Market sentiment and the British stock exchange market

Auteur : Hanser, Tanja
Promoteur(s) : Bodson, Laurent
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en sciences économiques,orientation générale, à finalité spécialisée en Economics and Finance
Année académique : 2017-2018
URI/URL : http://hdl.handle.net/2268.2/5669

Market Sentiment and the British Stock Exchange Market

Jury:	Dissertation by
Promoter:	**Tanja Hanser**
Laurent Bodson	For a
Reader:	M.Sc. in Economics and Finance
Thomas Lejeune	Academic Year:
Reader:	2017/2018
Prof. Dr. Robert Jung	

Contents

List of Figures

Acknowledgment

I want to thank Mr. Bodson for giving me the idea to look at sentiment analysis and get acquainted to machine learning techniques. Moreover, I want to thank him for receiving the opportunity to get more involved with the software R which meant to get frustrated and depressed when things were not working but at the same time to get happy and satisfied when they were working. Additionally, I want to thank Mr. Lejeune for providing me technical support when questions occurred from my side. Furthermore, I want to thank Mr. Jung for him being flexible and shortly available. Lastly, I want to thank Ms. Gillet for her endurance and organisation throughout the degree in Liège.

1 Introduction

Different stock exchange market indices compromise performances and beliefs about performances of diverse companies into an aggregate measurement of a market. In particular, Financial Times Stock Exchange 100 (FTSE 100) is a market-capitalisation weighted index of the United Kingdom's listed blue-chip companies. Hence, it aims to measure the performance of the 100 largest companies traded on the London Stock Exchange that cover industries like basic material, consumer goods, technology, financial services or healthcare. In February 2018, FTSE 100 encompassed more than two trillion pounds sterling of market capitalisation, covering 81 % of the whole market capitalisation in the British market relative to the market cap of FTSE-ALL shares. (LSEG FTSE100). Accordingly, FTSE 100 is a suitable measure to try to understand the movements in the performance of the British market. Particularly, it is of interest whether the beliefs about the performance, also called sentiment can predict the movement of the FTSE 100. Large literature has investigated the influence of sentiment on stock performance and its predictability primarily in the United States of America. Yet, a small attempt has been made to look at other sophisticated capital markets than the USA which is the matter of this master thesis.

The question that arises is what are those beliefs or sentiments? Beer & Zouaoui (2013) define sentiment as a belief about future cash flows and investment risks that is not warranted by fundamentals. Bormann (2013) points out that sentiment is a latent process, thus a process which is existent however not visible. Baker & Wurgler (2006) define sentiment, on the one hand, as the propensity to speculate. It can be understood as a market sentiment or feeling in the market that drives the relative demand for speculative investments and therefore cause cross-sectional effects even if arbitrage forces are the same across stocks. On the other hand, according to them sentiment can be understood as investors' feeling being either optimistic or pessimistic about a stock, so arbitrary waves of sentiment might affect the cross section.

Thus, seeing sentiment as either one of these definitions, two follow-up questions occur: firstly, is it possible to make the latent process visible and measure the impact of the sentiment and, secondly, does sentiment cause general pattern which can be observed as consequences on the market?

According to the classical theory where by assumption market participants with rational expectations act on efficient markets combined with the theory of efficient market hypothesis from Fama (1970), emotional distortion can happen once but on average are irrelevant due to perfect information (Bormann (2013)). Thus, prices move and fluctuate randomly because information is instantaneously incorporated into prices (Da et al. (2014)). As a consequence, sentiment as defined by Baker & Wurgler (2006) cannot occur.

However, contrary to that the field of behavioural finance suggests that emotions play an important role in financial decision making. De Long et al. (1990) developed a model where they assume a rational investor and an irrational to trade according to its biases. They result in, betting against sentiment of irrational investors is costly and risky for rational agents as limits of arbitrage occur because these rational investors, also called arbitrageurs are not aggressive enough to bring prices back to fundamentals (Baker & Wurgler (2006)).

Because of the alternative model proposition, diverse suggestions have occurred in the recent history how to measure the latent process of sentiment and what their implication might be. Direct and indirect attempts have been made to quantify sentiment. A direct approach is used by authors like Brown & Cliff (2004) who develop their sentiment measure from survey data provided by the Association of Individual Investors where households are directly questioned about their attitude. Other authors like Baker & Wurgler (2006) develop indirect approaches to measure sentiment which describe the manifested behaviour of investors. They derive measures from various assessable market variables. Other authors, though, focus more on potential causes why sentiment waves or changes towards securities might occur. One of such potential causes is that emotions are triggered by the availability of information, general knowledge and the conceivability of a certain event (Bormann (2013)). Due to that, authors like Tetlock (2007) or Smales (2015) analyse the impact of news on investor's behaviour because its main task is to convey information or to convince the reader. Thus, news or, precisely, the tone of the news can either be positively or negatively inclined towards the issue which might influence the decision making of market participants. Nowadays, information transmission is augmented by social media. The more and more accelerating usage of social media and the desire to share opinion on the internet is seen as a potential source for the sentiment derivation from authors like See-To & Yang (2017) or Checkley et al. (2017). Microblogging websites like Twitter where users can write messages up to 140 characters have captured large interest because posting activity can be seen as real-time data provision, displaying sentiment hidden in the text. Using methods of natural language processing (NLP) and machine learning, authors try to reveal sentiment from such information and measure its contemporaneous impact as well as predictive power on the stock market. All in all, some authors are successful in identifying significant impact and predictive power of sentiment. Moreover, they are able to derive some common pattern and consequences of them. At the same time, other authors can not find any impact of sentiment. Therefore, the idea of sentiment is controversial.

Focusing on the British market instead of the US market in this thesis might give a slight indication whether sentiment has a universal notion. Additionally, it is important to understand whether sentiment plays a role in market movement from the perspective of practitioners as they can adjust and implement more sophisticated investment strategies to achieve higher goals. Moreover, finding support for behavioural finance might lead to the claim to further reconsider classical theory.

Due to that, this thesis follows the ideas of sentiment analysis by trying to retrieve a sentiment index from Twitter for FTSE 100 and check whether in the short-term sentiment it can predict the return and the volatility of the British market or not.

In the following, firstly, a detailed literature review is presented for the sentiment analysis. Secondly, Twitter sentiment derivation for the British market is described. Thirdly, other commonly used sentiment measures by practitioners are presented and, fourthly, all sentiment measures are set into relation with FTSE 100's return and realised volatility. Specifically, a vector autoregressive is identified to conduct Granger causality and, finally, predictability is assessed with forecasting by looking at the behaviour of error metrics compared to the baseline which is run without sentiment. Thirdly, critical issue about the results are discussed and lastly, a conclusion is drawn.

2 Literature Review

Following the classical theory where rationality is assumed combined with the efficient market hypothesis, sentiment should be ruled out by definition. Distortions that might occur due to feelings like euphoria or fear would "at best" (Bormann (2013)) only have short-term impact and occur only as noise if at all. Since all information will be incorporated into the price development on average the right expectation will always be formed (Bormann (2013)). Therefore, sentiment indicators should not exhibit any kind of predictive power (Feldman (2010)). On the contrary, De Long et al. (1990) discard the assumption of rationality and assume some part of the economic subjects to be emotionally driven which leads to other results than predicted by the classical theory. According to them, fluctuations in stock prices are derived from the irrational behaviour of traders who are prone to sentiments and, therefore, exhibit unpredictable irrational behaviour.

From that point on, the field of behavioural finance has evolved and developed various sub categories that try to prove the shortcomings of the efficient market hypothesis. They try to find other explanations for movements in stocks rather them just being random. The following review summarises the results from some subcategories. Firstly, the idea of the efficient market hypothesis and De Long et al. (1990)'s model are presented. Secondly, literature review on the development of sentiment indices derived from various market variables is made. Thirdly, the area of news sentiments and, lastly, the most recent work in news sentiments on word mining, specifically from social media platforms are outlined. Table 1 gives an overview over the literature with concrete dates.

Fama (1970) states that "a market in which prices always "fully reflect" available information is called "efficient"". It is efficient because only then "accurate signals" for resource allocation could be sent to the market to make proper production-investment decisions. In his means, the term "fully reflect" signifies that any kind of information is completely used in determining equilibrium of expected returns like presented in equation 1

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

where p_j signifies the price of a specific object, r the return, t today and $t+1$ tomorrow, E the operator for the expected value and Φ the information set. Thus, the price can then and only then evolve after information is processed or reflected in the formation of the current price. This, though, rules out the possibility of trading systems where expected profits or returns can be achieved in excess of equilibrium expected profits. Fama concludes that the development of prices is a fair game with no option of trading. Chocrane (2013) points out concisely that price developments are unpredictable because any kind of information is already incorporated fully into the price development. This represents an efficient market in the means of Fama.

De Long et al. (1990) highlight that according to the efficient market hypothesis assets should be sold at their fundamental values. However, they develop a model where prices can deviate from their fundamentals due to unpredictable irrational behaviour. Their model is based on an overlapping-generations model with two periods, two types of traders and two assets, a risky and risk free one, from which they derive their conclusions. De Long et al. differentiate between a rational trader, who trades according to fundamentals and whom they call the arbitrageur and an irrational trader whom they call the noise trader. Noise traders have the possibility to exhibit sentiment which manifests in their beliefs of price development. The noise trade can either be bullish (optimistic) or bearish (pessimistic) about the future price development. According to that inner attitude, the noise trader can either push the price up or down. His perceptions or beliefs, though, are not predictable which serves the highest risk to the model and make price fluctuate heavily. The arbitrageur, yet, has to react to price development by either buying or selling the asset before the disbeliefs become even stronger and the arbitrageur suffers high loss which can occur due to short-term limitation. Thus, because of misperception of the noise trader, arbitrage "deters" and it leads to significant divergence form fundamental values even if there is no fundamental risk. In his setting, arbitrageurs are not able to offset the effects of noise traders. Instead the noise traders experience higher expected return than fundamentals traders because they overestimate returns and underestimate their risk. Therefore, they invest more in the risky assets than the fundamental traders who refrain from trade because of their risk-aversion. De Long et al. emphasise that the most interesting part is not the relation between the return and risk but the rise in risk which is self-produced and destabilising. At last, arbitrageurs bear the risk that is made up as they react more to the noise traders than trade on fundamental. They try to assess their movement to bet against them which turns out to be costly and risky.

Based on these contrasting theories, a field of literature emerged to prove the importance of feelings by developing sentiment measures. Notably is that nearly all the reviewed papers covered the US stock market through the analysis either of the Dow Jones Industrial Average (DJIA), National Association of Securities Dealers Automated Quotations (NASDAQ) or Standard & Poor's 500 (S&P 500) with only a few exceptions of Germany and China.

Baker & Wurgler (2007) argue that the Great Crash of 1929 or the Dot.com bubble of the 1990s could not be explained by the standard financial models which is why new solution are needed to explain these crashes. According to them, the consistent rise in stock prices around these extreme events are driven by the price rise of speculative and difficult-to-value stocks for which investor's sentiment has risen and burst at some point. Due to that, they do not question the fact whether sentiment has an impact on stock movement but rather asked the question how to measure it. They suggest extracting the latent process from market variables of trading volume, dividend, premium, closed-end fund discount, the number of first-day returns on initial public offerings and the equity share in new issues by applying principal component analysis (PCA). Their index is able to capture sentiment volatility around speculative major events like the Dot.com crisis. On top of that, they identify that the hard-to-value stock which are for instance young, unprofitable, high-volatility growth stocks are the most prone to investors' sentiment because these types of companies are difficult to arbitrage and, therefore, difficult to value.

Baker & Wurgler (2006) arrive to the same conclusion in 2006 where high-to-value stock are more prone to investors' sentiment whereas older firms with high valuation history are not highly affected by the sentiment. Further, they identify a pattern that when sentiment are low relative to the average, subsequent returns are relatively high (to their average) whereas when sentiment are high, subsequent returns are relatively low. Controlling for the three-factor model of Fama and French, Baker & Wurgler find significant predictability power of their sentiment index which stands in contrary to the efficient market hypothesis. This is intriguing as Baker & Wurgler covered a long-term period of forty years. Hence, according to their analysis sentiment plays a role not even "at best" in the short term but also in the long term.

Baker et al. (2010) augment their model to international trade activities by applying their sentiment approach to six large western counties. These indices represent local sentiments that could capture individual country-specific movements. The global sentiment is developed as a linear combination of principal component of the local sentiments. They base their valuation on the dual-listed shares which are pairs of securities that claim equal cash flows but trade in different markets and sometimes at substantially different prices. They point out that country-level results are significantly driven by the global index whereas local indices impact the hard-to-value firms. Thus, this paper hints to the fact that sentiment may have multi-layered characteristics with different kind of influence power.

Finter et al. (2012) follow the approach of Baker & Wurgler (2006) and develop a sentiment index for Germany. They find significant contemporaneous relation between sentiment and the excess return even when controlling for Fama French factor model. However, this relation does not lead to severe relative mispricing over time. Further, their result deviate from Baker & Wurgler as they do not find "much predictive power of sentiment for future stock returns". They reason that the proportion of institutional traders (arbitrageurs) compared to retail trades (noise traders) is bigger in Germany compared to the USA where it is roughly equal. This finding shows that the importance of the relation between rational and irrational investors as described in De Long et al. (1990) impacts the characteristics of the sentiment and its enforcement.

Shen et al. (2017) also use the developed sentiment of Baker & Wurgler (2006) and try to identify pattern in returns behaviour between companies that are highly exposed to market risk to low exposed market risk. They find out that following low sentiment periods high-risk companies earn more than low-risk companies from which they conclude that during this period time the market seems to function efficiently. However, following high sentiment high-risk firms do not earn higher returns than the low-risk companies. During this time high-risk companies are more likely to be overpriced than low-risk companies which leads in the correction process to smaller returns for high-risk companies than for the low-risk companies. Thus, they reason that during this time markets do not work efficiently because sentiment-driven investors undermine the traditional risk-return trade off.

Feldman (2010) dares to compare different existent sentiment indices. He takes the University of Michigan consumer confidence index where households in the USA are surveyed about their consumer confidence, the sentiment developed by Baker & Wurgler (2006), the CBOE's Implied Volatility (VIX) which is seen as the fear gauge by practitioners, put-call ratio and advance to decline ratio derived from option trading into consideration. Moreover, he develops the so-called perceived loss index which is based on negative weekly returns of mutual funds which are exponentially averaged to put more emphasis on the most current performances. The obtained weights are multiplied with the total assets which construct the perceived loss index that, consequently, represent a bearish sentiment. Controlling for the Fama French three-factor model, Feldman finds no significant power in explaining contemporaneous returns for neither of the indices. Nevertheless, he figures out that the perceived loss index outperforms the other indices in predicting returns in the middle run, especially for one- and two-year horizons.

Brown & Cliff (2005) do not use data provided as market variables but rather use the information provided by the investor's intelligence that tracks several market newsletters that are valued as being either bullish, bearish or neutral. Their idea is that investors might be influenced through the provided information such that their inner attitude might change towards some securities. Given these newsletter evaluations Brown & Cliff calculate the sentiment index as the bull-bear spread which is defined as the percentage of newsletters being bullish minus the percentage of newsletters being bearish. Their main goal is to identify whether the market is overvalued during optimistic periods and if so whether the following periods are characterised by low cumulative long-run returns because prices revert to their intrinsic values. They test their hypothesis by relating the level of sentiment to market mispricing proxied by the DIJA pricing errors. They test two approaches, firstly, they relate the level of the sentiment to the pricing error, and, secondly, they cointegrate these time series. Both tests lead to the same result that the market is overvalued during periods of optimism and undervalued during period of pessimism. Furthermore, high levels of sentiment result in significantly lower returns over the next two or three years. They confirm the model of De Long et al. (1990) that optimistic (pessimistic) investors drive prices above (below) fundamental values that revert to fundamental value at some point.

Interestingly is that these papers look a long-term horizon with at least 15 years perspective in Finter et al. (2012) to 50 years in Shen et al. (2017) and the cross-sectional impact. They seem to find some important role of sentiment in the long-run. During high sentiment companies seem to be overvalued which leads to low return in the subsequent time, whereas during low sentiment periods companies seem to be undervalued which leads to high return in the subsequent years. This pattern seems to be dependent on the structure or the type of the company. Young, unprofitable, high-volatility growth stocks are more affected by the mood swings because they are more speculated about than long-lived sophisticated once. The exposure to market risk shows that high-risk companies exhibit anomalies during high sentiment by achieving less higher returns than the low-risk companies. Moreover, the countries' investor structure seems to have a high impact on the severity of the sentiment characteristics. But at the same time, it is doubtful whether sentiment constitutes in these way as some author do not find significant influence of sentiment.

Tetlock (2007) uses the same idea as Brown & Cliff that news content might be related to individual investor's psychology, thus, the media might impact investor's decision making and consequently influence the stock returns. He examines this relation by developing a sentiment index which is based on the column from the Wall Street Journal. Using the General Inquirer (GI), a quantitative content analysis program, he identifies word frequencies for 77 predefined lexical categories.

Applying PCA to these categories, he receives a single media factor that captures the maximum variation in the GI categories. This media factor is strongly related to pessimistic words in the column which he calls the media factor a pessimism factor. To obtain intertemporal relation between the pessimism factor and the stock movements, he applies the vector autoregression model (VAR) which finds that the pessimism factor significantly predicts downward pressure on market prices and trading volume which successively revert to fundamentals. To the other side of the coin, he also ascertains that market returns drive high media pessimism. Finally, he confirms that the noise trader theory holds. According to Tetlock, opinion media content has sudden influence on investors' sentiment that affects their trading behaviour and that evokes noise on the markets. Hence, media content is in such a way powerful that it can guide investor's mood waves which is why it is not necessary to observe the manifested sentiment via market variables but rather look at the source for sentiment shifts which occur due to the shifts in information provision.

Garcia (2013) also analyses the relationship between news and stocks. He develops a media sentiment index from general financial news which are published in the New York Times and the Wall Street Journal. He counts positive and negative words from the articles and assigns positive and negative sentiment as a fraction of the counted words. Contrary to Tetlock, Garcia also identifies the importance of positive words and not only negative ones that can predict the market. It seems that investors are sensitive to news during recessions. Furthermore, he realises that predictability effect is stronger during weekends because investors have time to read articles and take decisions based on them. Nevertheless, the impact of news vanishes within four days which might be a sign for non-informational impact of news.

Yuan (2015) examines the impact of as he calls the "market-wide attention-grabbing events" on the trading behaviour of the investors and, market returns. Precisely, he looks at record-breaking events for the Dow, Nasdaq Composite Index, the NYSE Composite Index and the S&P 500 Index and front-page articles that report about these events in the New York Times and the Los Angeles Times at the same time and tries to find a relation and trading pattern with market returns. He proxies investor's behaviour with aggregate order flow, aggregate daily mutual fund flows and individual trading records from a large brokerage firm. His results show that following Dow record events or news events when the market index is high, high levels of individual-investor aggregate net selling flow are observed. Moreover, Dow record events predict negative market returns as the market drops. When the market is high, front-page news events predict markets returns to be negative comparable to that of Dow record events but news shows little predictive ability when the market is low. Shortly, aggregate as well as household-level data reveal that high-market wide attention events lead investors to sell their stocks holdings dramatically when the level of stock market is high which reduces returns.

Smales (2015) investigates the relationship between time-variation of beta with the time variation of industry-specific news sentiment effects. He constructs a news sentiment and compares it to investor's reaction which is proxied by the VIX, investor's fear. The news sentiment is calculated given the data from Thomson Reuters News Analystics that rates news either to be positive (+1), neutral (0) or negative (-1) as a weighted average of the probabilities from these ratings. Smales identifies a clear time-varying pattern between news and investor's sentiment. First, he confirms like Tetlock that the magnitude of response to negative news is greater than to positive words. During high fear, industry-specific news does not influence market returns instead they are driven by systematic factors like general financial news. When fear is low, though, Smales finds a significant relationship between industry-specific news sentiment and returns for most of the industries.

Tetlock (2011) raises the hypothesis that investors overact to financial news that are stale and persistent in the market rather than immediately react to the latest news. He analysis the content of public news events from the Dow Jones newswires. If textual similarity to previous ten stories about the same firm is existent this information is defined as stale. Firm stock returns react less to news stories that contain stale information than to news with new information. When news is stale, it predicts negative movement of future returns at different horizons from two days to two weeks. Moreover, the drop after stale news is larger for stocks with above-average individual trading activity on news days.

Schumaker et al. (2012) want to figure out whether the choice of words and the tone used by the authors of financial online news articles correlate to stock price movement and whether it can predict the stock price movements. They use news published on Yahoo! Finance at a frequency of 20 minutes whereas all the mentioned above papers for news sentiment are analysed at a one-day rate. The authors use the so-called Arizona Financial Text (AZFinText) method where they compare three different stages. Firstly, the AZFinText system without sentiment information which only identifies the proper nouns in the articles. Secondly, they identify the tone of the article being either objective, subjective or neutral with the help of OpinionFinder (a program that identifies the document's sentiment and objectivity) and, thirdly, the model of polarity is identified by dividing the articles into positive, negative and neutral sentiment. They find that subjective news articles predict price direction better than objective news. Articles with a negative sentiment predict the price direction the best compared to positive and neutral news. Furthermore, they find that positive articles follow reduction in returns whereas negative and neutral articles follow increase in returns.

Groß-Klußmann & Hautsch (2011) examine the relation between stock-specific news and the market reaction at an even higher frequency of 20 seconds. They obtain news from the Reuters NewScope Sentiment Engine that extracts news and automatically categorises them into being positive, neutral or negative. Specifically, they use firm-specific news and refrain from using earning announcements. They relate this news to high-frequency returns, volatility, trading intensity, trade sizes, trade imbalances, spreads and market depth. Using VAR model to identify the relation between news sentiment and returns, they find that the high-frequency trading activity significantly reacts to intraday company-specific news items which are identified as relevant. The strongest effect is achieved for volatility and cumulative trading volume. Bid-ask spreads, trade sizes and market depth do not necessarily react directly but rather indirectly through cross-dependencies to volumes and volatilities. According to them, the sentiment has some predictive power for price movements around news arrivals.

Yu et al. (2013) compare the effect of, on the one hand, conventional media that consists of major newspapers, television broadcasting companies and business magazines, and on the other hand, of social media that consists of data from blogs, forums, news and micro blogs like Twitter. They aim to understand their interrelatedness on short term firm stock performances. Yu et al. define sentiment index as a fraction of positive minus negative words. Social as well as conventional media have a strong interaction with stock performance but social media seems to have an even stronger relationship than conventional media. Specifically, blog and Twitter sentiment are found to have positive effect on risk.

Mao et al. (2011) also attempt to compare different kind of sources of sentiment development and try to assess their predictive power for the DIJA price, trading volumes and market volatility measured by the VIX and the price of gold. They consider surveys, news headlines, search engine data and Twitter feeds as their ground for analysis. The news sentiment is calculated from emotional words in the financial headlines of different kind of newspapers. For Twitter sentiment they take the ratio of tweets with bullish words to bearish words as well as the volume of 26 search queries. For the search engine sentiment, they use the same 26 queries and calculate the average of the search volume from Google. Finally, they use VAR models and Granger causality to check whether these sentiment measures predict financial indicators. They find that survey sentiments do not have any predictive power contrary to Google search volumes, Twitter Investor Sentiment and the frequency of occurrence of financial terms on Twitter in the previous one to two days. As a whole, Twitter seems to outperform the other measures.

Q. Li et al. (2014), firstly, examine whether fundamental information in firm-specific news articles impact trading activities of investors. Specifically, they use online financial news that is related with the companies listed on China Securities Index (CSI 100). Secondly, they investigate whether sentiment defined as emotions that are evoked via news and public social media impact on stocks. They use the model called electronic-media-aware quantitative trader, termed eMAQT that builds a weighted term vector with proper nouns that should represent firm's fundamentals from news and sentiment term that should represent the mood of the news. Particularly, they use the standard part-of-speech tagger to extract nouns from the articles. To capture sentiment words, they construct a topic based model from discussion forums in China sina.com and eastmoney.com. The non-linear model of Support Vector Regression model is used to assess the predictability of their sentiment index. They find that fundamental information particularly on restructuring and earnings issues of firm-specific articles affect the trading behaviour. Moreover, they assess relatively reliable results in predicting high trading volume stocks that are in high attention of news reports. Risky stocks measured by the β values are better predictable than low-risk stocks because they are more influenced by daily news. Additionally, stocks in consumer-related industries tend to be better predictable because consumer news articles are more rigorously considered.

Shifting the perspective to the trigger of sentiments, firstly, it is notable that time perspective changes. Authors look only at least three-month news up to one year with mainly one-day time frequency. They all seem to agree to find some impact of news on the investor's inner mood and their decision to change their trading behaviour. Analysing subjective vs. objective types of news, it seems that subjective news has a greater impact on the investor's attitude. However, it is not clear whether investors react more to persistent and old information like Tetlock suggested or whether latest news. Moreover, it seems that sentiment index obtained from Twitter can be the best related to predict stock moves. The more a stock is covered by media the better it is to model its fluctuations via the news sentiment which seems to confirm the model of noise traders from De Long et al. (1990). At the same time, it is unclear whether news predictability is truly given because in some case it occurs only around news arrivals (Yu et al. (2013)) and in the other case it vanishes within days (Garcia (2013)).

Checkley et al. (2017) study Twitter tweets and posts on StockTwits, a microblogging website like Twitter but just for traders where they can share their opinions and estimation about the movement in stock. They examine stocks of Amazon, Apple, Goldman Sachs, Google and IBM as these stock experience high social media attention at a two minutes interval rather than one-day frequency like the other authors do. The sentiment index is generated from the commercial firm called PsychSignal that uses Linguistic Inquiry and Word Count (LIWC) framework to assess the author's mood. LIWC applies word dictionaries to the tweets to extract the frequency of sentiment words that reflect emotions, thinking styles or social concern to calculate relative appearance in the text and set them into relation. They use Granger causality to identify correlation between returns, volatility and traded volume with the tweets sentiment. Particularly, they find a relationship between the sentiment and volatility and the traded volume but not to the returns. Moreover, they assess forecasting errors are materially smaller with the sentiment index compared to a baseline model without sentiment. Nevertheless, these are only modest improvements. Lastly "sentiment and market behaviour are found in times of strident and discordant sentiment" indicating that traders form a "mob rather than a wise crowd".

Antweiler & Frank (2004) examine messages posted on Yahoo!Finance and Raging Bull. They aim to figure out whether the sentiment of these messages can predict DIJA's returns and volatility and whether disagreement in the messages induces trade. Their sample size of 1.5 million messages is evaluated with the help of Naive Bayes to categorise the messages into buy, hold or sell. Given that information, a bullishness and agreement index is developed and set into relation with the aimed question by applying fixed effects regression to the data. They find out that the messages can predict return for the next day. Even though, this effect is "statistically significant but economically quite small in comparison to plausible transaction costs". They confirm that posting volume as well as the disagreement in messages induce trade contemporaneously. However, greater disagreement on one day predicts fewer trades on the next day and not more trades. Finally, they also find that these messages are able to forecast volatility.

Sprenger et al. (2014) raise the same question as Antweiler & Frank (2004) and augment the scope to the points whether message volume increases trading volume returns or volatility. Moreover, they try to analyse the importance of investment advice by looking at the reach of the individual follower base. Like Antweiler & Frank (2004) they use Naive Bayes to categorise tweets and define a bullishness and agreement index. They find relation between the sentiment and stock retuns, message volume and trading volume as well as between disagreement and volatility. Finally, according to them users that provide above average investment advise are retweeted and more followed.

Oliveira et al. (2013) analyse Apple, Amazon, Google, IBM, Goldman Sachs and Standard and Poor's index from the microblogging website StockTwits as these stock experience high attention. Particularly, they analyse whether the extracted sentiment can predict returns, volatility and trading volume. The sentiment itself is developed by counting words "bullish" and "bearish" which label the feeds of StockTwits and set into relative relation. To assess the importance of the sentiment, Oliveira et al. define different specifications for of multiple linear regression models e.g. returns depend on previous returns or returns depend on previous returns and previous sentiment measure. Furthermore, to evaluate the quality of the prediction from the specified model they apply Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). Observing a longer time period than the previous authors, Oliveira et al. do not find evidence that the tested sentiment index is able to predict returns and volatility. Nevertheless, they do find some prediction for the trading volume.

See-To & Yang (2017) analyse the role of sentiment dispersion of the DIJA from Twitter and its information content about future stock returns and realised volatility. They find that sentiment indeed contains information about stock realised volatility which can increase the prediction accuracy. Firstly, See-To & Yang accurately clean tweets and train Naive Bayes to categories tweets into being bullish or bearish. Secondly, the sentiment dispersion index is assigned to be the standard deviation of the score (relation between bullish and bearish tweets) of the sentiment per day. Thirdly, the use pooled regression to set the realised volatility and return in relation to the sentiment and sentiment dispersion. Fourthly, they check for the prediction accuracy with RMSE and, fifthly, they apply non-linear estimation technique of Support Vector Regression. Lastly, they ascertain that sentiment dispersion raises realised volatility on the same day and then reduces on the following several days but no such relation is found to returns.

Nguyen et al. (2015) aim to develop a model that can predict the direction of stock price movements (up or down) via social media message board information. Specifically, they study Yahoo Finance Message Board for 18 stocks among others Apple, Amazon, Cisco Systems and NVIDIA Corporation. They compare several sentiment measures e.g. the human sentiment which represents the percentage of classes like strong buy or sell of the annotated messages, the so-called JST sentiment where the method of joint sentiment/topic model is applied which can grasp the topic of the plain text message either it being on products, service, dividend or something else and asses the sentiment to the relevant topic or the aspect-based sentiment that extracts explicit topics via consecutive nouns and the sentiment simultaneously. Applying, non-linear Support Vector Machine to the specifications and comparing the accuracy of predictability of them, the aspect-based sentiment outperforms the other measures in accuracy.

The analysis of social media and its content ranges between three months to one year at a frequency of one day. However, it is most notable that the techniques applied to extract sentiment from the texts of either Twitter feeds or Yahoo Finance message boards get more and more advanced as they not only count the words being either positive or negative but also apply machine learning techniques from computational linguistic as Naive Bayes to categorise the feed being positive or negative, or obtain the aspect that the sentiment word is related to. Besides that, authors assume that the relation between sentiment and returns is not necessarily linear but rather non-linear which is why they use techniques as Support Vector Machines or text to vector matrices. However, they do not seem to find a concrete relation between sentiment and returns. There are rather some indications that sentiment have an impact on the volatility on the stock.

As a whole, the literature on the sentiment analysis evolved from capturing the market sentiment that are represented through market variables and use cross sectional data to explain the effects of sentiment on them to investor's sentiment measures which occurred, firstly, through the analysis of conventional news to news, information broadcasted vie social media and their impact. The main focus in these studies is put on the noise traders, individual traders as they have the potential to bring the market to deviate from the fundamental prices as De Long et al. (1990) defined. Concluding it is difficult to state whether the sentiment has an impact on returns due to variation of news types, different time spans and different techniques to extract the sentiment. At the same time, individual traders

	Author	Market	Market index	Market variables / News type - news	Time frame	Period	Number of items
Market sentiment	Baker & Wurgler (2007)	USA	NYSE, NASDAQ	NYSE turnover, dividend premium, closed-end fund discount, NIPOs, RIPOs and equity share in new issues	Yearly	01.01.1966 - 31.12.2005	
	Baker & Wurgler (2006)	USA	CRSP - Compustat database	Closed-end fund discount, NYSE share turnover, NIPOs, RIPOs, the equity share in new issues and the divend premium	Yearly	1961 - 2002 2002	
	Baker, Wurgler & Yuan (2010)	CA, FR, DE, JPN, UK, USA	Siamese twin pairs companies	Volatility premium, total volume of IPOs, RIPOs and market turnover	Yearly	1980 - 2005	
	Finter et al. (2012)	Germany	955 German stocks listed at the FWB	Net fund flows, consumer confidence survey, survey, RIPOs, net fund flows, inverted put-call ratio and trading volume	Yearly	1993 - 2006	
	Shen et al. (2017)	USA	NASDAQ	NIPOs, RIPOs, dividend premium, closed-end fund fund discount, NYSE turnover, equity share in new issues	Monthly	01.07.1965 - 31.12.2014	
	Feldman (2010)	USA	Mutual fund data from CRSP and Bloomberg	Negative weekly returns which are exponentially averaged	Weekly	01.01.1984 - 31.12.2007	
	Brown & Cliff (2005)	USA	DJIA	Newsletter - Survey data II that categorises appr. 130 market newsletters	Monthly	01.01.1963 - 31.12.2000	
News sentiment	Tetlock (2007)	USA	DJIA	Newspaper - Column from the WSJ	Daily	01.01.1984 - 17.07.1999	
	Garcia (2013)	USA	CRSP covered companies	Newspaper - General financial news from the NYT and WSJ	Daily	01.01.1905 - 31.12.2005	55,307
	Yuan (2015)	USA	DJIA, NASDAQ, S&P 500	Newspaper - The NYT and LAT	Yearly	01.01.1983 - 31.12.2005	
	Smales (2015)	USA	NYSE, NASDAQ ans AMEX	Online industry - specific news - Thomson Reuters News Analytics	Daily	02.01.2004 - 31.12.2010	
	Tetlock (2011)	USA	DJIA	Newspaper - DJIA newswires	Daily	01.11.1996 - 01.10.2008	
	Schmuker et al. (2012)	USA	S&P500	Online news - Yahoo! Finance	20 min.	26.10.2005 - 28.11.2005	
	Groß-Klußmann & Hautsch (2011)	UK	Stocks traded at the LSE	Online stock - specific news - Reuters NewScope Sentiment Engine	20 sec.	01.01.2007 - 01.06.2008	29,497 headlines
	Mao et al. (2011)	USA	DJIA and gold price	Online news and social media - News headlines, Twitter, volume of Google search queries	Daily	01.07.2010 - 30.09.2011	
	Yu et al. (2013)	USA	Randomly selected 824 companies	Social media and conventional media - Social media: blogs, forums, Twitter Conventional media: major newspapers, television, broadcasting companies and business magazines	Daily	01.07.2011 - 30.09.2011	52,746 messages
	Mao et al. (2011)	USA	DJIA	Micro blog - Twitter	Daily	28.02.2008 - 19.12.2008	9,853,498 tweets
	Li et al. (2014)	CH	CSI 100	Online financial news and socal media - Stock discussion forum sina.com and eastmoney.com	Minutes	01.01.2011 - 31.12.2011	
Social media sentiment	Checkley et al. (2017)	USA	Amazon, Apple, Goldman Sachs, Google, IBM	Micro blog - Twitter and StockTwits	2 min.		
	Antweiler & Frank (2004)	USA	DIJA	Social media - Yahoo! Finance and Raging Bull	Daily	01.01.2000 - 31.12.2000	1,5 mio. messages
	Sprenger et al. (2014)	USA	S&P 100	Micro blog - Twitter	Daily	01.01.2010 - 30.06.2010	250,000 tweets
	Oliveira et al. (2013)	USA	Apple, Amazon, Google, IBM, Goldman Sachs, S&P	Micro blog - StockTwits	Daily	01.06.2010 - 31.10.2012	390,000 stocktwits
	See-To & Yang (2017)	USA	DJIA	Micro blog - Twitter	Daily	20.01.2015 - 17.07.2015	1,170,414 tweets
	Nguyen et al. (2015)	USA	18 stocks e.g.: Apple, Amazon, Cisco Systems, and NVIDIA	Social media - Yahoo Fiance Message Board	Daily	23.07.2012 - 19.07.2013	

NIPO = number of IPOs, RIPO = first day returns on IPOs, II = Inteligent Investor, WSJ = Wall Street Journal, NYT = New York Times, LAT = Los Angeles Times,

Table 1: Literature overview

can be affected from different news sources. Nevertheless, there seems to be some evidence that sentiment next to other factors like the economy state can play a role in determining returns and their volatility.

3 Twitter sentiment derivation

The literature review has shown a trend from market based analysis to individual investors analysis, focusing on the noise traders and assuming them to be represented the latest via social media interaction. The sentiment that these social media interactions reveal is mined with computational linguistics and machine learning approaches.

The aim of this master thesis is to develop such a sentiment index for the British market from the social media platform Twitter for the short-term period. Moreover, it aims to assess whether the sentiment index can add information to the predictability of FTSE 100's returns as well as its realised volatility. It would have been interesting to check the presented methods from the long-term approaches. But, due to time restriction and data availability, these methods are not taken into account.

Particularly, Twitter is taken as a base for analysis due to its advantages over other sources like blogs, columns or messaging boards. The limited length of 140 characters to express sentiment holder's opinion is one of the advantages of Twitter. Technically, Liu (2015) defines opinion as a quadruple

$$(g, s, h, t) \tag{2}$$

where g stands for the sentiment target which can be further divided into an entity and an aspect where entity describes for example a class of a product like a camera and aspect describes a feature of the entity like the length of the camera. s represents the sentiment of the opinion, h stands for the opinion holder and t for opinion's posting time. Due to the restriction in length, the opinion holder is less likely to talk about different sentiment targets other than his feelings or beliefs about FTSE 100's performance compared to blogs or discussion messing boards where entities and the aspects of the issue might change constantly. Because of this, a simplified assumption is made that all tweets represent the same entity and aspect being an opinion or a feeling towards FTSE 100' stock performance. Furthermore, Liu (2015) defines s, the sentiment as an underlying emotion associated with an opinion which can be represented as a triple from the text

$$(y, o, i) \tag{3}$$

where y signifies the type of the sentiment. It is either a rational sentiment, expressing no emotion or an emotional sentiment expressing emotion. o represents the orientation of the sentiment being either positive or negative which is also called polarity and i represents the intensity of the emotion where the orientation is quantified with a number. Twitter feeds have a unique character as they depict sudden exclamation either of joy or sadness which is retweeted if shared the same opinion or commented but rarely scrutinised or discussed in such a way like in blogs or messaging boards. Thus, Twitter expresses less rationalised sentiment feeling rather purer unpredictable exclamation of emotional movement. Apart form that, Twitter is the most popular daily used social media platform among microblogging websites and messaging boards. Approximately 100 million people use Twitter actively on a daily base and post around 500 million tweets per day (www.omnicoreagency.com/twitter-statistics). Due to that amount of posting activity Twitter might serve as a good sample selection resource to represent noise investor traders' behaviour.

Consequently, the underlying sentiment in the means of Liu (2015) is derived with the help of the free software environment for statistical computing and graphics R which is used throughout the analysis.

3.1 Twitter data collection and description

Twitter data can be collected with two options: on the one hand, you can use the streaming application programming interface (API, a computer communication system) where tweets are real-time downloaded according to the searching queries or, on the other hand, the historical API can be used where you can download the last nine days Twitter history for free. The access to tweets that were posted prior to nine days can be purchased costly. In this case, the historical API is used to collect Twitter feeds starting on the 28.05.2018 until the 31.07.2018. As a whole, a period of two month with 47 trading days is covered.

Specifically, search queries follow the same pattern $\$tickersymbol$. Like the hashtag #, the cashtag \$ tags a tweet with an important category or topic but it is only related to financial topics. This search pattern is used to avoid noise data that is not related to the defined sentiment target e.g. hiring opportunities (Sprenger et al. (2014)).

Searching for the query $\$FTSE$, on average results in a volume of 400 tweets per week. Compared to the corresponding US index like the S&P, the search query $\$SPX$ on average results in 9,500 tweets per week. Relying only on the search

$FTSE, $III, $ASBFY, $ADM, $AAL, $ANFY, $AHT, $AZN, AVIVA PLC, $BA, $BDEV, $BKG, $BLT, $BP, $BTI, $BLND, $BTA, $BUNZL, $BRBY, $CCL, $CNA, $CCHGY, $CPG, $CRH, $CRDA, $DCC, $DGE, $DLG, $EZJ, $EVR, $EXPN, $FERG, $FRES, $GFS, $GSK, $GLEN, $HLMA, $HL, $HSBC, $IMB, $INF, $IHG, $ITRK, $IAG, $ITV, $JMAT, $JSTTY, $KGF, $LAND, $LGEN, $LYG, $LSE, $MKS, $MDC, $MRO, $MFGP, $MNDI, $MRW,,$NGG, $NMC, $OML, $PPB, $PSO, $PSN, $PRU, $GOLD, $RDS.A, $RDS.B, $RB, $REL, $RTO, $RIO, $RR, $RBS, $RMG, $RSA, $SGE, $JSAIY, $SDR, $SMT, $SGRO, $SVT, $SHPG, $SKY, $SNN, $SMDS, $SMIN, $SKG, $SSE, $STJ, $STAN, $SLA, $TWODF, $TSCO, $TUI, $UL, $UUGRY, $VOD, $WTB

Table 2: Search queries

query $\$FTSE$ would result in a very small sample size from which it is going to be even more difficult to construct a sentiment index. Due to that, attempts are made to increase the sample size by searching for all the companies that are represented by the FTSE 100 because beliefs about the individual stock might also influence the aggregate index and, therefore, the British market. Nevertheless, it is important to highlight that this approach undermines cross-sectional dimension of companies. Therefore, it must be kept in mind when interpreting. Table 2 presents all the search queries that are made, only the query for the Next PLC is excluded because the query for $\$NXT$ resulted in extracting all tweets with the word *next* inside but not necessarily tweets about the company Next PLC. The overall sample size with the search for all companies is 88,136 tweets large and without only 6,534.

Figure 1 shows the absolute frequency of the posting activity during the whole period for the companies in the FTSE 100. The range of the positing activity is quite large, stretched between only two posts for Intertek PLC, a multinational inspection, product testing and certification company and Scottish Mortgage Investment Trust, a publicly traded investment trust and 17,479 tweets for Sky PLC,

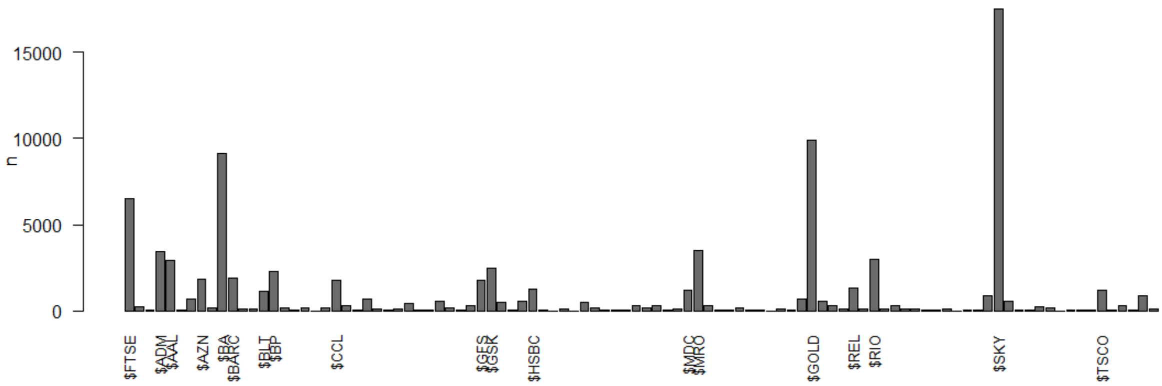

Figure 1: Absolute tweets frequency

a media and telecommunication company. Sky PLC, Randgold Resources Ltd, gold mining company, Barclays, a financial service company and FTSE 100 are the dominating outliers in the sample with tweets activity being above 6,500. 15 % of companies range between 1,001 and 6,500 tweets and 81 % between 0 and 1,000. Thus, 81 % of British companies do not stand in high attention of the Twitter community at least in financial terms. Consequently, the final sentiment index is going to be primarily dependent on the unpredicted emotion shifts of the highest outliers which might bias the results and must also be kept in mind.

An important feature of Twitter is the possibility to retweet other uses tweets by leaving them as the original one, modifying them slightly or commenting them. Roughly 29 % of the sample size are retweeted tweets. Contrary to See-To & Yang (2017), retweets are kept in the data sample because they show that the person who retweets the post agrees with the initial writer and feels the same way towards the issue. Thus, this person has the same underlying opinion.

3.2 Twitter data preparation

It is necessary to adjust Twitter data, firstly, because posting tweets is not limited to time whereas trading is and, secondly, because tweets are characterised by unstructured data which means it covers data elements that are different from numbers.

Hence, in the first case, tweeting activity is matched with the trading hours of the London Stock Exchange. So, tweets that are published after 16:30 o'clock are counted to the following day since the potential trade can only be expressed on the next day. On top of that, week-end data is also matched with Monday trading hours. In the second case, data of tweet feeds needs to be prepared and structured before being analysed. This process is called text cleaning. Particularly, in sentiment derivation the focus is put on words that might transmit the value of the sentiment. Even though, tweets can look like the following one: "Rentokil Initial plc $RTO Increased 0.26 %" which was posted on the 22.06.2018 by heraldks. The sentiment polarity is transferred via the word "increased" rather than the plain digit. Due to that, firstly, all digit in tweets are removed. Secondly, URLs forwarding to another websites are redundant because they do not convey a particular sentiment which is why they are removed as well. Thirdly, tweets can either start with *RT* or with an @ which displays that theses tweets are retweeted or addressed to with a comment. The interesting part about it is that people share the same sentiment like the initial Twitter user rather than who they address to. Therefore, these text elements are also deleted. Fourthly, punctuation signs like points or commas are also removed. Although, removing questions marks might lead to the loss of sarcasm that a sentence might convey, it is ignored because sarcasm or irony is a field of NLP which is still researched and tried to be implemented to an algorithm (Liu (2015)). Nevertheless, it is a critical point that needs to be kept in mind. Fifthly, some words like *the*, *a*, *I*, $yourself$ or *them* appear the most frequently in texts which are summarised as stopwords. Stopwords, though, do not serve any valuable additional information to sentiment derivation which is why they are also removed from the text. Sixthly, words are lemmatised meaning that words are converted into their fundamental word form e.g. *were* or *is* are transformed into *be* or words like *buying*, *bought* are transformed to *buy*. Lastly, the remained words might serve well the purpose but still some words might still be irrelevant as they can be ambiguous to being positive or negative. Therefore, the tf-idf measure is applied.

Silge & Robinson (2017) point out that the tf-idf meausre intends to identify how important a word is to a document in a collection of documents, so how important a word is to a Twitter feed in a collection of Twitter feeds. Tf-idf is a product of tf, shortly for term frequency, how often a word occurs in the document and *idf*, shortly for inverse document frequency which decreases the weight for commonly used words and increases the weight for words that are not used very much in a collection of documents (Silge & Robinson (2017)). Particularly, tf-idf is calculated according to the equation 4 to 6.

$$tf - idf(t, d, D) = tf(t, d, D) * idf(t, D) \tag{4}$$

$$tf = \frac{f_{t,d}}{\Sigma_{t \in d} f_{t,d}} \tag{5}$$

$$idf(t, D) = log(\frac{N}{|d \in D : t \in d|}) \tag{6}$$

where t stands for term that appears in d the document. D stands for the maximum number of documents which is specified by N and $f_{t,d}$ is the term frequency in a document. Terms with low scoring in tf-idf are added to stopwords because they occur too often in documents serving no additional information to sentiment derivation and, thus, they are also removed from the text.

3.3 Twitter data categorisation

The cleaned data is, finally, used to categorise tweets into their polarity being either positive or negative. To categorise tweets, two approaches are used: on the one hand, the dictionary approach as in Checkley et al. (2017) and, on the other hand, the Naive Bayes algorithm that exploits conditional probabilities as in See-To & Yang (2017), Antweiler & Frank (2004) or Schumaker et al. (2012).

The dictionary approach categorises terms with the help of a predefined dictionary which are publicly available. The choice of such a dictionary is essential since important information value might be lost when using inappropriate one. Due to that, terms in the analysis are matched to the dictionary of Loughran and McDonald which is specifically built to match financial language (Loughran & McDonald (2015)). In particular, they define terms into five categories of being either, positive, negative, litigious, superfluous or contrasting. Aiming to draw a clear distinction in investor's sentiment, so them to be either bullish or bearish towards the British market, only the positive and negative category is used to match the cleaned text data.

To match Loughran & McDonald's dictionary to the unstructured data, data is, firstly, tokenised, meaning that the sentences are broken down into their individual words and represented in a vector. Liu (2015) reports that unigram (individual word representation) and n-gram (n-subsequent words of the sentence) text representation "have been shown to be highly effective for sentiment classification". Particularly, the bi-gram representation is used in this analysis which means that two following words are embedded into a matrix of two vectors. It is of high interest to include negation words that might change the polarity direction of the sentiment. Thus, when a preceding word is part of negation terms like $isn't$ or $aren't$ the sentiment direction which is identified through the match of the last word in the bi-gram with the dictionary is changed to the opposite polarity.

The advantage of a dictionary approach is that terms are already categorised. However, the shortcoming of this method is the static behaviour of it because abbreviations, modern language usage or the social media language usage are categorically left out. For instance the term like *omg* which is a short form for *ohmygod* usually expresses a positive sentiment, might be excluded with this method. Due to that, the dictionary is augmented by twenty terms from the tf-idf calculation that have high scores. For example the term oof is identified to be highly important in the documents which is added as *negative* to the dictionary as it describes an

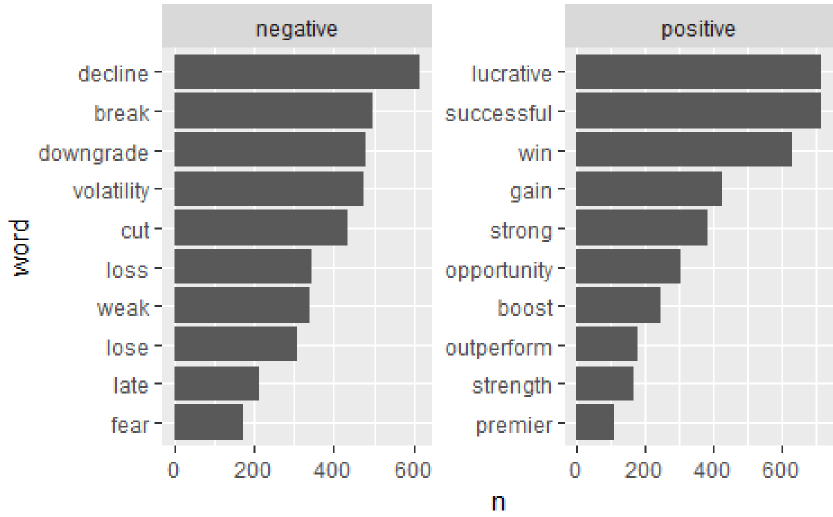

Figure 2: The most frequent words

expression that is mainly used in spoken language. Another example is the term *roar* that is used in combination with economy which is a metaphocial way to say that the economy is expanding, thus, it is added as a *positive* word.

Figure 2 shows the ten most used words either in the negative category, a bearish sentiment or a positive category, a bullish sentiment that are identified by the dictionary approach. Words like *decline* or *break* are used the most to signify bearish sentiment whereas words like *lucrative* or *succesful* are used the most to signify bullish sentiment.

Still, the dictionary approach even if being effective in sentiment classification, it still cannot capture the whole diversity of a language. This deficit is attempted to the minimised via the training of a Naive Bayes algorithm that does not consider any external dictionaries but base the categorisation given conditional probabilities of data at hand. Even though, Naive Bayes algorithm is "the oldest" of the algorithms used to classify documents, it is also the most successful natural language algorithm in doing so (Antweiler & Frank (2004), Liu (2015) and See-To & Yang (2017)).

Jurafsky & Martin (2017) formulate in equation 7 that the Naive Bayes is a probabilistic classifier which means that for the document d out of all classes $c \in C$ returns the class $\hat{c}$ which has the maximum posterior probability given the document, so "our estimate of the correct class".

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(c|d) \tag{7}$$

Thus applying Bayes' rule to the equation 7 leads to

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} \frac{P(d|c)P(c)}{P(d)}. \tag{8}$$

Since the document probability $P(d)$ does not change for each class, equation 8 can be simplified to

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(d|c)P(c) \tag{9}$$

where $P(d|c)$ is the likelihood of a document given the class and $P(c)$ is the prior probability of the class. Thus, it leads to

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} P(f_1, f_2, ..., f_n|c)P(c). \tag{10}$$

However, the computation of the likelihood in such a representation is cumbersome which is why the so-called naive Bayes assumption is made to reduce the complexity. The naive Bayes assumption assumes that the probabilities $P(f_i|c)$ are independent of each other which leads to the following representation:

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{f \in F} P(f|c). \tag{11}$$

Hence, walking through each word (w_i) in a sentence means that it does not dependent on the preceding or the following word. It is formulated as

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{i \in positions} P(w_i|c). \tag{12}$$

and expressed in the log space leads to

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) + \sum_{i \in postions} logP(w_i|c) \tag{13}$$

where

$$P(c) = \frac{N_c}{N_{doc}} \tag{14}$$

is the fraction of classified documents N_c in documents N_{doc} and

$$P(w_i|c) = \frac{count(w_i, c) + 1}{\sum_{w \in V} count(w, c) + |V|} \tag{15}$$

is the fraction of each word in a category in all documents where V is the vocabulary that exists in all classes. In addition, Laplace smoothing is applied in 15 to elude

Twitter feed	Category
Initial #FTSE supports at 7625 & 7590 but I believe the US peak has turned so downward force shold be powerful https://t.co/GBRehPMkT	negative
RT @techniquant: $FTSE ENDS THE DAY ON A BEARISH NOTE CLOSING NEAR THE LOW OF THE DAY: https://t.co/fQBrUfwtlb	neagtive
As I was saying <U+001F9D0> *DJI*NDX *SPX*FTSE $DAX https://t.co/UCeK0jLCAq	neutral
$FTSE Here is latest updated 1 hour chart from London update presented to clients @ https://t.co/c2fCVssy9t #FTSE https://t.co/91SGv1ZdV	neutral
Watch the Video on $FTSE: Next Bullish Cycle Has Started https://t.co/G1ysC4HHfW #elliotwave	positive
$FTSE How we saw it back in January 2018. It pulled back in wave "b" as expected and have already made new hights https://t.co/eGQ8SRdFh	positive

Table 3: Tweets categorisation examples

errors.

Still, tweets come as uncategorised data which is the challenging part to identify labelled data such that the algorithm works. Antweiler & Frank (2004) and Sprenger et al. (2014) manually labelled data as being either positive, negative or neutral. Following their approach, 1060 randomly picked tweets are labelled manually according to the same categories. Table 3 shows some example of the categorisation. In total, 20.5 % of the documents are labelled as negative, 30.8 % as positive and 48.7 % as neutral. After cleaning the data, it is not tokenized but represented as a document term matrix where rows depict all documents, columns each words that is used in the document and cells present the frequencies of a word that is included in the document. This representation is used to more easily assess word frequencies. Moreover, words that occur more than five times are included into the evaluation to obtain more secure results. The prediction of the Naive Bayes classifier is assessed with a confusion matrix. To predict categories, data is partitioned into in to training and testing data set of 70 % and 30 % and cross validated with a 10-fold cross validation technique.

Applying the algorithm in the means of Jurafsky & Martin (2017), the fourth table shows that the classifier predicts the testing test given the training set with

	Reference		
Prediction	negative	neutral	positive
negative	25	25	31
neutral	8	94	11
positive	15	19	60

	Class: negative	neutral	positiv
Sensitivity	0.521	0.681	0.588
Specificity	0.767	0.873	0.817
Accuracy	0.621		

Table 4: Confusion matrix for manually labelled tweets

	Reference		
Prediction	negative	neutral	positive
negative	11	48	18
neutral	27	69	66
positive	10	21	18

	Class: positive	neutral	negative
Sensitivity	0.229	0.5	0.177
Specificity	0.725	0.38	0.833
Accuracy	0.34		

Table 5: Confusion matrix with 10-fold for manually labelled tweets

Prediction	Reference negative	positive
negative	1975	171
positive	638	2228
Sensitivity	0.756	
Specificity	0.929	
Accuracy	0.839	

Table 6: Confusion matrix for proxied categorisation

Prediction	Reference negative	positive
negative	1004	973
positive	1609	1426
Sensitivity	0.384	
Specificity	0.594	
Accuracy	0.485	

Table 7: Confusion matrix with 10-fold for proxied categorisation

an overall accuracy of 62 %. Predicting the right individual category can, though, only be achieved with 52 % for the positive category, 68 % for the neutral category and 59 % for the negative category. The individual categories are correctly not predicted in 77 %, 88 % and 82 % of the time. Thus, these results are fair enough, unfortunately though, still not satisfactory for a classifying performance.

Cross validation is achieved via larger training of the classifier. The training set is subdivided into ten equally large sets of which nine sets are trained on the tenth with ten iterations. On average 70 % of tweets are well categorised. However, when applying the classifier to the testing set the overall accuracy shrinks from 62 % to 34 %. Also the sensibility and the specificity diminish as shown in the fifth table. Thus, these randomly picked data poorly classify the categories which should not be used to classify tweets.

Checkley et al. (2017) use another method to classify their Naive Bayes, they download Twitter data with the words *bullish* and *bearish* to represent positive and negative sentiment and then successfully train a Naive Bayes. Following their idea, 10,000 tweets for each category are downloaded from Twitter. However, as shown in the seventh table, results are not satisfying as well. After the cross validation, the overall accuracy decreases to 49 % from 84 %, sensitivity from 75 % to 38 % and specificity from 93 % to 60 %.

Due to that poor performance of the Naive Bayes in the two cases, only the dictionary approach is used in the following evaluations.

3.4 Sentiment calculation

Finally, sentiment is defined in three different ways as given in equation 16 to 18 following the specifications of Oliveira et al. (2013).

$$BIND_t = log(\frac{1 + Bull_t}{1 + Bear_t}) \tag{16}$$

$$TIS_t = \frac{Bull_t + 1}{Bull_t + Bear_t + 1} \tag{17}$$

$$RTIS_t = \frac{TIS_t}{TIS_{t-1}} \tag{18}$$

The bullishness indicator $BIND_t$ is calculated as the log of the ratio between the bullish sentiment and the bearish sentiment. Specifically, $Bull_t$ represents the counted categories of tweets being positive on the day t and $Bear_t$ represents the counted categories of tweets being negative on the day t. Adding one to either the $Bull$ or the $Bear$ prevents the equation not to be unsolvable if no sentiment is counted on the day for either of the categories. The Twitter Investor Sentiment TIS_t is counted as the fraction of bullish sentiment from the whole sentiment at each day during the week. The ratio of the Twitter Investor Sentiment $RTIS_t$ aims to measure intraday shifts in the sentiment, thus, representing the change in sentiment waves.

Specifically, for the bullishness indicator holds if $BIND_t$ is above zero than bullish investor sentiment is present, whereas if $bind_t$ is smaller than 0 than the investor sentiment is bearish. If the sentiment equals to zero than the investor are indifferent in their beliefs about the performance about the market at the specific point in time. The same holds true for the TIS with the exception that the threshold lies at 0.5 and not zero. The change in the sentiment is shown

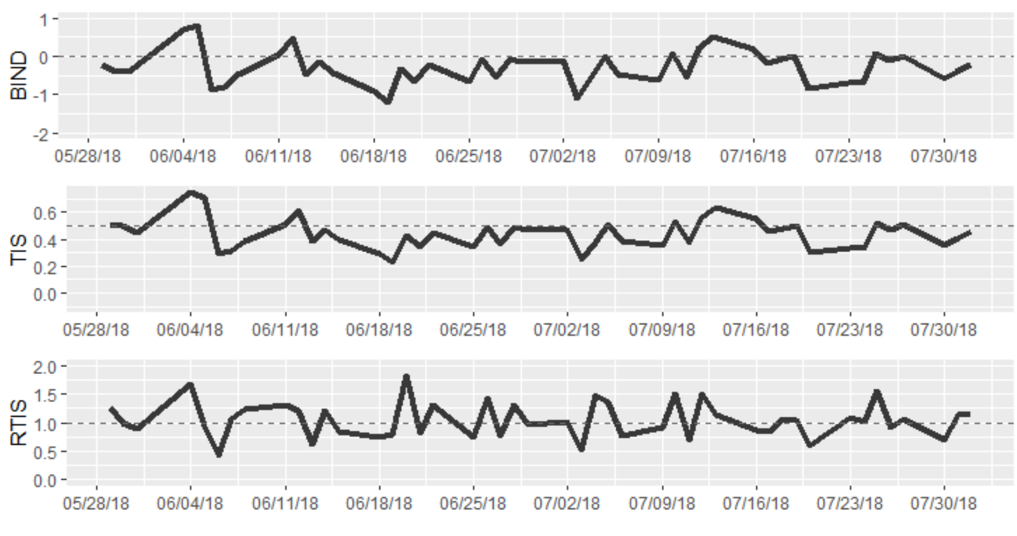

Figure 3: The sentiment in the British market

through the deviation from one. If market participants have common beliefs about the market than sentiment does not change which is true at the threshold of one. Deviating from one means that market participants have different understanding about market's performance.

Figure 3 depicts investors' sentiment that is retrieved from Twitter. The two measures either as $BIND_t$ or as TIS_t show nearly the same sentiment development during the whole period. According to the specification in 16 and 17 the British market experiences more bearish beliefs about the FTSE 100 than bullish one. Especially, during the first two weeks the sentiment oscillates between bullish and bearish sentiment. From the third to the fourth week, sentiment stays slightly bearish. From the fourth to the six week sentiment is even less bearish which can be seen as indifferent. From the sixth till the eighth week sentiment oscillates in small bounces either displaying a bullish or a bearish sentiment. The $RTIS$ depicts the constant change in sentiment within the period. However, at the end of the fourth as well as at the end sixth week the same sentiment beliefs persist.

4 Other sentiment measures

The question that arises is whether the above derived sentiment from Twitter are well performing or not. To identify the relation between the Twitter sentiments and the stock performance, other market variables that are commonly viewed as sentiment proxies are considerate and compared to in the next section. Like in Feldman (2010) and Simon & Wiggins (2001) the put-call ratio as well as the trading index, ARMS index, is considered and like in Smales (2015), Mao et al. (2011) implied volatility index is also taken into account. The data for these three time series is retrieved from the data provider Bloomberg for the same period as for the constructed sentiment form Twitter.

The put-call ratio, abbreviated as RCR, is derived from option trading and equals to the total trading volume of puts divided by the total trading volume of calls. Particularly, the $UKXPCG$ is used which is the put-call ratio of the FTSE 100. Simon & Wiggins (2001) explain that market participants view the put-call ratio as a fear indicator where levels higher than one reflect bearish sentiment and level lower than one reflect bullish sentiment. They reason that when market participants become more and more bearish, they tend to buy puts either to hedge their portfolios or to make bearish bets which is depicted through higher put-call ratios. The lower the put-call ratio, the lower is the demand for puts. Thus, the

	Mean	Sd	Min	Max	Skew	Kurtosis
PCR	1.50	0.08	1.34	1.61	-0.69	-0.7
ARMS	1.36	0.91	0.21	3.86	1.02	0.09
VFTSE	12.52	1.39	9.75	15.93	0.69	-0.20
BIND	-0.29	0.45	-1.2	0.79	0.29	-0.21
TSI	0.44	0.11	0.23	0.75	0.52	0.18
RTIS	1.05	0.32	0.41	1.84	0.36	-0.47

Table 8: Descriptive statistics

put-call ratio is driven by demand rather than supply.

The trading index, abbreviated as ARMS, equals to the number of advancing stocks scaled by the trading volume of advancing stocks divided by the number of declining stocks scaled by the trading volume of declining issues (Simon & Wiggins (2001)). The trading index tends to be high when the number of advancing stocks is low relative to the number of declining stocks. Thus, representing a bearish sentiment. Simon & Wiggins (2001) explain that advancing stocks go along with lower trading volume of advancing issues relative to the trading volume of declining issues which means that the market participants sell their stocks. Bullish sentiment is observed when the trading index tends to be low. Specifically, the $UKXARMS$ index is downloaded.

Volatility indices that use implied option volatilities information are sometimes referred to as "the investor fear gauge" (Areal (2008)), meaning the higher the index the greater the fear of investors is. These type of volatilities indices represent the market consensus on future stock market volatility. By construction they are forward-looking and have a constant forecast horizon (Areal (2008)). In particular, the volatility index FTSE 100, shortly $VFTSE$ is used.

Table 8 depicts the descriptive statistics of the different sentiment measures. During the whole period the mean on the put-call ratio lies at 1.5 reflecting bearish sentiment towards the FTSE 100. Moreover, the standard deviation changes slightly with the minimum and the maximum ranging only in the bearish region. The trading index also indicates on average bearish sentiment. However, according to the minimum and maximum, sentiment varies between bullish and bearish. The bullish sentiment as well as the Twitter investor sentiment also display on average bearish sentiment during the observed period. Like the trading index, though, they do vary in sentiment polarity as depicted in figure 2. Even though, Twitter sentiment waves change as shown by the RTIS, on average there is only a slight diversion in beliefs of FTSE's behaviour as the mean is only 1.05. The implied volatility is difficult to interpret as on average the VFTSE is 12.52. Therefore, investors exhibit fear about the market when being above the mean and are more

	PCR	ARMS	VFTSE	BIND	TIS	RTIS
PCR	1					
ARMS	-0.234*	1				
VFTSE	-0.013	-0.100	1			
BIND	-0.259*	0.086	-0.090	1		
TIS	-0.357*	0.101	-0.043	0.983*	1	
RTIS	-0.053	-0.081	-0.033	0.561*	0.562*	1

*Significance at the 5 % level

Table 9: Correlation between sentiment measures

confident when market is beneath it.

As shown in the descriptive statistics, sentiment measures indicate a common belief that the British market is bearish on average during the observed time. Hence, do these different measures also correlate with each other. Table 9 shows the pairwise Pearson correlation. The different measures correlate with each other only to small degree. The put-call ratio is negatively correlated with all the other sentiment. This relation is only expected for the bullish as well as the Twitter investor sentiment. Since an increase in put-call ratio would mean to become more bearish which is reflected by the negative sign of the BIND and TSI. This relation is the strongest compared to other measures and significant at 5 % level. Put-call ratio is also significantly related to the trading index, however, they lead to diverging implications. A more bearish sentiment from the put-call ratio would lead to a bullish sentiment from the ARMS index. The ARMS variable does not correlate significantly with any sentiment measures except for the put-call ratio. It is positively correlated with the BIND and TIS directing to opposing sentiments. When ARMS indicate more bearish sentiment, the Twitter sentiment measures indicate bullish sentiment. It is negatively correlated to the implied volatility and the change in the Twitter sentiment. Overall, the correlation is low. The implied volatility correlates negatively the lowest with all the other measures at no significance level. The negative sign is expected for the Twitter sentiments because the higher the fear about the market means becoming more bearish in the Twitter sentiment whereas it does not hold true for the other measures. The highest and significant relation at 5 % is found between the BIND and the TIS index. Since both measures display the same movement only BIND is evaluated in the following section. Moreover, the pairwise Pearson measure identifies positive and significant relationship to the change in the Twitter sentiment for BIND as well as TIS.

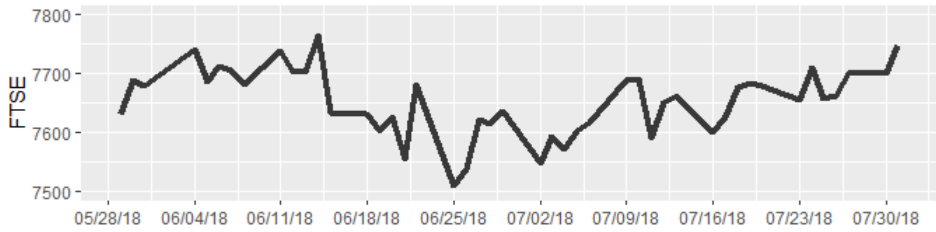

Figure 4: Price development FTSE 100

5 Financial data

The different sentiment measures are set into relationship with FTSE 100's performance. In particular, following Checkley et al. (2017), Oliveira et al. (2013), Mao et al. (2011), See-To & Yang (2017) and Bollen et al. (2011) performance is defined as the FTSE 100 return and its realised volatility. Data of FTSE 100's price development is retrieved at daily closing price as well as at five minutes frequency from the data platform Bloomberg at the same time range as for the sentiments.

FTSE 100's price development during the time period is depicted in figure 4 that displays an erratic price movement. Nevertheless, two trends can be observed, a declining trend starting from the second till the fourth week where the lowest price is reached on the 25.06.2018 and an upward trend from the fourth week until the end of the observation period. Checking the price development for stationary by applying Dickey Fuller test with a drift and trend, it results in non-stationarity as shown in the appendix A. To circumvent potential problems that might go along with that observation, first differencing is applied to the data, leading to the returns of the FTSE 100 which are stationary (see appendix A).

Particularly, returns are calculated as following

$$r_t = log(F_t) - log(F_{t-1}) \tag{19}$$

where F_t is FTSE's price at time t and r_t is FTSE's return at time t. Moreover, returns are standardised to make results between sentiments measures more comparable.

The second stock performance measure being the realised volatility is calculated as in equation 20. In general, volatility measures equity risk. A simple volatility measure is the standard deviation, however, to obtain a more precise ex-post estimator, the realised volatility RV_t is used which is the summation of intra-day squared returns $r^2_{t,i}$ (Bauwens et al. (2012)). M the amount of high frequency data. Since the given data is at five minutes frequency M equals to 102 for the 8.5 trading hours of the London Stock Exchange.

$$RV_t = \sqrt{\sum_{i=1}^{M} r_{t,i}^2} \tag{20}$$

6 Analysis

To understand the relation between the sentiment measures and the stock performance represented either as the return or the realised volatility, Granger causality analysis is applied and in the next step, stock performances are predicted with sentiment measures and compared to the baseline equation without sentiment. Finally, with the help of the error metrics the predictability of the sentiment specifications is evaluated like in Mao et al. (2011), Bollen et al. (2011), Checkley et al. (2017) and See-To & Yang (2017).

6.1 Granger causality

Granger causality test is a statistical hypothesis test to determine whether a time series X is useful in forecasting another time series Y. If the null hypothesis is rejected than X does not help to predict or Granger cause Y. The alternative hypothesis is adding X does help predict Y (Mao et al. (2011)) Therefore, a variable X is said to "Granger cause Y" if Y can be better predicted using the histories of both X and Y than by using the history of Y alone (Bollen et al. (2011)). However, Granger cause does not mean causation it is rather a correlation relationship that is established between the two time series. Like in Bollen et al. (2011) or Checkley et al. (2017), this thesis does not attempt to test for causality but rather for the predictive information of the time series.

Specifically, the tested Granger causality test can be expressed as the following

$$Y_t = \sum_{i=1}^{m} \beta_i Y_{t-i} + \sum_{i=1}^{m} \delta_i X_{t-i} + \epsilon_t \tag{21}$$

where the hypothesised relation is that variable X Granger causes Y as it brings additional information to the series. In the first case, X represents the sentiment that Granger causes Y the returns or the realised volatility. However, news can react more to the swings in returns rather than them causing the returns swings as

	p-values		p-values
Returns → PCR	0.304	PCR → returns	0.829
Returns → ARMS	0.897	ARMS → returns	0.64
Returns → VFTSE	0.03*	VFTSE → returns	0.256
Returns → BIND	0.32	BIND → returns	0.336
Returns → RTRIS	0.689	RTIS → returns	0.124

*at 5 % level

Table 10: Granger causality for returns

indicated by Tetlock (2007). Therefore, this observation is checked as well. Thus, in the second case, X is either the return or the realised volatility that Granger causes Y the sentiment.

To conduct Granger causality, the endogenous relationship between returns and sentiment and realised volatility and sentiment is each specified by a vector autoregressive model where the length of lags is identified with the help of selection model criteria like Akaike information criteria (AIC), Schwarz criterion (SC), also known as Bayesian information criteria or Hannan-Quinn information criteria (HQ). From the identified specification error terms are checked for being autocorrelated or not. Specifically, the Portmanteau test as well as the Breusch-Godfrey test are run where for both cases the null hypothesis of no serial correlation of the error term is checked against the alternative hypothesis of error term being autocorrelated. All the results are provided in the appendix B and C.

The VAR selection model identifies VAR(1) for the endogenous relation between FTSE's returns and all the other sentiment measures with no autocorrelation in the error terms. For the endogenous relationship between the realised volatility and the different sentiment measures also VAR(1) with no autocorrelated error terms is identified. Only for the implied volatility relationship VAR(2) specification with no autocorrelation in the error term is identified.

Table 10 shows the p-values for the Granger causality test for the endogenous relation between the returns and the specified sentiment measures. As identified by the selection model, Granger causality is run for one lag. Thus, low p-value means that the null hypothesis is rejected whereas high value in p means that the null hypothesis cannot be rejected. The → signifies the Granger causality direction. Returns do not Granger cause the sentiment specification as the p-values are high with the exception for the implied volatility where the p-value is 0.03. Hence, price changes in the FTSE 100 do Granger cause the implied volatility significantly at the 5 % level. Thus, it indicates that investor's fear emerge in accordance to returns. Looking in the opposite direction whether sentiment measures Grange cause returns, p-value are all high such that there is no evidence that the senti-

	p-values		p-values
Realised volatility → PCR	0.619	PCR → realised volatility	0.911
Realised volatility → ARMS	0.28	ARMS → realised volatility	0.998
Realised volatility → VFTSE	0.003*	VFTSE → realised volatility	0.293
Realised volatility → BIND	0.842	BIND → realised volatility	0.679
Realised volatility → RTRIS	0.281	RTIS → realised volatility	0.528

*at 5 % level

Table 11: Granger causality for realised volatility

ment measures either proxied from the market variables or the constructed Twitter measures Granger cause returns.

Table 11 shows the p-values for the Granger causality test for the endogenous relation between the realised volatility and the specified sentiment measures. Like in the case for the returns, realised volatility does not Granger cause either the put-call ratio, the trading measure or one of the sentiment index from Twitter. However, realised volatility does Granger cause the implied volatility significantly at 5 % level. Moreover, none of the different sentiment measures Granger cause realised volatility.

These results show that there is no linear impact going from the sentiment direction to the FTSE 100's returns or the realised volatility. Only, a Granger causality relation is identified from the stock's performance to the implied volatility.

6.2 Forecasting

Even though, the Granger causality test did not find Granger causality relation between sentiment measures and stock performance, with the only exception for the implied volatility, still all of the sentiments are checked whether they can serve some prediction power to the movement of FTSE 100's returns or volatility. Therefore, a one-step ahead prediction is conducted and forecasting errors are checked where they are the lowest. Specifically, the data is divided into a training set, the estimation period of 43 days, and a testing set, the prediction period of five days.

In particular, two models as shown in equation 22 and 23 are compared with each other. $M0$ is an autoregressive model and $M1$ is augmented to sentiment specifications. Y represents stock performances, thus, the return and the realised volatility, X represents the different sentiment measures, β and γ are the equivalent weights, i is the specified lag as in the Granger causality so in nearly all cases it is one, except for the implied volatility for the realised returns where it has two lags. ϵ signifies the error term.

$$M0 : Y_t = \alpha + \sum_{i=1}^{j} \beta Y_{t-i} + \epsilon_t \tag{22}$$

$$M1 : Y_t = \alpha + \sum_{i=1}^{j} \beta_i Y_{t-i} + \sum_{i=1}^{j} \gamma_i X_{t-i} + \epsilon_t \tag{23}$$

To evaluate the quality of the prediction of the models in 22 and 23, two error metrics firstly, the Mean Absolute Percentage Error, measured as in 24 and, secondly, the Root Mean Squared Error, measured as in 25 are used. In both cases $\hat{y}$ represents the fitted values and y the targeted values. h is the prediction period, being maximal five days. The lower the value of RMSE and MAPE the better is the model (Oliveira et al. (2013)).

$$MAPE = \frac{1}{h} \sum_{t=T}^{T+h} | \frac{y_{t+h} - \hat{y}_{t+h}}{y_{t+h}} | * 100 \tag{24}$$

$$RMSE = \sqrt{\frac{1}{h} \sum_{t=T}^{T+h} (y_{t+h} - \hat{y}_{t+h})^2} \tag{25}$$

Results for both models are shown in the table 12. In the case Y_t being the returns, the baseline model outperforms the models that are augmented with sentiment measures because the error metrics measured either as MAPE or RMSE are the smallest. Only the specification for the change in the sentiment that is derived from Twitter, MAPE and RMSE are even smaller. This observation indicates even though no Granger causality was observed before and sentiment change does not enter any specification significantly at any critical level, there is still some information in the data that might help to predict returns. In the case where Y_t is the realised volatility, the baseline model outperforms the specification with put-call ratio, the trading index and the change in the sentiment. The implied volatility

Y_t = Returns	MAPE	RMSE	Y_t = Realized volatility	MAPE	RMSE
M0	8.982	0.417	M0	0.156	0.0827
M1 with PCR	9.759	0.435	M1 with PCR	0.159	0.082
M1 with ARMS	12.629	0.46	M1 with ARMS	0.158	0.083
M1 with VFTSE	12.737	0.482	M1 with VFTSE	**0.129**	**0.072**
M1 with BIND	10.764	0.424	M1 with BIND	**0.154**	0.088
M1 with RTIS	**7.586**	**0.346**	M1 with RTIS	0.171	0.09

Table 12: Accuracy measure for the forecasting MO and M1 model

measured either as MAPE or RMSE are smaller than the baseline, indicating that the implied volatility brings information to the realised volatility. Moreover, the MAPE measure is slightly smaller for the bullishness sentiment compared to the baseline equation. However, it is higher when estimated with RMSE.

All in all, even if the Granger causality test only found Granger causality relation between returns and implied volatility, there still seems to be some information in the change of the Twitter sentiment that lowers the error metrics. In the case for the realised volatility the implied volatility and to some small extend the bullishness indicator lowers the error metrics beneath the baseline. Notwithstanding, these results should be viewed critically and with doubts due to the model setting and the previous results which are outlined in the next section.

7 Critical issues

In this master thesis, sentiment analysis is conducted for the United Kingdom. Specifically, sentiment indices are derived from the advancing microblogging website Twitter and set into realtion with FTSE 100's returns and realised volatility. The results of the analysis, though, should be viewed with doubts due to several reasons.

Firstly, even though millions of people use the micro blogging website Twitter on a daily base, these users seem to pay less attention to the British market in financial terms. It is evidenced by the low results for the query $\$FTSE$ which is on average 25 times smaller than for the corresponding query $S\&P$ of the US market. The relatively small sample size from this search query would lead to problems in applying natural language processing techniques as these substantially reduce it while cleaning data.

Secondly, the attempt to enlarge the sample size by searching for all companies that are comprised in the FTSE 100 leads to further issues which undermine the validity of the results. For instance the cross sectional interdependences are left aside. Moreover, although, the sample size is increased it depends to the most degree on four outliers. Thus, potential beliefs about FTSE 100's performance are set in the most relation with four queries. However, it is doubtful that these four queries represent on average the sentiment about FTSE 100's performance.

At the same time, this insight shows that the British market is not much covered by the social media. As Checkley et al. (2017) conclude from their analysis the more an asset is covered by social media the more it will shift to social media opinions. This conclusion, though, raises issues in this analysis. If the British market is not well covered by social media can it be influenced by sentiment shifts to any extend? The less attention is paid to a market in terms of social media the smaller is the proportion of individual investors who trade on such platforms and, thus, the smaller is the proportion of people who spreads sentiments and who pursues them. Therefore, the smaller is the likelihood that prices deviate from their fundamental prices. The Office for National Statistics of Great Britain publishes every second year a statistical bulletin where it conducts statistics on the ownership of UK quoted shares. The bulletin from 2017, states that in 2014 only 11.9 % of UK shares by value were owned by individuals whereas the other 34.3 % were held by domestic and 53.8 % by foreign institutional traders. Given this statistic, it is very questionable whether that low proportion of individual traders can impact the market severely through their feelings about the market. Due to that, it is questionable whether De Long et al. (1990)'s model holds true in all circumstances.

Thirdly, the option to retweet Twitter feeds shows that the opinion holder shares the same feelings towards the issue like the previous writer. However, as **?** explain retweeting is conducted in diverse ways with different kind of intentions. Some just copy the original tweet and retweet it whereas others for instance cut "unnecessary" worlds like adjectives or adverbs. These parts of sentences, though, are essential for the sentiment analysis when applying the dictionary approach or even Naive Bayes classifier. On top of that, the act of retweeting follows two opposing purposes. On the one hand, Twitter users simply want to share information and their inner conditions which is why they use Twitter as a "communication tool". On the other hand, Twitter users intend to use retweets strategically to increase their reputation which the authors call "a selfish act of attention seekers" while, in fact, being indifferent to the content of the retweet. Consequently, leaving out retweets from the sample sizes would mean to leave out common shared opinions but at the time when leaving them in questions rise how useful the provided information is.

Fourthly, classifying the polarity of Twitter feeds is a difficult challenge. The dictionary approach standardises the process of sentiment categorisation to a certain degree but at the same time it looses a lot of information. The Naive Bayes algorithm, though, needs to be optimised to the considerable data until results are satisfying for which great computational skill are needed.

Fifthly, even though, the error metrics are slightly low for the prediction of returns in the case of the change in the sentiment. It is very unlikely that this result can be reasoned by the sentiment itself. Considering, the first and the second critical point as well as the issue that no Granger causality relation was identified beforehand, this fact can be reasoned simply as a coincidence.

Sixthly, Y.-M. Li & Li (2013) account in their analysis for consumer opinions on products for credibility of the Twitter user by considering two factors, source and content of the tweet. In the first place, they aim to reduce the amount of spammers that use Twitter and assess opinion on products only from credible Twitter users. This raises, though, the question if social media sentiment has the power to drive markets to some extend, cannot market be manipulated easily via purposed posting activity of fake accounts that cannot be identified as fake.

Considering all these critical points, the conducted sentiment analysis on the British market for short time period from 28.05.2018 until 31.07.2018 does not show any linear sentiment influence. There is still some chance that sentiment is related to returns or volatility in non-linear way but according to this analysis there is little evidence that sentiment plays a role in the United Kingdom and, therefore, questions the ideas of behavioural finance.

8 Conclusion

The efficient market hypothesis states that prices fully reflect available information on the market. As a consequence, predicting prices is impossible due to the fact that all information is already incorporated into the price setting. However, the field of behavioural finance questions the assumptions of the efficient market hypothesis and, thus, its results. A model developed by De Long et al. (1990) assumes a rational and an irrational investor who trade on the same market. The rational trader trades on the fundamentals of assets whereas the irrational one trades on sentiments which manifests in their beliefs about the price development, so they feel either bearish or bullish about assets' price development. These feelings, though, are unpredictable and cause prices to fluctuate heavily. As a consequence, rational traders pay attention to irrational traders and try to bet against them and not trade on fundamentals which turns out to be costly and risky.

Due to this model's results, several literature niches have evolved trying to measure the sentiment of irrational traders. In the most recent history, the fast growing social media is considered as real-time data provider from which sentiment measures are derived. Ambiguous results can be drawn from the literatures suggesting that sentiment measures can predict the returns and volatility of assets as well as suggesting that they cannot. Primarily, the US stock market has been analysed so far in the literature.

This master thesis conducts a sentiment analysis for the British market. It derives sentiment indices with the help of natural language processing methods from the micro blogging website Twitter which are described as the bullishness index and the change in the sentiment. On the one hand, the Loughran & McDonald dictionary is applied to a bi-gram representation of tweets and, on the other hand, a Naive Bayes classier is applied to identify the categories of sentiment polarity. Since the Naive Bayes classifier does not perform satisfying in categorising, the dictionary approach is used to develop the sentiment indices.

The developed sentiment indices from Twitter as well as other commonly used sentiment measures like put-call ratio, trading index (ARMS) and implied volatility measure are set into relation with FTSE 100's returns and realised volatility. Specifically, Granger causality analysis is conducted checking for both directions whether stock performance Granger causes sentiment or sentiment Granger causes stock performances. The only identified direction is that returns and volatility Granger causes implied volatility. In the next step, a one-step-ahead prediction is run and evaluated with the error metrics, RMSE and MAPE. It shows that the change in the sentiment leads to the lowest metrics in predicting the returns and implied volatility predicts volatility with the lowers error metrics. Even if, a positive relation is established between the sentiment and stock performances, they should be viewed critically due to the potential biases that arise from the low coverage of the British market by social media. Conclusively, no impact of sentiment in the British market is identified.

A Stationarity

Dickey-Fuller test	for FTSE 100's price		
	phi3	phi2	tau3
Value of test-statistic is:	-2.5715	2.8972	4.3415

Critical values for test statistics:			
	1pct	5pct	10pct
tau3	-4.15	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61
*phi3 = none , phi2 = drift, tau3 = trend			

Table 13: Stationarity test for the FTSE 100's prices

Dickey-Fuller test	for FTSE 100' returns		
	phi3	phi2	tau3
Value of test-statistic is:	-6.7087	15.0356	22.5436

Critical values for test statistics:			
	1pct	5pct	10pct
tau3	-4.15	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61
*phi3 = none, phi2 = drift, tau3 = trend			

Table 14: Stationarity test for FTSE 100's returns

B VAR selection for returns

VAR selection for return and PCR			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-7.19500079	-7.11888017	-6.91937305	-6.7255644	-6.6366111	-6.43381474	-6.58050991	-6.57459096	-6.49084826	-6.34796337
HQ(n)	-7.10314209	-6.96578233	-6.70503608	-6.44998829	-6.29979585	-6.03576036	-6.1212164	-6.05405831	-5.90907647	-5.70495245
SC(n)	-6.92564305	-6.66995061	-6.29087166	-5.91749118	-5.64896605	-5.26659786	-5.23372122	-5.04823044	-4.78491591	-4.4624592
FPE(n)	0.00075102	0.00081316	0.00100032	0.00123123	0.00137625	0.0017433	0.00157981	0.00169959	0.00202607	0.00264844

Autocorrelation test for the VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 25.846, df = 36, p-value = 0.8946	Chi-squared	= 38.309, df = 40, p-value = 0.5465

Table 15: VAR selection for returns and put-call ratio

VAR selection for return and ARMS			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-0.25263412	-0.10099203	0.07653833	0.3011422	0.3351582	0.3926535	0.4804264	0.6678496	0.5653713	0.2888068
HQ(n)	-0.16077542	0.05210581	0.2908753	0.5767183	0.6719735	0.7907079	0.9397199	1.1883823	1.1471431	0.9318177
SC(n)	0.01672362	0.34793754	0.70503972	1.1092155	1.3228033	1.5598704	1.8272151	2.1942102	2.2713037	2.174311
FPE(n)	0.77746715	0.90783223	1.09251166	1.3867561	1.4672313	1.6071976	1.8413208	2.3751824	2.3503415	2.0197752

Autocorrelation test for the VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 31.673, df = 36, p-value = 0.6745	Chi-squared	= 42.26, df = 40, p-value = 0.3736

Table 16: VAR selection for return and ARMS

VAR selection for return and VFTSE			
AIC(n)	HQ(n)	SC(n)	FPE(n)
2	2	1	2

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-0.784548	-0.9355499	-0.7785922	-0.6968709	-0.7373359	-0.7139945	-0.7438455	-0.7055225	-0.8989649	-0.8700153
HQ(n)	-0.6926893	-0.7824521	-0.5642553	-0.4212948	-0.4005207	-0.3159401	-0.284552	-0.1849898	-0.3171932	-0.2270043
SC(n)	-0.5151902	-0.4866204	-0.1500909	0.1112023	0.2503091	0.4532224	0.6029432	0.8208381	0.8069674	1.0154889
FPE(n)	0.456746	0.3940594	0.4645662	0.5111737	0.5020192	0.5314448	0.5412961	0.6015186	0.5434728	0.6339179

Autocorrelation test for the VAR(2)			
Portmanteau Test (asymptotic)		Breusch - Godrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 23.258 , df = 32, p-value = 0.87	Chi-squared	= 48.718, df = 40, p-value = 0.1623

Autocorrelation test for the VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 28.155 , df = 36, p-value = 0.8216	Chi-squared	= 37.476, df = 40, p-value = 0.5845

Table 17: VAR selection for returns and VFTSE

VAR selection for return and BIND			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-1.7921405	-1.7234136	-1.5498615	-1.3862318	-1.3506863	-1.3785897	-1.2239967	-1.0140144	-0.8228773	-0.60312913
HQ(n)	-1.7002818	-1.5703158	-1.3355245	-1.1106557	-1.0138711	-0.9805353	-0.7647032	-0.4934817	-0.2411056	0.03988179
SC(n)	-1.5227828	-1.274484	-0.9213601	-0.5781586	-0.3630413	-0.2113729	0.122792	0.5123461	0.883055	1.28237504
FPE(n)	0.1667566	0.1792243	0.2148274	0.2565564	0.2718602	0.2734186	0.3348944	0.4418478	0.5864382	0.82782813

Autocorrelation test for the VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 19.666, df = 36, p-value = 0.9877	Chi-squared	= 24.609, df = 40, p-value = 0.9734

Table 18: VAR selection for returns and BIND

VAR selection for return and RTIS			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-2.39777259	-2.2652083	-2.3063055	-2.1428683	-2.0976501	-1.935128	-1.7455686	-1.7406303	-1.560445	-1.3366718
HQ(n)	-2.30591389	-2.1121105	-2.0919685	-1.8672922	-1.7608349	-1.5370736	-1.2862751	-1.2200977	-0.9786732	-0.6936608
SC(n)	-2.12841485	-1.8162788	-1.6778041	-1.3347951	-1.1100051	-0.7679111	-0.3987799	-0.2142698	0.1454873	0.5488324
FPE(n)	0.09100395	0.1042554	0.1008255	0.1203871	0.1288081	0.1567208	0.1987889	0.2136522	0.2804792	0.397527

Autocorrelation test for the VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 27.816, df = 36, p-value = 0.8337	Chi-squared	= 31.032, df = 40, p-value = 0.8445

Table 19: VAR selection for returns and RTIS

C VAR selection for realised volatility

VAR selection for the realised volatility and PCR			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-1.18E+01	-1.17E+01	-1.15E+01	-1.13E+01	-1.12E+01	-1.10E+01	-1.10E+01	-1.10E+01	-1.10E+01	-1.11E+01
HQ(n)	-1.17E+01	-1.15E+01	-1.13E+01	-1.10E+01	-1.08E+01	-1.07E+01	-1.05E+01	-1.05E+01	-1.04E+01	-1.05E+01
SC(n)	-1.15E+01	-1.13E+01	-1.09E+01	-1.05E+01	-1.02E+01	-9.89E+00	-9.63E+00	-9.48E+00	-9.28E+00	-9.28E+00
FPE(n)	7.77E-06	8.31E-06	9.93E-06	1.24E-05	1.49E-05	1.71E-05	1.95E-05	2.02E-05	2.24E-05	2.09E-05

Autocorrelation test for VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 26.634, df = 36, p-value = 0.8722	Chi-squared	= 38.408, df = 40, p-value = 0.542

Table 20: VAR selection for realised volatility and put-call ratio

VAR selection for the realised volatility and ARMS			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-4.78E+00	-4.67E+00	-4.50E+00	-4.28E+00	-4.17E+00	-3.97E+00	-3.95E+00	-4.04E+00	-4.15536621	-4.16522854
HQ(n)	-4.69278963	-4.51716145	-4.28743109	-4.00234175	-3.8300528	-3.57350038	-3.49169877	-3.51894965	-3.57243977	-3.52094142
SC(n)	-4.51819955	-4.22617798	-3.88005424	-3.47857151	-3.18988918	-2.81694337	-2.61874837	-2.52960587	-2.4667026	-2.29881086
FPE(n)	0.00836259	0.00940393	0.01120648	0.01419522	0.01618769	0.02028767	0.02163143	0.02103372	0.02036364	0.02253603

Autocorrelation test for VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 31.46.4, df = 36, p-value = 0.6842	Chi-squared	= 46.124, df = 40, p-value = 0.2338

Table 21: VAR selection for realised volatility and ARMS

VAR selection for the realised volatility and VFTSE			
AIC(n)	HQ(n)	SC(n)	FPE(n)
2	2	2	2

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-4.88856141	-5.13525347	-4.99957309	-4.858637	-4.96705477	-4.95252172	-4.84804235	-4.66169039	-4.64467141	-5.03059609
HQ(n)	-4.79652039	-4.98185178	-4.78481072	-4.58251394	-4.62957104	-4.55367731	-4.38783727	-4.14012463	-4.06174497	-4.38630897
SC(n)	-4.62193031	-4.69086831	-4.37743387	-4.05874371	-3.98940741	-3.7971203	-3.51488687	-3.15078085	-2.9560078	-3.16417842
FPE(n)	0.00753861	0.00590877	0.00681491	0.00794651	0.0072771	0.00761284	0.00882871	0.01130168	0.01248398	0.00948534

Autocorrelation test for VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 33.631, df = 32, p-value = 0.3884	Chi-squared	= 52.843, df = 40, p-value = 0.084

Table 22: VAR selection for realised volatility and FVTSE

VAR selection for the realised volatility and BIND			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-6.42780522	-6.34931423	-6.13891844	-5.94190433	-5.92164127	-6.03467268	-6.10985724	-5.9647921	-5.85872401	-5.77866331
HQ(n)	-6.3357642	-6.19591253	-5.92415607	-5.66578128	-5.58415754	-5.63582827	-5.64965216	-5.44322634	-5.27579757	-5.13437619
SC(n)	-6.16117412	-5.90492907	-5.51677922	-5.14201104	-4.94399392	-4.87927126	-4.77670176	-4.45388256	-4.1700604	-3.91224564
FPE(n)	0.00161736	0.00175484	0.00218096	0.0026898	0.00280148	0.00257973	0.00249976	0.00307053	0.00370763	0.00448923

Autocorrelation test for VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 33.631, df = 32, p-value = 0.3884	Chi-squared	= 52.843, df = 40, p-value = 0.084

Table 23: VAR selection for realised volatility and BIND

VAR selection for the realised volatility and RTIS			
AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-7.07258212	-6.92783516	-6.84238162	-6.63204913	-6.66356861	-6.622118	-6.44699715	-6.3644232	-6.34893823	-6.56621378
HQ(n)	-6.9805411	-6.77443347	-6.62761925	-6.35592608	-6.32608488	-6.22327359	-5.98679206	-5.84285743	-5.76601179	-5.92192666
SC(n)	-6.80595102	-6.48345	-6.2202424	-5.83215584	-5.68592125	-5.46671658	-5.11384166	-4.85351365	-4.66027462	-4.6997961
FPE(n)	0.00084876	0.00098399	0.00107929	0.00134894	0.00133405	0.00143367	0.00178435	0.002059	0.00227091	0.00204241

Autocorrelation test for VAR(1)			
Portmanteau Test (asymptotic)		Breusch - Godfrey LM Test	
data:	Residuals of VAR object	data:	Residuals of VAR object
Chi-squared	= 29.884, df = 36, p-value = 0.7538	Chi-squared	= 44.687, df = 40, p-value = 0.2814

Table 24: VAR selection for realised volatility and RTIS

References

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*(VOL. LIX, No. 3).

Areal, N. (2008). Ftse-100 implied volatility index.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross section of stock returns. *The Journal of Finance*.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*(2), 129–151.

Baker, M., Wurgler, J., & Yuan, Y. (2010). Global, local, and contagious investor sentiment. *Journal of Financial Economics*.

Bauwens, L., Hafner, C. M., & Laurent, S. (2012). *Handbook of volatility models and their applications* (Vol. 3). John Wiley & Sons.

Beer, F., & Zouaoui, M. (2013). Measuring stock market investor sentiment. *Journal of Applied Business Research*, *29*(1), 51.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. doi: 10.1016/j.jocs.2010.12.007

Bormann, S. (2013). Sentiment indices on financial markets: What do they measure? *Economics Discussion Papers*.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1–27. doi: 10.1016/j.jempfin.2002.12.001

Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *The Journal of Business*, *78*(2), 405–440. doi: 10.1086/427633

Checkley, M. S., Higón, D. A., & Alles, H. (2017). The hasty wisdom of the mob: How market sentiment predicts stock market behavior. *Expert Systems with Applications*, *77*, 256–263. doi: 10.1016/j.eswa.2017.01.029

Da, Z., Engelberg, J., & Gao, P. (2014). The sum of all fears investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1–32.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*(Vol. 98, No. 4), 703–738.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *73*(Vol. 25 No.2), 383–417. doi: 10.1111/jofi.12608

Feldman, T. (2010). A more predictive index of market sentiment. *Journal of Behavioral Finance*, *11*(4), 211–223. doi: 10.1080/15427560.2010.526892

Finter, P., Niessen-Ruenzi, A., & Ruenzi, S. (2012). The impact of investor sentiment on the german stock market. *Zeitschrift für Betriebswirtschaft*, *82*(2), 133–163. doi: 10.1007/s11573-011-0536-x

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*(Vol. LXVIII, No. 3).

Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, *18*(2), 321–340. doi: 10.1016/j.jempfin.2010.11.009

Jurafsky, D., & Martin, J. H. (2017). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd edition ed.). [India]: Dorling Kindersley Pvt, Ltd.

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826–840. doi: 10.1016/j.ins.2014.03.096

Li, Y.-M., & Li, T.-Y. (2013). Deriving market intelligence from microblogs. *Decision Support Systems*, *55*(1), 206–217. doi: 10.1016/j.dss.2013.01.023

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions.* Cambridge University Press.

Loughran, T., & McDonald, B. (2015). The use of word lists in textual analysis. *Journal of Behavioral Finance*, *16*(1), 1–11. doi: 10.1080/15427560.2015.1000335

Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. *arXiv preprint arXiv:1112.1051*.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611. doi: 10.1016/j.eswa.2015.07.052

Oliveira, N., Cortez, P., & Areal, N. (2013). On the predictability of stock market behavior using stocktwits sentiment and posting volume. In D. Hutchison et al. (Eds.), *Progress in artificial intelligence* (Vol. 8154, pp. 355–365). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-40669-0_31

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, *53*(3), 458–464. doi: 10.1016/j.dss.2012.03.001

See-To, E. W. K., & Yang, Y. (2017). Market sentiment dispersion and its effects on stock return and volatility. *Electronic Markets*, *27*(3), 283–296. doi: 10.1007/s12525-017-0254-5

Shen, J., Yu, J., & Zhao, S. (2017). Investor sentiment and economic forces. *Journal of Monetary Economics*, *86*, 1–21. doi: 10.1016/j.jmoneco.2017.01.001

Silge, J., & Robinson, D. (2017). *Text mining with r: A tidy approach.* " O'Reilly Media, Inc.".

Simon, D. P., & Wiggins, R. A. (2001). S&p futures returns and contrary sentiment indicators. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *21*(5), 447–462.

Smales, L. A. (2015). Time-variation in the impact of news sentiment. *International Review of Financial Analysis*, *37*, 40–50. doi: 10.1016/j.irfa.2014.11.019

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957. doi: 10.1111/j.1468-036X.2013.12007.x

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*(Vol. LXII, No. 3).

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies*, *24*(5), 1481–1512. doi: 10.1093/rfs/hhq141

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. doi: 10.1016/j.dss.2012.12.028

Yuan, Y. (2015). Market-wide attention, trading, and stock returns. *Journal of Financial Economics*, *116*(3), 548–564. doi: 10.1016/j.jfineco.2015.03.006