 and the EUR/USD time series, but not for the rest of the market returns time series. Both the Pearson correlation and Granger causality analysis focused on inspecting the relationship for the whole time period of 8 months. Although this gave us some insight about the significance of the analyzed relationship, it was not very informative about what was happening on the shorter horizon. The analysis did not consider the strength of tweet polarity or tweet volume peaks, but rather looked into the average effect over the whole time period.

The second method used in this study addressed the relationship between the tweet polarity and the index market returns over the shorter horizon. More specifically, the event study focused on the abnormal market returns observed during external events. In this case the external events were the Boris Johnson's tweets for the period of 8 months, with polarity score above 0.5 or below -0.5 and frequency above 7. The results obtained from the analysis lead towards a conclusion that the events have a significant impact on the FTSE 100 index returns. This held true for both event window lengths of (-2,2) and (-1,1), when inspecting the effect of all events together, as well as when inspecting the effect of positive and negative events separately. For the FTSE Euro 100 Index, however, the events proved to have an insignificant impact. This served as a robustness check against the possibility that the UK abnormal stock returns were caused by some bigger events that impacted the whole European stock market. When looking at the effect of the positive and negative events separately, however, the interpretation of the results gets contra-intuitive for the negative events. Namely, for both positive and negative events, we have a significant impact resulting into positive abnormal returns for the FTSE 100 index returns. This might be due to several reasons:

1. The effect of the events might be more short-lived. This would mean that the participants reacted more quickly to the new information included in the tweets and as a consequence its effect was visible the same day as the event.

2. The number of negative events considered in this analysis (8 negative events) might have been too small, thus leading to biased results.

3. The sentiment classification method used in this analysis for the purpose of determining the tweet polarity (the LM dictionary-based approach) might have incorrectly classified the sentiment of the tweets. Another approach would be to use the machine learning method.

4. Market participants might not have taken the tweets as new information.

5. Some other events might have happened at the same time as the negative tweets impacting the market participant's decision, and thus leading to the significant positive abnormal returns for the analyzed event window lengths.

Table 12: Results with event window (0,0)

	FTSE 100			**FTSE Euro 100**		
	(N=34)			(N=34)		
	Mean	**T-test**	**Signed-rank test**	**Mean**	**T-test**	**Signed-rank test**
CAR (0,0)						
All events	0.0017	2.4443	(0.0125)*	0.0003	0.4665	(0.3930)
		(0.0600)			(0.6438)	
Positive events	0.0014	1.8163	(0.0495)*	0.0001	0.1894	(0.4077)
		(0.0813)			(0.8512)	
Negative events	0.0028	1.6299	(0.1762)	0.0010	0.5432	(0.8657)
		(0.1471)			(0.6037)	

Source: Own work.

In order to address the first possible issue that might have lead to the non-intuitive results for the negative events, I repeated the analysis while narrowing the length of the event window solely to the day of the event (0, 0). The results obtained from the analysis are presented at Table 12. From the table it can be observed that when looking only at the day of the tweet event, we do not get any significant results for either of the indices, while the sign of the effect remains the same. More precisely, both positive and negative events lead to insignificant increase in the abnormal returns for the FTSE 100 index and FTSE Euro 100 index. This leads to the conclusion that the tweet effect is not reflected in the returns the same day and there might have been some other reason behind the positive effect for the negative events.

In order to address the second possible issue, I repeated the analysis for the event window length (-1, 1) while considering all tweets that had non-neutral sentiment, regardless of the frequency or strength of the signal. The total amount of such tweets is 157, 111 of them are positive and 46 negative. The results are presented at Table 13. As it can be observed the sign of abnormal returns for the FTSE Euro 100 Index is negative for the negative events, leading to the intuitive conclusion that the negative events lead to negative abnormal returns. However this time, the effect of both positive and negative events is insignificant. The insignificant results might be due to the fact that all dates were classified as events even when the tweet polarity was very close to 0, or not sufficient enough.

Table 13: Results with increased number of events

	FTSE 100			**FTSE Euro 100**		
	(N=157)			(N=157)		
	Mean	**T-test**	**Signed-rank test**	**Mean**	**T-test**	**Signed-rank test**
CAR (-1,1)						
All events	0.0007	1.1078	(0.2316)	0.0004	1.0787	(0.3584)
		(0.2696)			(0.2823)	
Positive events	0.0004	0.0495	(0.5069)	0.0008	1.6265	(0.1025)
		(0.6215)			(0.1066)	
Negative events	0.0016	1.2087	(0.2490)	-0.0003	-0.4185	(0.4675)
		(0.2330)			(0.6775)	

Source: Own work.

CONCLUSION

In this master thesis I investigated the relationship between the public sentiment derived from Boris Johnson's tweets and the financial market movements. The results have statistically shown that when looking at the relationship over the shorter horizon, positive tweet events seem to be having a significant positive impact on the market stock returns. On the other hand, however, negative tweet events were not followed by negative abnormal returns. The abnormal returns appeared 1-2 days around the events. The relationship on the longer horizon, however, proved to be small and insignificant.

Considering this and other studies with similar outcomes, it becomes very clear how important the role of sentiment caused by social media has become in the search for the creation of more accurate forecasting models.

Nevertheless, it is important to note that there are some things not considered in this research, which can be improved upon in the future work. First, while this thesis offered a more unique approach in terms of determining the public sentiment, by focusing on the Twitter feed of one influential person instead of deriving a public sentiment measure from a large Twitter feed data, it simultaneously limited the amount of analyzed data possibly leading to less accurate results. Extending the number of analyzed correlated Twitter profiles, as well as the time period can be some possible directions for future research. Second, for the purpose of analysing the tweet sentiment the dictionary based approach was used, because of its simplicity and convenience. However, this method does not take into account the context of the text, which can lead to a loss of important information. More advance sentiment analysis methods, like the machine learning method for example, can be used in future studies in order to improve the precision of the results. Finally, the models used in this thesis assume a linear relationship between the variables, which can be quite restrictive for the financial market movements. Other, more sophisticated methods, which capture the non-linear relationship of the analyzed variables, can be used instead in the future.

REFERENCE LIST

1. Akaike, H.(1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/tac.1974.1100705
2. Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross section of stock returns. *The Journal of Finance*, *61*(4), 1645-1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x
3. Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, (2), 129–151. https://doi.org/10.2139/ssrn.962706
4. Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, *6*(2), 159–78. https://doi.org/10.2307/2490232
5. Blasco, N., Corredor, P., & Ferreruela, S. (2011). Market sentiment: A key factor of investors' imitative behavior. *Accounting and Finance*, *52*(3), 663-685. https://doi.org/10.1111/j.1467-629X.2011.00412.x
6. Bollen, J., Mao, H., & Zeng, X. (2010). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, *2*(1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007
7. Bormann, S. (2013). Sentiment indices on financial markets: What do they measure? *Economics Discussion Papers*. Retrieved January 15, 2021 from https://www.researchgate.net/publication/273450297_Sentiment_Indices_on_Financial_Markets_What_Do_they_Measure
8. Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance, 11* (1), 1–27. https://doi.org/10.1016/j.jempfin.2002.12.001

9. Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *The Journal of Business, 78*(2), 405–440. https://doi.org/10.1086/427633
10. Brown, S. J., & Warner, J. B. (1980). Measuring Security Price Performance. *Journal of Financial Economics, 8*(3), 205–258. https://doi.org/10.1016/0304-405X(80)90002-1
11. Brown, S. J., & Weinstein, M. I. (1985). Derived Factors in Event Studies. *Journal of Financial Economics*, *14*(3), 491–495. https://doi.org/10.1016/0304-405X(85)90010-8
12. Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, *18*, 373-399. https://doi.org/10.1016/0304-405x(87)90045-6
13. Campbell, J. Y., & Shiller. R. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies, 1*, 195-228. https://doi.org/10.2307/2961997
14. Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics, 70*(2), 223-260. https://doi.org/10.1016/S0304-405X(03)00146-6
15. Corrado, C. J. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics, Elsevier, 23*(2), 385-395. https://doi.org/10.1016/0304-405X(89)90064-0
16. Corrado, C. J., & Zivney, T. L. (1992). The specification and power of the sign test in event study hypothesis test using daily stock returns. *Journal of Financial and Quantitative Analysis*, *27*, 465–478. https://doi.org/10.2307/2331331
17. Cowan, A. R. (1992). Nonparametric event study tests. *Review of Quantitative Finance and Accounting*, *2*, 343–358. https://doi.org/10.1007/bf00939016
18. Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies, 28*(1), 1-32. https://doi.org/10.1093/rfs/hhu072
19. Daniel, W. W. (1990). *Applied Nonparametric Statistics*. Boston: PWS-Kent.
20. Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science, 53,* 1375-1388. https://doi.org/10.1287/mnsc.1070.0704
21. De Long, J. B., Shleifer, A., Summers, L. H. & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy, 98*(4), 703–738. https://doi.org/10.1086/261703
22. D'elia, V.V. (2005). Measuring the impact of environmental news in Argentina: A non-parametric approach. *Universidad del CEMA*. Retrieved September 23, 2020 from https://acp.copernicus.org/preprints/preprints.html
23. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association, 74*(366), 427–431. https:// doi.org/10.2307/2286348
24. Dolley, J. C. (1933). Characteristics and Procedure of Common Stock Split-Ups. *Harvard Business Review,* 316–26. https:// doi.org/10.12691/jfe-8-2-4

25. Fama, E. F., Fisher, L., Jensen, M., & Roll, R. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review, 10*(1), 1–21. https://doi.org/10.2139/SSRN.321524
26. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417. https://doi.org/10.2307/2325486
27. Fama, E. F. (1976). Foundations of Finance. *The Journal of Finance, 32(3), 961-964 Inc*. https://doi.org/10.2307/2326337
28. Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics, 5*(2), 115–146. https://doi.org/10.1016/0304-405x(77)90014-9
29. Fama, E. F., & French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy, 96*, 246-273. https://doi.org/10.1086/261535
30. Fama, E.F., & French, K.R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, 33*(1), 3-55. https://doi.org/10.1016/0304-405x(93)90023-5
31. Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82–89. https://doi.org/10.1145/2436256.2436274
32. Ferguson, N., Philip, D., Lam, H. Y. T., & Guo, J. M. (2015). Media Content and Stock Returns: The Predictive Power of Press. *Multinational Finance Journal, 19*(1), 1-75. https://doi.org/10.17578/19-1-1
33. Finter, P., Niessen-Ruenzi, A., & Ruenzi, S. (2012). The impact of investor sentiment on the german stock market. *Zeitschrift f ̈ur Betriebswirtschaft, 82*(2), 133–163. https://doi.org/10.1007/s11573-011-0536-x
34. Garcia, D. (2013). Sentiment During Recessions. *Journal of Finance, 68*, 1267–300. https://doi.org/10.1111/jofi.12027
35. Gibbons, J. D. (1976). Nonparametric Methods for Quantitative Analysis. *University of Michigan: Holt, Rinehart and Winston, 53*(1), 391-392. https://doi.org/10.2307/2533133
36. Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica, 37*(3), 424-438. https://doi.org/10.2307/1912791
37. Jackson, A. B., Dinh, T., & Coulton, J. J. (2016). The impact of sentiment on price discovery. *Accounting and Finance, 56*, 669-694. https://doi.org/10.1111/acfi.12128
38. Jegadeesh, N., & Wu, D. (2013). Word Power: A New Approach for Content Analysis. *Journal of Financial Economics, 110*(3), 712–29. https://doi.org/10.1016/j.jfineco.2013.08.018
39. Johnman, M., Vanstone, B. J., & Gepp, A. (2018). Predicting FTSE 100 returns and volatility using sentiment analysis. *Accounting and Finance, 58*(1), 253-274. https://doi.org/10.1111/acfi.12373
40. Kearney, C., & Liu, S. (2014). Textual Sentiment in Finance: A Survey of Methods and Models. *International Review of Financial Analysis, 33*, 171-185. https://doi.org/10.1016/j.irfa.2014.02.006

41. Keim, D. B., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics, 17*(2), 357–390. https://doi.org/10.1016/0304-405x(86)90070-x
42. Kothari, S. P., & Warner, J. B. (2004). The Econometrics of Event Studies. *Handbook of Corporate Finance: Empirical Corporate Finance,* 1. https://doi.org/10.2139/ssrn.608601
43. Kumar, A., & Lee, C. M. C. (2006). Retail Investor Sentiment and Return Comovements. *The Journal of Finance, 61*(5), 2451–2486. https://doi.org/10.1111/j.1540-6261.2006.01063.x
44. Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54,* 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y
45. Loughran, T., & Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance, 66,* 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x
46. Loughran, T., & Mcdonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research, 54*(4), 1187-1230. https://doi.org/10.1111/1475-679X.12123
47. MacKinlay, A. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature, 35*(1), 13-39. Retrieved September 18, 2020 from http://www.jstor.org/stable/2729691
48. Mao, H., Counts, S., & Bollen, J. (2011). *Predicting financial markets: Comparing survey, news, twitter and search engine data*. Retrieved September 23, 2020 from http://arxiv.org/abs/1112.1051
49. Mushtaq, R. (2011). Augmented Dickey Fuller Test. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1911068
50. Lo, A. W., & MacKinlay, A. C. (1988). Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. *Review of Financial Studies, 1*(1), 41–66. https://doi.org/10.1093/rfs/1.1.41
51. Luoma, T. (2011). *Nonparametric Event Study Tests for Testing Cumulative Abnormal Returns*. Retrieved December 23, 2020 from https://www.researchgate.net/publication/226896915_NonParametric_Event_Study_Tests
52. Li, F. (2010a). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature, 29,* 143–165. Retrieved January 24, 2020 from https://www.semanticscholar.org/paper/Textual-Analysis-of-Corporate-Disclosures%3A-A-Survey-Li/b9778f0f90e9cc294a8b3747007ebe82eeec329d
53. Li, F. (2010b). The Information Content of Forward-Looking Statements in Corporate Filings—A Naive Bayesian Machine Learning Approach. *Journal of Accounting Research, 48,* 1049–1102. https://doi.org/10.1111/j.1475-679X.2010.00382.x

54. Olaniyan, R., Stamate, D., Ouarbya, L., & Logofatu, D. (2015). Sentiment and stock market volatility predictive modelling — A hybrid approach. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 1-10. https://doi.org/10.1109/DSAA.2015.7344855
55. Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 1345-1350. https://doi.org/10.1109/SCOPES.2016.7955659
56. Papaioannou, P., Russo, L., Papaioannou, G., & Siettos, C. I. (2013). Can social microblogging be used to forecast intraday exchange rates? *NETNOMICS: Economic Research and Electronic Networking, 14*, 47-68. https://doi.org/10.1007/s11066-013-9079-3
57. Pearson, K. (1920). Notes On The History Of Correlation. *Biometrika*, *13*(1), 25–45. https://doi.org/10.1093/biomet/13.1.25
58. Piškorec, M., Antulov-Fantulin, N., Novak, P. K., Mozetič, I., Grnčar, N., Vodenska, I., & Šmuč, T. (2014). Cohesiveness in Financial News and its Relation to Market Volatility. *Scientific Reorts 4***,** 5038. https://doi.org/10.1038/srep05038
59. Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PloS one*, *10*(9). https://doi.org/10.1371/journal.pone.0138441
60. Samuelson, P. A. (1965). Rational Theory of Warrant Pricing. *Industrial Management Review, 6* (2), 13-39. Retrieved September 20, 2020 from https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1346056
61. Samuelson, P. A. (2015). Rational Theory of Warrant Pricing. *Henry P. McKean Jr. Selecta*, 195–232. https://doi.org/10.1007/978-3-319-22237-0_11
62. Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. Tokyo: McGraw-Hill Book Company.
63. Smailović J., Grčar M., Lavrač N., & Žnidaršič M. (2013). Predictive Sentiment Analysis of Tweets: A Stock Market Application. *Human - Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data,* 77-88. https://doi.org/10.1007/978-3-642-39146-0_8
64. Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance, 62*, 1139-1168. https://doi.org/10.2139/ssrn.685145
65. Uhl, M., Pedersen, M., & Malitius, O. (2015). What's in the News? Using News Sentiment Momentum for Tactical Asset Allocation. *The Journal of Portfolio Management, 41*, 100-112. https://doi.org/10.3905/jpm.2015.41.2.100
66. Vu, T., Chang, S., Ha, Q., & Collier, N. (2012). An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter. *Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data,* 23-38. Retrieved September 18, 2020 from https://www.aclweb.org/anthology/W12-5503

67. Wang, Y., Keswani, A., & Taylor, S. J. (2004). The Relationships between Sentiment, Returns and Volatility. *SSRN Electronic Journal, 21*(1), 1-34. https://doi.org/doi:10.2139/ssrn.499502
68. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*, 80–83. https://doi.org/10.2307/3001968
69. Wu, D., Zheng, L., & Olson, D. (2014). A Decision Support Approach for Online Stock Forum Sentiment Analysis. IEEE Transactions on Systems, Man, and Cybernetics: Systems, *44*,1077-1087. https://doi.org/10.1109/TSMC.2013.2295353
70. Wüthrich, B., Permunetilleke, D., Leung, S., Cho, V., Zhang, J., & Lam, W. (1998). Daily Prediction of Major Stock Indices from Textual WWW Data. *HKIE Transactions, 5*(3), 151-156. https://doi.org/10.1080/1023697X.1998.10667783
71. Xu, Y., & Cohen, S. (2018). Stock Movement Prediction from Tweets and Historical Prices. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1*, 1970-1979. https://doi.org/10.18653/v1/P18-1183
72. Yuan, Y. (2012). Market-Wide Attention, Trading, and Stock Returns. *SSRN Electronic Journal, 116*(3), 548-564. https://doi.org/10.2139/ssrn.1105532
73. Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems, 55*(4), 919–926. https://doi.org/10.1016/j.dss.2012.12.028
74. Zhang, X., Fuehres, H., & Gloor, P. (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. https://doi.org/10.1016/j.sbspro.2011.10.562
75. Zhang, X., Fuehres, H., & Gloor, P. A. (2012). Predicting Asset Value through Twitter Buzz. *Advances in Intelligent and Soft Computing, 113*. https://doi.org/10.1007/978-3-642-25321-8_3
76. Zhang, W. & Skiena, S. (2010). Trading Strategies to Exploit Blog and News Sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, *4*(1), 375-378. Retrieved September 18, 2020 from https://ojs.aaai.org/index.php/ICWSM/article/view/14075

APPENDICES

Appendix 1: Povzetek (Summary in Slovene language)

V moji magistrski nalogi sem raziskovala razmerje med javnim sentimentom, povzročenim s strani tvitov Borisa Johnsona ter med premiki finančnega trga. Rezultati so statistično pokazali, da imajo tviti pri opazovanju tega razmerja skozi krajši časovni okvir pomemben pozitiven vpliv na donos trga vrednostnih papirjev. Na drugi strani pa negativnim tvitom ni sledil negativen abnormalen donos. Abnormalni donosi so se pojavili šele v okviru 1-2 dni pred in po pojavu samega tvita. Opisano razmerje se je na dolgi rok izkazalo kot brez pomena.

Ob upoštevanju te in ostalih študij s podobnimi izidinam postane zelo jasno, kako pomembna je postala vloga čustev sproženih s strani socialnih medijev, ko iščemo nove natančne napovedovalne modele.

Kljub navedenemu je pomembno omeniti, da nekatere stvari v moji nalogi niso bile upoštevane, luknje pa lahko v prihodnosti zapolnijo nove raziskave. Kot prvo stvar bi izpostavila dejstvo, da je moja magistrska naloga ponudila nekoliko bolj unikaten pristop v sklopu določanja javnega sentimenta, in sicer z osredotočanjem na Twitter novice samo ene vplivne osebe namesto pridobivanja javnega sentimenta iz velike količine podatkov iz Twitter novic. Tako je bila naloga rahlo omejena pri količini podatkov, ki so bili analizirani, kar je lahko potencialno vodilo do manj natančnih rezultatov. Povečevanje števila analiziranih koreliranih Twitter profilov ter povečevanje časovnega obdobja sta lahko možni usmeritvi prihodnjim raziskavam. Kot drugo stvar bi rada izpostavila, da je za namen analiziranja sentimenta tvitov bil zaradi preprostosti in priročnosti uporabljen pristop baziran na slovarju, vendar pa ta metoda ne vzame v zakup konteksta v tekstu, kar lahko vodi do izgube pomembnih informacij. Bolj napredne metode analize sentimenta, kot je na primer metoda strojnega učenja, so lahko uporabljene v prihodnjih študijah z namenom izboljšanja natančnosti rezultatov. Za konec bi rada poudarila še to, da modeli, uporabljeni v magistrski nalogi, domnevajo linearno razmerje med spremenljivkami, kar je lahko precej restriktivno za gibanje na finančnem trgu. V prihodnosti so lahko namesto le-teh uporabljene bolj sofisticirane metode, ki zajemajo ne-linearno razmerje analiziranih spremenljivk.