VRIJE UNIVERSITEIT AMSTERDAM

BACHELOR THESIS

The Effect of Social Media Sentiment on the European Stock Market

Author:
Mark de Kwaasteniet

2649271
M.G.de.Kwaasteniet@student.vu.nl

Supervisor:
Lowie Dangermond

Amsterdam, June 25, 2021

Contents

1 Abstract

This study is interested in examining investor sentiment about European indices and using this sentiment to determine its effect on the returns and trading volume of the index. Investors are constantly looking at new ways to beat the market, with various studies supporting the claim that investor sentiment can predict future market movements. This study uses VAR models, Granger causality tests, and Impulse Response Functions to determine whether this effect of investor sentiment is consistent on the European index market.

This study concludes that there is no effect of social media sentiment, about the European indices, on the returns and trading volume of the European indices. This study accepts the null hypothesis, indicating that it is impossible to study investor sentiment about European indices on Twitter to predict future returns or trading volume of European indices. The findings cannot reject the traditional asset pricing theories and their assumptions, implying that even though there might be anomalies in the European market, like size, market-to-book, and momentum effects, investor sentiment is not likely to be one in the European index market. However, the study can reject the weak form of the efficient market hypothesis and the random walk theory, as multiple VAR models indicate autocorrelation.

In contrast to the expectations, this study finds that index returns and trading volume significantly influence the sentiment measured among investors. These are insignificant findings in the field of finance, as they will not directly result in increased returns on the European index market. These findings implicate that investment strategies, including investor sentiment, cannot outperform the market, as investor sentiment holds no predictability over future price movements or trading volume of the European indices. However, since the study has several limitations affecting the findings, future research needs to tackle these deficits, to entirely exclude the effect of investor sentiment on the returns and trading volume of the European indices.

2 Introduction

"Dumb Money Is on GameStop, and It is Beating Wall Street at Its Own Game" was the New York Times headline on January 27, 2021 (Phillips and Lorenz, 2021). It was one of many about the "David and Goliath story" between the individual investors buying and Institutional Investment companies shorting GameStop stock. Individual investors were able to lift the price of GameStop's stock far above its fundamental value by collectively buying the stock on the market. The leading cause of this unexpected individual investor movement was shared social media sentiment, simultaneously believing in and discussing GameStop. Social media has laid a foundation for small individual investors to compete with the "big guys". Therefore, it is essential to examine the effects of these social media developments on the stock market to determine the relevancy of traditional stock market theories. Do they still hold or have trading conditions changed to such an extent that they need to be adjusted?

Retail investors are considered individual investors or traders who participate in the stock market on their behalf instead of investors who work for institutions like pension, mutual, or hedge funds. Even though retail trading has long been possible, online retail trading became increasingly popular in the last decade (FinSMES, 2019). Showing a significant peak in online trading activity at the start of 2020. The AFM argues that there have been two primary macroeconomic factors before this peak, i.e., low-interest rates and the outbreak of COVID-19 (Autoriteit Financiële Markten, 2020). Studies show that these factors cause an increase of participants and general activity in the stock market as people will have relatively more money to spend and more time to spend it, considering the unemployment rate that increased due to COVID-19 (Chordia et al., 2001; Talwar et al., 2021). This visible shift has been made possible because of the functionality of the internet, which increased the transparency and accessibility of the stock market, making it a common ground for both institutional investors and the new retail investors and simultaneously creating some changes regarding the process of trading.

Considering the origin of retail trading, it is safe to say that traditional retail trading was fairly complicated. Every step had to be taken manually and offline, like looking up companies in the paper and calling brokers to invest in certain companies. This type of trading meant a lack of transparency in the market and created a barrier for retail investors who did not want to pursue such a time-consuming practice. This resulted in general retail investors willing to make this cumbersome effort to understand the market and its investing options. Currently, the entry barrier for retail investors is much lower due to technological innovations, like the internet and the introduction of smartphones. People can buy stocks on margin with their smartphone, their +18 identification, and some spare money (Robinhood, 2021). The trading platforms use gamification, making an activity look like a game to stimulate ease, excitement, and activity in trading. These innovations regarding these new trading methods have the following results for the stock market conditions.

The stock market has increased in transparency, and because of this increased transparency, it is has become easier to observe the behaviour of stocks and investors in the market. Not only is it possible to track stock prices, volatility, and trading volume per second, to improve trading strategies or find anomalies in the market. There are also market participants that are tracking the behaviour of investors in terms of actions and emotional states. Tracking the emotional states of investors has become increasingly accurate as the internet is a virtual place where everyone can share their emotions and thoughts.

Taking these market changes into consideration, there is a possibility of observing investors and examining their emotions to determine the future actions of these investors. By knowing investors' emotions, it does not matter if the investor is rational or irrational; if studies can predict the actions of investors based on the emotions they express, it will enable the predictability of future demand for a particular stock. The idea behind this course of the reason is the following: for a stock to increase in price, investors have to start believing that the stock has more value than the current price, increasing the demand and thus the price. Similarly, people need to stop believing in a particular company or stock to demand and thus the price to decrease. Therefore, investors with favourable emotions towards a stock also referred to as positive sentiment, show some belief in that stock. If investors are expressing emotions on the internet before purchasing stock, then by knowing the behaviour associated with positive/negative investor sentiment, the stock prices will be predictable. Therefore, this study will determine whether investor sentiment on social media can explain price fluctuations in the stock market.

Even though researchers have already done extensive work in this field, these studies mainly focus on the U.S. stock market. This is partly because the U.S. stock exchanges have a larger market capitalization than the European[1] stock exchanges (Winck, 2020). This begs the question of whether this significant relationship between investor sentiment on social media and stock returns and trading volume, which these studies have found (Bollen et al., 2011; Rao et al., 2012), is also detectable on the European stock market.

Therefore, to partly fill this empirical gap, this study aims to determine whether investor sentiment has a significant effect on the returns and trading volume of the European indices, using messages about the European indices on Twitter as a proxy for investor sentiment on social media. The main research question in this study is an all-encompassing question that is answered through multiple sub-questions. The main research question is as follows

> **RQ:** *What is the effect of investor sentiment on Twitter, about European indices, on the returns and trading volume of the European indices?*

In order to answer the main research question, this study constructs the following sub-questions, arranged in table 1, to determine the specific effect social media sentiment has on European indices. The first sub-question is constructed to determine if there is a positive relationship between Twitter sentiment about European indices and the returns of European indices. This relationship would indicate that positive (negative) investor sentiment on Twitter would result in positive (negative) returns of the European index.

[1]This study uses European as a reference to continental Europe, thus including the United Kingdom.

The second sub-question is looking at the volume of Twitter content expressed by investors, which is used to measure the involvement of investors, where high Twitter volume represents high investor involvement. In order to determine whether high investor involvement results in higher index returns, the sub-question examines if Twitter volume about European indices has a significant positive relationship with the European index returns. It is necessary to divide Twitter sentiment and Twitter volume into separate variables since the former measures both the polarity and magnitude of the sentiment directly from Twitter content and the latter only measures the magnitude of the investor involvement indirectly from Twitter volume.

At last, this study examines the relationship of Twitter sentiment and Twitter volume separately with the trading volume of the European indices. Examining these relationships could help predict the trading activity of investors and, therefore, the volatility of stock since trading volume and return volatility are closely linked (Jones et al., 1994). In order to know whether positive investor sentiment leads to more trading, this study uses sub-question 3 to determine if there is a significant positive relationship between Twitter sentiment and trading volumes of the index. In order to know whether high investor involvement results in more trading, this study uses sub-question 4 to determine if there is a relationship between Twitter volume and the trading volume of the index.

Table 1: **Sub-research questions**

No.	
1	*What is the effect of Twitter sentiment on the returns of European indices?*
2	*What is the effect of Twitter volume on the returns of European indices?*
3	*What is the effect of Twitter sentiment on the trading volume of European indices?*
4	*What is the effect of Twitter volume on the trading volume of European indices?*

All the sub-questions follow a similar hypothesis, where the null hypotheses denote that the specified Twitter variable does not have a significant effect on the specified European index variable, using a significance level of 5% for rejecting the null hypothesis and accepting the alternative hypothesis (i.e., there is a significant effect). This study follows a quantitative research design to test the hypotheses, incorporating data from multiple online sources to conduct the analyses needed to answer the research question.

Furthermore, this study analyses European indices instead of individual companies. Not only because the trading volume of European indices is much larger than individual stocks in these European indices, but also because these indices are less exposed to firm-specific risk as they are diversified by combining 30 or more individual stocks into an index. However, what needs to be noted is that this study only examines the tweets directly about the European indices and not the tweets about individual companies in the index. This decision is made since incorporating tweets about individual companies into the process would make the analysis much more complicated and infeasible for the time frame of this study. However, this study acknowledges the limitation of

only using index-related tweets, further explaining the limitation in the section 4.1.1.
The effect of investor sentiment and the stock market has been sought-after for quite some time now, the result of which could be essential in the field of finance. Currently, there are well-established studies that fill the empirical gap in the U.S.; however, the European stock markets have not been studied to our knowledge. By examining the European stock market, this study can make conclusions about the anomalies that could or could not exist in the European stock market. Furthermore, finding an effect could be helpful in algorithmic trading strategies to predict future returns of stocks. It could also be relevant on an economic scale, helping to understand and predict potential signs of inflated markets.

This research starts in section 3 by introducing well-established stock market theories and evaluating their assumptions and relevance, in which market anomalies are taken into account. To continue in section 3.4 with a comparison of economic and behavioural economic theories that build the foundation of this study, i.e., investor sentiment—providing empirical evidence in section 3.5 about investor sentiment and its relationship with investors behaviour and indirect consequences to the stock market.
Section 4 is dedicated to the methodology of this research, justifying the research design and its methods. In addition, explaining the processes for cleaning and analyzing the data variables, discussing their validity. Section 4.5 provides the assumptions used to make reliable interpretations of the analyses in section 4.5.2 - 4.5.4. Section 5 reveals the objective results of this study, incorporating the estimates of the analyses mentioned above. Then, interpreting these results and answering the research questions in section 6.

In conclusion, this study has not been able to find the effect that was expected when formulating the research questions. Instead of Twitter variables significantly affecting index returns, this study found that the returns of European indices have a significantly positive effect on the sentiment scores of the tweets. This result indicates that European index returns are not predictable using Twitter sentiment about the concerned index. Even though this result might not be helpful for algorithmic trading strategies that seek to predict future movements of the stock market, it could be of value to some seeking to understand investors' sentiment and expressions. However, since this study is limited by only using index related tweets, further research, including individual company tweets, is needed to completely dismiss the effect of Twitter sentiment on European index returns.

3 Theoretical Background

3.1 Asset Pricing Theories

This study is interested in finding the relationship between social media sentiment and index returns and trading volume. Social media platforms were introduced in 2000; therefore, measuring social media sentiment and its effect on stock prices is relatively new. However, some traditional frameworks studied the stock prices by looking at various factors that might relate to the stock returns. In order to show the relevance of including social media sentiment into these frameworks, this study will discuss the concept of these traditional frameworks and their flaws.

3.1.1 Capital Asset Pricing Model

In 1961 Treynor had developed a framework, the CAPM, to analyze expected asset return using the associated risk of the asset. The model is a general equilibrium model built on the earlier work (Markowitz, 1952) about portfolio selection. The model rests on the primary assumptions that there are only risk-averse investors in the market, that all investors are rational investors with similar expectations towards risk and returns and, at last, that all information in the market is available to every investor. Based on these assumptions and some additional ones, they have constructed a model that will describe the relationship between systematic risk and expected return. The following equation describes the CAPM

$$R_i = R_{rf} + \beta_1 (R_{mkt} - R_{rf}) + e_i \tag{1}$$

where R_i, R_{rf} and R_{mkt} are the expected return of asset i, the risk-free rate and the expected return of the market, respectively and $E[e_i] = 0$. β_i is the volatility of asset i; to be more specific, it measures the systematic risk of the asset.

Systematic risk is generally called market risk, which in short, is the risk of losses due to fluctuations in the overall market prices. On the other hand, unsystematic risk is considered firm-specific risk because it relates to the individual company's risk. The CAPM only incorporates systematic risk to compute expected returns since modern portfolio theory has shown no risk premium for unsystematic risk (Markowitz, 1952). By combining different assets, investors can diversify all the unsystematic risk, leaving only systematic risk in a portfolio.

3.1.2 Arbitrage Pricing Theory

In order to improve the restricted one-factor model, that is, the CAPM, Ross (1976) developed the Arbitrage Pricing Theory to incorporate multiple macroeconomic variables into the equation to capture the systematic risk. Due to multiple factors in the model, it is far less restricted by assumptions than the CAPM. Nevertheless, both the CAPM and the APT are widely used frameworks that support the following two most important economic and financial hypotheses.

3.2 Efficient Market Hypothesis and the Random Walk Theory

The efficient Market Hypothesis (EMH) and the Random Walk Theory (RWT) are traditional market theories suggesting that investors make rational investment decisions in the stock market. These theories are well established and emphasize the efficiency and unpredictability of the stock market. Efficiency means that the market value of a stock (i.e., the market price) would represent the fundamental value (i.e., the true value), resulting in a market with almost no undervalued stocks. The general idea behind these theories is that the market is unpredictable, and even if it were predictable, investors would not be able to systematically outperform them since they are efficient most of the time. The RWT of Fama (1965b) describes the unpredictability of the stock market due to the stock prices that follow a Random Walk. The theory argues that stock prices today are independent of both stock prices tomorrow and prices of other stocks. Therefore, it would be impossible to predict stock movements based on past or present observations. The EMH states that all information about a certain stock is already incorporated into the stock price on the market (Fama, 1965a). There are different forms of this hypothesis, but the strongest one assumes that the prices of a company stock even reflect inside information about the company. Therefore, both the RWT and the EMH conclude that investors are better off investing in a passive investment portfolio because systematically outperforming the stock market with active portfolio management is impossible when there are no information advantages, and predictions are worthless. These theories dismiss the ability of social media to provide information advantages for meaningful predictions and therefore assume no effect of social media sentiment on stock prices. This study is designed to determine whether this assumption still holds or if the market circumstances have changed to such extend that we can reject these theories.

3.3 Market Anomalies

Since the theories, as mentioned earlier, and hypotheses are based on the efficiency of the market, they have a common challenge in explaining market anomalies. Market anomalies are occurrences in the stock market that are inconsistent with the aforementioned theories and hypothesis expectations. There is empirical evidence that questions the efficiency of the market and provides factors that have a specific correlation with stock prices and their returns. For example, Lo and MacKinlay (1988); Kavussanos and Dockery (2001) show, respectively, that stocks have positive autocorrelation between past returns and correlated with other stocks on the stock exchange. This motivated other academics to test the RWT, resulting in various studies where the RWT could not hold (Gibbons and Hess, 1981; Gallagher and Taylor, 2002; Qian and Rasheed, 2007). Over time, several studies have also questioned the relevance of the EMH by describing different anomalies in the stock market. The list of accepted anomalies consists of the 'size effect', 'Book-to-Market effect' and 'momentum effect'. The 'size effect' has been introduced by Banz (1981), finding a 'size effect' in the stock market, meaning that portfolios with small-sized stocks tend to have significantly higher average stock returns compared to portfolios with big

sized stocks.
Furthermore, Eugene and French (1992) found that the Book-to-Market ratio of a company significantly influences the average stock returns. High Book-to-Market ratio stocks, referred to as growth stocks, often have higher average stock returns than low Book-to-Market ratio stocks.
At last, Jegadeesh and Titman (1993) accepted a market anomaly called 'momentum effect by finding a one-year momentum anomaly in the stock market. Meaning, stocks that performed well in the past would perform better in the future compared to stocks that performed poorly in the past. These three anomalies suggest that we can not only reject the EHM and the RWT but also need to adjust the asset pricing models, like the CAPM, in order for them to still be relevant with these anomalies in the market. This supports the necessity for this study since there might be even more anomalies, like social media sentiment, that academics just have not created enough evidence for to be widely accepted.

3.3.1 Adjusted Asset Pricing Models

In order to adjust the theoretical models to the established market anomalies, Fama and French (1993) tried to improve the CAPM by introducing two additional market factors. These factors are used to capture both the aforementioned 'size effect' and 'Book-to-Market effect'. This model has proven to be more accurate in predicting expected excess returns of an asset because it incorporates the market anomalies that the CAPM and APT cannot explain. In 1997, Carhart tried to improve the Fama and French Three-Factor model by including a fourth factor to adjust for the 'Momentum effect', mentioned in section 3.3. Although these adjusted asset pricing models can explain a fair part of the market anomalies, they tend to focus on the technical aspects of the stock market and disregard behavioural factors influencing the investors in the market.

3.4 Behavioural Finance

Behavioural finance is a separate branch in the finance field dedicated to explaining behavioural traits of people in situations with financial choices. An important observation was made by Shiller (2000b), who found that investors are often irrational when investing—creating so-called' bubbles' in the market, which means that the prices do not represent the fundamental value. This contradicts the RWT and contradicts the implication of the EMH that all stock prices reflect the fundamental value of the stock, meaning that the market is not efficient in every aspect.
According to Shiller (2000b), this inefficiency is partially caused by millions of people being too optimistic about their investment lacking a foundation of fundamental value. This emphasizes the behavioural aspect in financial markets that is generally overlooked by the aforementioned traditional and adjusted asset pricing models. The main reason for the gap in these frameworks is that they were developed on the concept of rational investors. Rational investors would not invest in assets that would be overvalued, meaning that they would only invest

if the fundamental value is equal to the asset price. In such a market, overvalued assets would automatically decrease in price to maintain the market equilibrium.

However, people are not as rational as we would like to believe. Kahneman and Tversky (1982) suggested that investors are constantly prone to biases, like overestimating their ability to predict future stock prices, which causes the investors to be overconfident and irrational in their investment choices. Along with other behaviour psychology theories, this suggests that investors are not making rational choices, which could cause inefficiencies in the stock market. A great example of an inefficient market is shown by the collapse of the housing market in 2008 in the United States. According to Shiller (2000b), besides very supportive monetary policy, one of the leading causes was too optimistic, emotional investors that overestimated the future returns of the housing market. A significant percentage of this effect can be allocated to emotional contagion, which is the phenomenon of people affecting each other's emotions and behaviour. Not only investors but also stock analysts are prone to this psychological bias, creating irrational equity reports. This, in order, leads to herd behaviour and incorrect valuation creating an upward spiral of irrational expectation inflating the asset prices in the market (Shiller, 2000a). This example, among other studies, shows enough empirical evidence for the stock market's inefficiency (De Long et al., 1990; Baker and Wurgler, 2006). Shleifer and Vishny (1997) go a step further and argue that the rational investors in the market cannot force prices of the assets to their fundamental value like the CAPM and APT would suggest.

These arguments are essential in the way that they stress the importance of rejecting the assumption of rational investors, that the asset pricing theories make use of, and examine the behaviour of both rational and irrational investors to incorporate into asset pricing models, in order to make accurate expectations about the stock market. This does not only apply to the, so far mentioned, stock returns. It also plays a big role in predicting the trading activity of investors and thus the trading volume of stock

3.4.1 A Model of Competitive Stock Trading Volume

The model of Wang (1994) describes the fluctuations in trading volume as a reaction to the activity of rational investors. Rational investors trade when two changes in the market occur. First of all, they trade when the expected return of their investment changes to take advantage of this change. Second of all, when they observe that the price deviates from its fundamental value, caused by irrational investors. In this case, they are taking a position to speculate against the expected future correction of the market.

The model indicates that investors balancing their portfolio is automatically associated with trading. Therefore, when investors find the need to re-balance their portfolio due to the reasons mentioned above, it will automatically increase trading. Due to the assumption of this model that investors are risk-averse, trading in stock will always be accompanied by stock price fluctuations, otherwise called volatility. This model gives a clear insight into how the behaviour of rational investors accompanies trading volume; however, it is somewhat outdated. As mentioned earlier, these theoretical models hold very well when fundamental value is identifiable and indicating a clear difference between rational and irrational investors. However, nowadays, it is hard to identify a stock's actual fundamental value because many companies rely on technology that is far less comprehensible to investors than traditional financial statements. Therefore, this section boils down to the conclusion that both this model and the asset pricing models should incorporate a factor for investors' behavioural traits to adjust for anomalies in the market.

3.5 Sentiment

In order to make sense of the behavioural traits of investors in the stock market, academics have been starting to use "top-down" approaches to study the emotional states of investors in combination with stock market fluctuations. In order to study the emotional states of people, academics have constructed a variety of sentiment proxies to establish the sentiment among the general public or investors specifically. Hirshleifer and Shumway (2003) use the weather condition as a proxy to determine the mood of the investors. This study shows significant correlations between the weather and stock returns. This indicates that a positive mood among investors caused by pleasant weather results in higher returns on the stock market. Lemmon and Portniaguina (2006) have used consumer confidence to measure sentiment among investors and have shown significant relationships between investor sentiment and stock returns. Others have constructed indices using multiple proxies to function as investor sentiment (Baker and Wurgler, 2006; Baker et al., 2012; Lee et al., 2002). The index of Baker and Wurgler (2006), for example, measures trading volume, dividend premium, closed-end fund discount, IPO volume, the returns on the first day of the IPO's and equity shares in new issuance's. They are averaging the six proxies, giving the study an index that would indicate the sentiment state and the sentiment change at a given period. Academics used the index of Baker and Wurgler (2006) in different situations and found the same significant relationships between the sentiment index and stock aspects, like returns and volatility (Sibley et al., 2016; Huang et al., 2014).

Since 2000, studies have focused more on measuring sentiment directly from internet content to improve the accuracy of the sentiment proxies. By using internet content, these studies can directly measure sentiment among the general public or the investors by looking at the polarity of the textual content. Antweiler and Frank (2004) processed all textual content on the message board of Yahoo Finance and the Raging Bull, determining the sentiment among investors by classifying selected messages into three categories, "buy", "hold", and "sell". This method measured the sentiment that investors expressed every 15 minutes. Using this sentiment proxy, Antweiler and Frank (2004) was not only able to find a significant positive effect of investor sentiment on stock returns. They also found that polarization among investors resulted in a significant increase in the stock's trading volume. These studies that use internet-based sentiment data collectively find significant relationships between investor sentiment and stock returns and trading volume (Antweiler and Frank, 2006; Schumaker and Chen, 2006). Internet-based sentiment data is often obtained from two sources, news websites and social media platforms. The aforementioned studies were primarily using news websites because social media started creating data from 2004 and 2006 with the introduction of Facebook and Twitter, respectively (Maryville-University, 2021). That being said, it also took some time for the social media platforms to grow significantly.

Due to technological innovations, academics are now able to collect social media data per second and use this to construct new proxies for investor sentiment. These proxies for investor sentiment is therefore instantly collected from the investor itself. Using the messages or textual emotions that investors release on social media platforms, studies can determine the sentiment polarity of this content. Bollen et al. (2011) extracted six different mood dimensions from each Twitter post in their data set, Rao et al. (2012) used a Naive Bayesian classification method to determine the number of positive and negative Twitter posts certain period. Even though these studies have different approaches to determining investor sentiment, they all try to interpret the textual content directly from the investor, making the sentiment proxy for investor sentiment more accurate than the news article-based sentiment proxies. This not only resulted in social media-based studies finding significant relationships between social media content and stock market returns, but some were also even able to find the predictability of future returns (Bollen et al., 2011). Many of these social media-based studies have been using Twitter data about the stock market due to the size of Twitter and the type of messages on Twitter, called' tweets' (Bollen et al., 2011; Rao et al., 2012; Ranco et al., 2015). These tweets are highly informative because they are public and often describe emotions, current thoughts and/or reactions.

These aforementioned studies are all done on the U.S. stock market. This study is interested in examining the same relationship using the European stock markets. Even though the European stock market is much smaller than the U.S., European markets are also subject to investors who tend to trade irrationally, creating an anomaly in the European market. In order to determine the exact relationship of investor sentiment with the price fluctuations and trading volume of the European indices, this study will use similar approaches as the aforementioned U.S. studies, using Twitter data to measure investor sentiment directly.

4 Methodology

This study uses a quantitative research design, processing data samples, to perform statistical analysis on the effect of investor sentiment on trading volume and returns of European indices. The numerical results are further translated into comprehensive relationships that provide the foundation for answering the sub-questions in this research.

4.1 Data Collection

A data sample from Twitter is collected to construct the variables 'Twitter sentiment' and 'Twitter volume' (Twitter, 2021). The financial data sample is drawn from the First Rate Database, a paid data set that contains intraday data about European indices (First Rate, 2021). The data sample includes three different European Indices; the German DAX, the British FTSE 100 and the French CAC 40, selected on both size and tweet volume to ensure a sufficient amount of data.

4.1.1 Twitter Data

A data set from Twitter includes numerous tweets scraped from the Twitter platform, with each tweet having an identification number, time, date, textual information, username. The data is scraped from the Twitter platform using the Python[2] package 'Twint' (Van Rossum and Drake Jr, 1995), an advanced Twitter scraping tool that lets Python users scrape historical data from the Twitter platform (Zacharias and Poldi, 2021). Many data analysts use Twint, it can retrieve data freely beyond the last 3200 tweets on the subject, and it has got a better rating on GitHub than the official Twitter scraper (i.e., Tweepy) (Hwang, 2020). Therefore, this study can analyze the data over a whole year, increasing the validity of this study, as a short time frame would be more susceptible to temporal abnormalities in the data.

To retrieve the tweets that relate to the concerned European indices, search operators in Python, like "since", "until", and "search", are used to create scraping boundaries. The data set only includes English tweets to ensure that the sentiment analysis tool, called VADER, correctly interprets the polarity of the textual content in the Twitter data. The DAX Twitter data set contains roughly 66.000 tweets, 37.000 of those are in English; for the FTSE 100 Twitter data set, this is 60.000 to 47.000, and for the CAC 40 Twitter data set, this is 61.000 to 15.000. This exclusion of non-English tweets could lead to missing values in the cleaned data set. Therefore, this study assumes that English tweets represent the sentiment of all investors (i.e., English and non-English investors) and that time intervals without tweets represent investors that are neutral in their sentiment. Even though both of these are not preferred assumptions, they are needed for the VADER sentiment analysis to accurately interpret the textual context since VADER has only proved accurate with English text. Furthermore, retweets are excluded

[2]The following GitHub page contains all the Python and R-studio code that has been used in this study to scrape, clean, process and analyse the data. https://github.com/MarkdeKwaasteniet/Thesis-Investor-Sentiment

from the data set, as they are shared emotions of the original tweet and therefore much harder to interpret. This poses a limitation for this study, but makes the interpretation of the tweets less biased.

Tweets are scraped from the 1st of January 2018 until the 1st of January 2019 because scraping Twitter data is cumbersome, making multiple years infeasible. Furthermore, this time frame entirely excludes the effects of COVID-19, which is essential as it severely impacted the stock market in 2020 (Baker et al., 2020). The time frame should pose no limitation to the study, as multiple similar studies had significant results using a timeframe of less than a year before 2018 (Bollen et al., 2011; Mittal and Goel, 2012; Rao et al., 2012).

This study does not use single keywords, as this includes tweets that are not related to a European index; a user could refer to Dax as a person instead of the index. Instead, Cashtags and Hashtags are used. Cashtags relate to the Dollar sign that investors combine with a stock ticker symbol to indicate that the tweet is about a certain company or index. The following sentence shows an example from the Twitter data set, highlighting the cashtag search operator.

*"Read this article to find our view when is a good time to buy World Indices again **$DAX**"*

Hashtags relate to the hash symbol that is used in front of a word to indicate an emotion or reference about a company or index. The following sentence shows an example from the Twitter data set, highlighting the hashtag search operator.

*"Horizons DAX Germany ETF Plans Special Dividend of $0.44 **#DAX**"*

Generally, indices are constructed by combining multiple company stocks into a single large portfolio. The German DAX index, for example, contains 30 large companies on the Frankfurt Stock Exchange, implicating that indices are directly affected by the performance of the underlying stocks. However, this study only uses tweets directly regarding the considered European index since incorporating tweets about individual companies makes the analysis much more complicated and infeasible for the time frame of this study. This implicates that the study is measuring delayed investor sentiment, as investors express sentiment about the individual companies first, then the sentiment about the index second. Even though this poses a limitation, in terms of incomplete investor sentiment, further discussed in section 6.1.1, the data set containing only index tweets is still large enough to perform the primary analyses in this study.

4.1.2 Financial Data

Financial data, scraped from the 'First Rate' database, provides historical intraday data on multiple European Indices, using 30-minute, 60-minute and daily intervals, since 1-, 5- or 10-minute intervals do not have a sufficient amount tweets. The data set contains the Opening price, the High, the Low and the Closing price, where the Opening and Closing prices refer to the first and the last price of the 30-minute, 60-minute or daily interval, respectively. The timezone of each index is set equal to the timezone in the country of origin, e.g., the

DAX index is set in timezone CET (CEST) and therefore opens on a trading day at 09:00 CET.

This study tested multiple data sources to obtain the data set with the highest quality. Even though studies used 'Dukascopy' for financial data, the data lacked an extensive quality review. Furthermore, this study performed validity tests by comparing the Dukascopy data with Yahoo Finance, Bloomberg and Investing.com, concluding that the data from Dukascopy deviated too much, compelling this study to pay for an intraday data set at First Rate. The quality of the First Rate data is checked by creating visual representations and calculating descriptive statistics. Figure 1 shows the visual representation of the data set with the DAX index, indicating an abnormality in trading volume of both the DAX and the CAC 40 on September 21, 2018. However, this represents a peak in the data instead of an error, consistent with other data sources and news articles on that date (Euronext, 2018).

Figure 1: Daily closing prices (top panel) and the corresponding trading volumes (bottom panel) of the European indices.

Table 11 shows the descriptive statistics to compare the First Rate data sets with the aforementioned sources. The First Rate data only deviated from the other sources at two decimals behind the dot, which is a sign that the data of First Rate is consistent, concluding that the quality is sufficient to continue with the further statistical analysis.

Table 2: Descriptive Statistics of the European indices

Daily DAX data set	N	Mean	St. Dev.	Min	25%	75%	Max
Open	251	12283.66	656.25	10452.02	11992.39	12690.56	13577.14
High	251	12346.62	651.98	10586.80	12041.00	12769.58	13596.89
Low	251	12199.62	664.90	10279.20	11883.12	12616.56	13517.81
Close	251	12269.59	664.22	10381.51	11964.95	12688.44	13559.60
Volume	251	1.0029e+08	3.2522e+07	5.0770e+07	8.3118e+07	1.0971e+08	4.0017e+08
Daily FTSE 100 data set	N	Mean	St. Dev.	Min	25%	75%	Max
Open	253	7366.68	304.88	6584.70	7105.30	7652.90	7877.50
High	253	7404.18	298.74	6721.20	7162.60	7685.70	7903.50
Low	253	7326.62	311.38	6536.50	7062.10	7615.60	7854.60
Close	253	7362.89	306.83	6584.70	7103.80	7651.30	7877.50
Volume	253	7.9596e+08	1.9295e+08	1.9281e+08	6.7906e+08	8.7251e+08	1.7116e+09
Daily CAC 40 data set	N	Mean	St. Dev.	Min	25%	75%	Max
Open	255	5297.45	214.66	4641.05	5148.71	5472.83	5637.94
High	255	5321.81	211.83	4664.20	5176.66	5491.81	5657.44
Low	255	5267.99	221.12	4555.99	5115.62	5448.60	5628.93
Close	255	5293.62	219.09	4598.61	5148.32	5473.34	5640.10
Volume	255	8.5231e+07	2.4040e+07	2.3250e+07	7.1796e+07	9.4334e+07	2.2251e+08

4.2 Data Cleaning

After data selection, cleaning the Twitter data is essential to make the data uniform and adjusted for the appropriate models. The financial data does not need much cleaning after retrieval, cleaning only the date, time, and timezone to match the Twitter data.

4.2.1 Twitter data

Irrelevant variables in the Twitter data set are excluded, keeping the tweet, date and time. Furthermore, it is essential to clean the textual part of the tweet variable since the VADER sentiment analysis uses the words and characters in a tweet to compute polarity scores. Improperly cleaned tweets result in incorrectly interpreting the polarity of the textual content, creating incorrect sentiment scores of the tweets. The following steps have been taken to clean the tweets.

- Deleting the hashtag symbol (#) to analyse the emotion of the hashtag message.
- Replacing the cashtag of a company or index (e.g., $DAX) with the symbol of the index in lower case.
- deleting the '@' symbol when there is a mention of another user in the tweet.
- Deleting website URLs that are mentioned in the tweet.

This study does not clean the data in terms of capitalization, punctuation or emoticons/emojis since the VADER sentiment analyzer is built for reading social media content, which is where these three grammatical applications are often used.

4.3 Data Processing

4.3.1 Tweet Volume

The variable 'Tweet Volume' is obtained by counting the number of tweets posted in the aforementioned time intervals. For example, on January 2th between 14:30-15:00, five tweets were posted about the DAX index, meaning that the Tweet Volume at 14:30 is equal to 5.

4.3.2 Twitter Sentiment

The VADER sentiment analysis is used on the tweets in the Twitter data to obtain the variable' Sentiment Score'. VADER stands for Valence Aware Dictionary and Sentiment Reasoner, a lexicon and rule-based sentiment analysis tool, which is mainly used for determining the sentiment of customer reviews on social media. The VADER tool has outperformed individual human raters on Twitter content and scored 0.96 of 1.0 on the F1 scale, measuring the mean of precision and indicating the overall accuracy (Hutto and Gilbert, 2014). Sentiment analyses belong to the field of Natural Language Processing (NLP), a subfield of linguistics combined with computer science to let computers interact with human languages. There are multiple ways of letting computers interact with natural languages, one of those is rule-based, making the computer interpret natural languages whilst using the set of rules that humans have incorporated. Over the years, these sets of rules have become more complex to capture language heuristics, like this contrastive conjunction

The indices are looking good, ***but*** *this might not last! #DAX #Euronext*

where *but* indicates a contrastive conjunction, that the VADER tool takes into account because of the rules it uses. These computers combine these set of rules with a specified lexicon, which is a list full of language words. All these words in the lexicon have received a specific valence value, varying between 4 (extremely positive) and -4 (extremely negative), 0 being neutral. VADER uses these valence values from the lexicon to compute a 'compound score' for the overall sentiment of the sentence. This compound score is calculated by summing up the valence values of each word in the sentence, adjusting them by the specific rules of VADER and then normalizing this summed value. This results in a compound score between 1 (extremely positive) and -1 (extremely negative). Therefore, these compound scores are officially called 'normalized, weighted composite scores'. This example shows the output of the VADER analysis, following the procedure in figure 2,

The indices are looking good, but this might not last! DAX Euronext

'negative': 0.0, 'neutral': 0.817, 'positive': 0.183, 'compound score': 0.3054

where 'neg', 'neu' and 'pos' are the ratios to which the sentence is negative, neutral or positive, respectively. Note, that hashtags, cashtags and URLs have been removed before calculating the score.

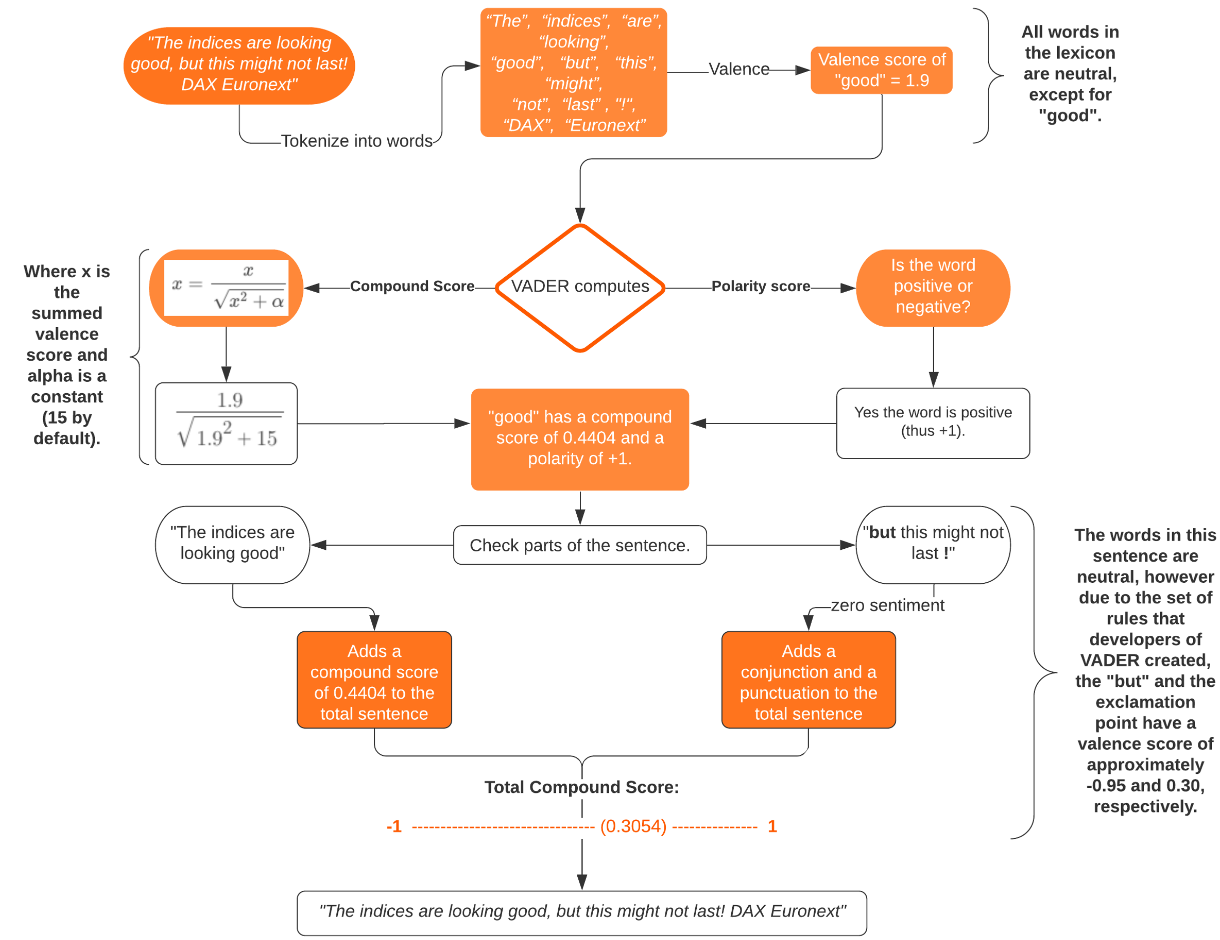

Figure 2: The process of calculating sentiment scores of tweets using the VADER analysis

Since VADER is lexicon and rule-based, the tool does not require training and can therefore be quickly applied to textual data. VADER works exceptionally well on social media content, compared to other sentiment analysis, as it has rules that are very accurate in detecting and interpreting emoticons, emojis and slang. These are often used in social media to express emotions but hard to interpret for alternative NLP tools. Additionally, VADER is trained on tweets that have been scraped directly from Twitter, using sentiment ratings from independent human raters to improve the accuracy of VADER on Twitter text.

However, the concept of causality and information flow needs to be addressed before using this computed sentiment score as a proxy for investor sentiment. When the tweets are analysed, they do not necessarily have to reflect current investor sentiment. The tweets could be time-delayed and represent past investor sentiment, or they could be a forecast function representing future investor sentiment. In order to simplify the statistical methods in this study, the tweets at time t are assumed to reflect the sentiment of investors at time t.

4.3.3 Returns and Trading Volume

The 'Index Returns' are calculated using the 'Closing prices' in the financial data set, representing the last price at the end of every time interval, i.e., 30-minutes, 60-minutes and daily. This study uses the index returns by combining the closing price at time t with the closing price at time $t-1$ because the 'Closing prices' are non-stationary. Non-stationary means that the variables follow a random walk and do not revert to a long term mean, figure 1 showing this trend. For the upcoming VAR models to be appropriate, the process of a variable needs to be stationary. By calculating the returns, the study takes the first difference of the closing prices, which transforms a non-stationary process into a stationary process that reverts to a long term mean of zero, shown in figure 3.

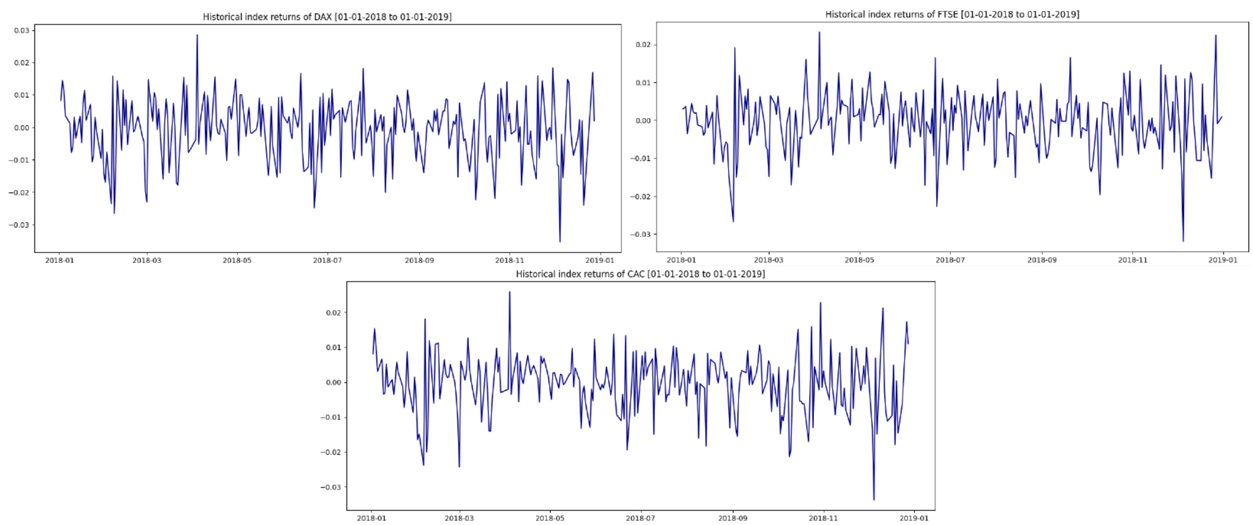

Figure 3: Daily European Index Returns

Furthermore, this study uses logarithmic returns because they are time consistent in contrast to discrete returns, making it convenient for summing index price changes in time series. In addition, logarithmic returns have the advantage of approximate raw-log equality, which means that if the index returns become very small, they return to the raw value of the returns. This is essential for intraday index returns because the returns can be very small within a 30-minute or 1-hour interval (Dougherty, 2011). The following equation is used to compute the short term logarithmic returns of the index

$$R_{it} = ln(\frac{P_t}{P_{t-1}}) \tag{2}$$

where i and t denote the European Index and the moment in time, respectively. This relates to the index returns within one time interval after the twitter sentiment has been measured.

For the primary analyses of this study, normal distributions of the variables are not necessary; however, it provides an understanding of the variables. Figure B7 shows that the distribution of the Log Returns has a normal curve, but it is much thinner than a log normal distribution, indicating positive kurtosis. This implicates high frequencies of extreme positive and negative values, which is often the result of a strong price reaction on information in the market (Lux, 1998). The probability plot in figure B7 shows deviations from theoretical distribution, indicating signs of skewness, which could result from several outliers. After manually checking these outliers, they are kept in the data set since they are not the result of inaccurate data but highlight important events for the statistical analyses to observe.

To construct the 'Trading Volume' variable, this study uses only the daily data, which provides the number of trades made within a daily time interval. This study prefers to analyze the trading volume with intraday data; however, First Rate data only provides daily trading volume, and the 'Dukascopy' data turned out to be inconsistent in section 4.1.2. Table B15 shows that the Trading Volume variable is statistically stationary; thus, there is no need to take the first difference in preparation for the VAR model. Figure B8 shows the distributions of the Trading Volume of the European indices, showing that the data distributions are similar to the index returns in terms of positive kurtosis and skewness.

4.4 Merging the variables

The Twitter variables are merged with the financial variables, resulting in three data sets for each index at a 30-minute, 60-minute and daily interval. It is difficult to determine if investor sentiment on a non-trading day represents sentiment about past or expected future index movement. Therefore, this study assumes that the tweets at time t reflect investor sentiment at time t, removing non-trading days after merging the data sets, to minimize interpretation bias of investor sentiment. Table 3 shows the number of missing values after merging the data sets, indicating that there are time intervals where no tweet has been placed about the index. The missing values are mainly the result of excluding non-English languages, resulting in the assumptions mentioned in section 4.1.1. Table 3 shows that the missing values of the FTSE 100 and the DAX are less than 5% of the total number of observations and are therefore considered neglectable. The CAC 40, however, has a significant number of missing values, indicating that the results from the analysis with this data set should be carefully interpreted since the estimates are prone to estimation errors, creating the possibility of biased inferences.

Table 3: Missing Values in the Data Sets

	Daily Data			30 Minute Data			60 Minute Data		
	DAX	FTSE 100	CAC 40	DAX	FTSE 100	CAC 40	DAX	FTSE 100	CAC 40
Total Observations	**251**	**253**	**255**	**4510**	**4537**	**4570**	**2255**	**2269**	**2285**
Sentiment Score	0	0	0	204	73	907	8	1	166
Tweet Volume	0	0	0	204	73	907	8	1	166
Log Returns	0	0	0	0	0	0	0	0	0
Trading Volume	0	0	0						

4.5 Statistical Analysis

4.5.1 General Analyses

The general analyses consist of both descriptive statistics as well as stationary tests for the upcoming statistical analyses. The descriptive statistics give insight into the different variables in terms of mean, standard deviation, min and max and the distribution of the time series found in table B11 - B13.

For the VAR model to be estimated, all variables in the equation need to be stationary. To test whether the variables are stationary, the augmented Dickey-Fuller test is used. The ADF equation is an extended version of the Dickey-Fuller test, where the change in the dependent variable is regressed on its lagged self to estimate the coefficients whilst allowing for autoregressive processes. The ADF test holds the following hypothesis

Null Hypothesis: The variable is unit root, non stationary.

Table B15 - B17 show that the null hypothesis can be rejected for all variables in every data set, accepting the alternative hypothesis that all variables are stationary.

4.5.2 Vector Autoregressive Models

This study constructs VAR models to obtain the effect of the Twitter variables on the Financial variables. The VAR model is an extended form of the Distributed Lag (DL) model. The core DL model is a linear regression model capable of determining the effect of lagged independent variables on the dependent variable. The Autoregressive Distributed Lag (ARDL) model is an extension of the DL model that incorporates an autoregressive process, including the model's lagged dependent variable. This is essential since this study uses time-series data, where the dependent variable is often correlated with its lag. Equation 3 provides an example of an ARDL(p,q) model

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_p x_{t-p} + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \cdots + \gamma_p x_{t-p} \epsilon_t \tag{3}$$

where p and q denote the number of lags of the independent and dependent variable, respectively.

A VAR model can be seen as multiple ARDL models where the regressand is stored into a vector. Constructing an ARDL model in this study does not suffice; since there is no certainty that the Sentiment Score affects Log Returns in this exact order, there is a possibility of Log Returns affecting the Sentiment Score. The following equation shows the basic bivariate VAR model that is used in this study

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}, \tag{4}$$

where y, t and p denote the variable, time and number of lags included in the model.

This study uses the Akaike Information Criterion (AIC) test statistic to determine the most suitable lag number to include in the VAR models, selecting the model with the lowest statistics. The AIC constructs a statistic based on both the number of parameters (lags) and the maximum likelihood of the VAR model. It penalizes the number of parameters since fewer parameters make the model more parsimonious; however, it penalizes less for the number of parameters compared to the BIC or SBIC, emphasizing more on model performance than on model complexity (Dougherty, 2011). This is preferred in this study, because the number of parameters in a VAR model increase twice as fast as the number of lags.

This study excludes any control variables in equation 4, because control variables, like changes in inflation, interest rates or GDP, are very hard to gather within a time interval of 30-minutes or 60-minutes. Control variables can be included in the daily interval VAR models, but this would implicate inconsistent methods; therefore, these are left out as well. Studies have shown that the aforementioned macroeconomic factors negatively correlate with investor sentiment (Yu, 2013), meaning that if interest rates increase, earnings of investors decrease, negatively affecting the overall sentiment of investors, resulting in a decrease in demand for the index and thus a price decrease. Therefore, because this study assumes that changes in macroeconomic effects need to affect investor sentiment to affect index returns or trading volume, the limitation of not including control variables in the VAR models is assumed to be minimal. This is consistent with the aforementioned studies that similarly exclude control variables (Bollen et al., 2011; Rao et al., 2012).

The following assumptions regarding the VAR models are required to compute unbiased coefficients and efficient standard errors.

1. There should exist no perfect multicollinearity, meaning that there should be no linear relationship between the independent variables. Multicollinearity causes the coefficient estimates to have increased variances, making them unstable and hard to interpret. However, this study uses a bivariate VAR model, which means that no additional variables pose the threat of perfect multicollinearity.

2. The error terms should have a conditional mean of zero in order to disregard the matrix with error terms in equation 4, leaving only the estimated coefficients for interpretation. This study plots the residuals in order to inspect the residuals and their conditional mean.

3. The error terms should be homoscedastic, meaning that the variance of the error terms is constant. If this is not the case, the error terms increase/decrease when the regressor increases, resulting in unbiased coefficients but inefficient standard errors. The coefficients are correct, but the p-values are invalid when determining whether the coefficients are significant. This is tested with the White test.

4. The error terms should be independent between observations, meaning that there is no autocorrelation. This study tests this assumption using the Breusch-Godfrey test. Even though it is prevalent for financial data to have autocorrelation, this study uses VAR models for the analyses, thus including several lagged dependent variables and automatically reducing the occurrence of autocorrelation.

5. The error terms should be normally distributed, i.e., the error term distribution should have a mean of zero and a standard deviation of 1. However, this assumption is the least restrictive since a normal distribution of the error term is not required for the OLS method to produce unbiased estimates. In order to test this assumption, the Jarque Bera test is used.

To answer the sub questions, the following VAR models (in matrix form) are estimated:

$$\begin{bmatrix} Return_t \\ SENT_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} Return_{t-p} \\ SENT_{t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} Return_t \\ TweetVolume_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} Return_{t-p} \\ TweetVolume_{t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \tag{6}$$

$$\begin{bmatrix} TradeVolume_t \\ SENT_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} TradeVolume_{t-p} \\ SENT_{t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} TradeVolume_t \\ TweetVolume_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} TradeVolume_{t-p} \\ TweetVolume_{t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \tag{8}$$

where p denotes the number of lags for the regressor and $Return$, $TradeVolume$, $SENT$ and $TweetVolume$ denote the log returns, trading volume, tweet sentiment and tweet volume of the index, respectively. Equation 9 and 10 show the VAR model in equation 5 and indicate the usage of 3 lags.

$$Return_t = c_1 + \beta_1 Return_{t-1} + \beta_2 Return_{t-2} + \beta_3 Return_{t-3} \tag{9}$$

$$SENT_t = c_1 + \beta_1 SENT_{t-1} + \beta_2 SENT_{t-2} + \beta_3 SENT_{t-3} \tag{10}$$

To determine the statistical significance of the estimated coefficients in the VAR models, the t-statistic shown in equation 11 is computed,

$$t = \frac{x - \mu}{S.E.(x)} \tag{11}$$

where x and μ are the coefficients and the population mean, respectively. In this study, the population mean is set to zero because the study is interested in determining whether the coefficients are different from zero and thus affect the dependent variable.

If the standard error is low compared to the coefficients, the t-statistic is high, implying that the probability of the coefficient is equal to zero is small. This study uses a significance level of 5%, meaning that the probability should be lower than 5% for this study to accept that the coefficient is significantly different from zero. However, the estimated coefficients of these models are not appropriate to interpret directly since many variables need to be controlled to make a valid conclusion regarding the effect of the coefficient. Therefore, the Granger Causality tests are performed, and impulse response functions are constructed to interpret the VAR models validly.

It is important to note that this study uses a VAR model instead of a VEC model, as variables in financial time series often need to be integrated before they can be analysed in a dynamic model. Integration means that the researcher has to take the first or second difference from the variable to make it stationary. When variables are integrated, they have the potential to be cointegrated with one another, resulting in two integrated variables having a significant correlation without them being directly related. Because this study uses index returns, which are first-order integrated, and considered cointegration. This study rejected the null hypothesis of the Johansen test because, according to Johansen (1992), different integration orders can never be cointegrated. Therefore, since the index returns are the only integrated variable, the VAR model is the preferred model for this study.

4.5.3 Granger Causality Test

Granger Causality tests are performed to determine the causality and significance of the relationships between the tested variables. Some events in life are always preceding each other in the same order, e.g., in order for the sun to shine, the clouds need to be absent. Not the other way around because the shining sun has no direct effect on clouds. However, in this study, the causality of the Twitter variables affecting the financial variables is not a certainty and must therefore be tested in both directions. The Granger Causality test does not directly test the causality; however, it analyzes whether one time series has specific predictive power of the other time series. The Granger causality test first constructs an autoregressive model with only a financial variable, like index returns. Then, if adding the Twitter variable, like sentiment score, has significant value to the R^2 of the model, the sentiment score will be Granger causal to the index returns. This study uses an F-test to determine the significance of Granger Causality statistic. The Granger Causality test is performed on all the aforementioned VAR models. All the models hold the same hypothesis.

Null Hypothesis: There is no causal effect of the independent variable on the dependent variable.

Using a significance level of 5% to determine whether this study is able to reject the null hypothesis and accept the alternative hypothesis, that there is a significant causal effect.

4.5.4 Impulse Response Functions

At last, this study constructs Impulse Response Functions (IRF) to determine the polarity of the significant effects given by the Granger causality results. These functions represent the effect that a shock in one variable has on another variable. The main idea behind this IRF is to observe the effect between the variables in the data set by creating a graphical representation of this shock. Because this study uses orthogonalized IRFs instead of generalized IRFs, the first parameter in the VAR model is not sensitive to a contemporaneous shock of any other parameters in the model. In addition, the last parameter in the model responds to shocks of all other variables. Furthermore, to correctly interpret the graph constructed by the IRF, an automatic bootstrap is constructed. This is the process of re-sampling the data of interest into smaller samples to achieve accurate confidence intervals of the population that the IRF shows.

5 Results

This section discusses the constructed VAR models and the validity tests on the residuals of the VAR models. Both of the analyses are essential for finding valid relationships in the data; however, the results cannot be interpreted directly. Therefore, Granger causality tests are performed and impulse response functions are constructed to interpret the coefficients in the VAR models and the direction of their effect.

5.1 VAR models

A total of 24 VAR models are constructed in 30-minute, 60-minute, and daily intervals with the three aforementioned European indices, using the four different variables shown in equation 5 - 8. Table 4 shows the VAR model of the daily DAX Log Returns and Sentiment Scores. The table contains the coefficients' estimates, the standard errors of the coefficients, the student t-statistics and the probability of the coefficient being equal to zero, based on the t-statistic. Both table 5 and 6 highlight the significant coefficients in the models, indicating that there is no lagged variable of Sentiment Score that significantly affects the Log Returns, but on the contrary, indicate that the first lagged variable of Log Returns significantly affects the Sentiment Score positively. Furthermore, both models have lagged dependent variables that are statistically significant, merely indicating the presence of autocorrelation in the data.

Table 4: VAR model daily DAX data

Table 5: Linear regression model on Log Returns

	Estimate	Std. Error	t value	Pr(>|t|)
Log_Returns.l1	-0.0718	0.0645	-1.11	0.2671
Sentiment_Score.l1	-0.0044	0.0166	-0.26	0.7938
Log_Returns.l2	0.1472	0.0688	2.14	0.0334
Sentiment_Score.l2	-0.0166	0.0165	-1.01	0.3143
Log_Returns.l3	0.0736	0.0678	1.08	0.2792
Sentiment_Score.l3	0.0106	0.0166	0.64	0.5234
Log_Returns.l4	-0.1583	0.0675	-2.35	0.0198
Sentiment_Score.l4	0.0209	0.0158	1.33	0.1855
const	-0.0015	0.0012	-1.23	0.2208

Table 6: Linear regression model on Sentiment Scores

	Estimate	Std. Error	t value	Pr(>|t|)
Log_Returns.l1	1.3633	0.2466	5.53	0.0000
Sentiment_Score.l1	0.1570	0.0636	2.47	0.0142
Log_Returns.l2	0.1443	0.2630	0.55	0.5838
Sentiment_Score.l2	0.2254	0.0629	3.58	0.0004
Log_Returns.l3	0.0484	0.2593	0.19	0.8520
Sentiment_Score.l3	0.1164	0.0633	1.84	0.0671
Log_Returns.l4	-0.2633	0.2578	-1.02	0.3082
Sentiment_Score.l4	0.1860	0.0602	3.09	0.0022
const	0.0173	0.0047	3.68	0.0003

Even though these coefficients of this daily DAX VAR model imply that lagged Log Returns significantly affect the Sentiment Score, it is not possible to validly interpret the direct impact of these coefficients on the dependent variables. First of all, because the residuals need to be tested on the assumptions in section 4.5.2 before the study can assume that the standard errors of the coefficients are efficient. Second of all, because the coefficients in the model have a joint effect on the dependent variable that cannot be interpreted separately. It is challenging to interpret the lagged variables whilst assuming the other lagged variables to remain constant. Therefore, as the Granger causality tests and IRFs provide the interpretable results of this study, no further models concerning the DAX, FTSE 100 and CAC 40 are presented in this section; however, they can be found in the appendix.

5.2 Testing the residuals

The validity of the estimated coefficients and thus of the Granger causality tests are based on the assumptions of the residuals being homoscedastic, non-autocorrelated and normally distributed. If these assumptions do not hold, the coefficients remain unbiased, but the standard errors will be inefficient. This results in invalid Granger causality test statistics as it uses the standard errors of the coefficients in the VAR model to determine whether or not there is a significant causal relationship between the two variables in question.

First of all, the assumption of the residuals regarding the conditional mean of zero is checked. Since it is impossible to test this assumption statistically, a plot is constructed to indicate whether the conditional mean is approximately zero. Figure 4 shows the residuals of the regression in table 4, indicating that even though there are outliers present in the residuals, there are no significant periods in which the residuals are only above or below zero; therefore, the conditional mean is assumed to be approximately zero. This holds for all the VAR models shown in figure B11.

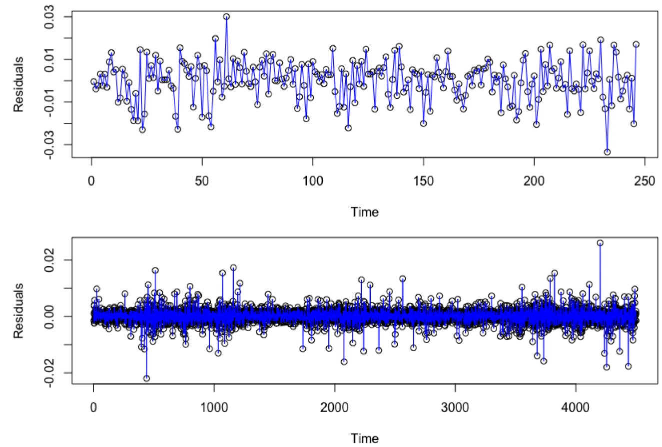

Figure 4: Residual plot of the regression in table 4 (top plot) and residual plot of the same regression with 30-minute data (bottom plot).

Table 7 contains all the p-values of the test statistics computed with the residuals of the VAR models to check for homoscedastic, non-autocorrelated and normally distributed residuals. The first row in every sub-table contains the p-values of the Breusch-Godfrey test, using an $\alpha < 5\%$ as the significance level. The table indicates no significant values for this test as all the values are larger than 0.05, meaning that there is no autocorrelation between the residuals in any VAR model. The second and third row of every sub-table is dedicated to the p-values of the White-test and Jarque-Bera test, respectively. Most of the p-values in the table are smaller than 0.05, indicating that there are heteroscedastic and non-normal distributed residuals present in the VAR models. However, the latter does not bear any significant challenges for the Granger causality test. According to the Gauss-Markov Theorem, OLS is still the best linear unbiased estimator (BLUE) if the residuals have a conditional mean of zero, are homoscedastic and have no autocorrelation. Therefore, homoscedasticity is the only assumption that is not yet satisfied, meaning that the standard errors of the VAR models will not be efficient and, therefore, would cause invalid Granger causality statistics.

Table 7: The P-values regarding the tests on the residuals of the VAR models.

	DAX Daily VAR models				DAX 30-Minute VAR models		DAX 60-Minute VAR models	
	Returns and Sentiment	Returns and Tweet volume	Trading volume and Sentiment	Trading volume and Tweet volume	Returns and Sentiment	Returns and Tweet volume	Returns and Sentiment	Returns and Tweet volume
Autocorrelation (BG-test)	0.365	0.431	0.790	0.683	0.554	0.238	0.402	0.345
Heteroskedasticity (White-test)	0.000	0.000	0.000	0.007	0.001	0.000	0.006	0.000
Normality (JB-test)	0.000	0.0001	0.000	0.000	0.000	0.000	0.000	0.000
	FTSE 100 Daily VAR models				**FTSE 100 30-Minute VAR models**		**FTSE 100 60-Minute VAR models**	
	Returns and Sentiment	Returns and Tweet volume	Trading volume and Sentiment	Trading volume and Tweet volume	Returns and Sentiment	Returns and Tweet volume	Returns and Sentiment	Returns and Tweet volume
Autocorrelation (BG-test)	0.868	0.902	0.716	0.565	0.051	0.415	0.093	0.330
Heteroskedasticity (White-test)	0.100	0.001	0.000	0.049	0.000	0.000	0.000	0.000
Normality (JB-test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	CAC 40 Daily VAR models				**CAC 40 60-Minute VAR models**		**CAC 40 60-Minute VAR models**	
	Returns and Sentiment	Returns and Tweet volume	Trading volume and Sentiment	Trading volume and Tweet volume	Returns and Sentiment	Returns and Tweet volume	Returns and Sentiment	Returns and Tweet volume
Autocorrelation (BG-test)	0.347	0.729	0.177	0.841	0.099	0.073	0.241	0.078
Heteroskedasticity (White-test)	0.004	0.015	0.151	0.295	0.000	0.000	0.000	0.000
Normality (JB-test)	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The Granger causality test can still be performed; however, it is essential to transform the inefficient standard errors into robust standard errors in the case of heteroscedasticity. Since the form of heteroscedasticity in this study is unknown, it is preferable to use a heteroscedasticity consistent covariance matrix (HCCM) (Long and Ervin, 2000). Over the years, multiple forms of the HCCM have been developed to adjust the standard errors for heteroscedasticity. This study uses "HC3" standard errors, developed by MacKinnon and White (1985). By taking heteroscedastic consistent standard errors, this study can validly continue the Granger Causality analysis.

5.3 Granger Causality Results

The Granger Causality test is a parametric test, meaning that the variables themselves need to be normally distributed for the test statistics to be accurate. Table B14 indicates that none of the concerned variables is statistically normal distributed, meaning that the distributions of the variables do not have a mean of zero and a standard deviation of 1. Even though this would be preferred, it does not restrict this study, as this study uses large sample sizes and can therefore rely on the Central Limit Theorem. According to this theorem, the distributions of the variables transform towards a normal distribution as the sample size grows larger. Thus, this study assumes that its data is approximately normally distributed, taking into account that the results are less accurate when the sample size is smaller, i.e., the 60-Minute and Daily data set.

Furthermore, this section is expressing the main results of this study. Since it is not possible to directly interpret the coefficients estimated by the VAR model, the significance of the relationships between the variables of interest in this study is estimated through Granger Causality tests. Table 8, 9 and 10 contain all the Granger Causality statistics, with the asterisk symbol indicating significant causal relationships. The statistics show no Twitter variables that have a significant Granger causal effect on the Index variables, except for the Daily FTSE 100 tweet volume on trading volume. This indicates that a change in the daily tweet volume about the FTSE 100 index Granger causes a change in the trading volume of the FTSE 100 index. This significant effect is in line with the expectations of this study.

Table 8: Granger Causality of Daily Data

***Cause**/Response*		Log Returns	Trading volume	Sentiment Score	Tweet volume
DAX	**Log Returns**			7.745***	3.665**
	Trading volume			1.995*	0.461
	Sentiment Score	0.7	0.742		
	Tweet volume	1.89	0.762		
FTSE 100	**Log Returns**			62.723***	2.5925*
	Trading volume			2.1006*	0.0045
	Sentiment Score	0.20686	0.1856		
	Tweet volume	1.2557	2.6582*		
CAC 40	**Log Returns**			16.553***	0.491
	Trading volume			2.6663**	2.0082*
	Sentiment Score	0.1685	1.189		
	Tweet volume	0.8961	0.9573		

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Granger Causality of 30 Minute Data

***Cause**/Response*		Log Returns	Sentiment Score	Tweet volume
DAX	**Log Returns**		3.371***	2.694***
	Sentiment Score	0.667		
	Tweet volume	0.742		
FTSE 100	**Log Returns**		6.943***	2.5068***
	Sentiment Score	1.1552		
	Tweet volume	0.3323		
CAC 40	**Log Returns**		2.4181***	0.6732
	Sentiment Score	1.1706		
	Tweet volume	0.742		

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Granger Causality of 60 Minute Data

***Cause**/Response*		Log Returns	Sentiment Score	Tweet volume
DAX	**Log Returns**		3.315***	2.3423***
	Sentiment Score	1.028		
	Tweet volume	0.256		
FTSE 100	**Log Returns**		15.612***	2.6622***
	Sentiment Score	1.8265		
	Tweet volume	1.0215		
CAC 40	**Log Returns**		3.0341***	1.1669
	Sentiment Score	0.417		
	Tweet volume	0.7325		

Note: *p<0.1; **p<0.05; ***p<0.01

The rest of the statistics in the Granger Causality tables show that the opposite effect can be found, meaning that changes in Log Returns or Trading volume is significantly Granger causing changes in the Sentiment Score or tweet volume. Table 8 for example, shows that changes in Log Returns are Granger causing changes in Sentiment Score, denoting an F-statistic of 62.723, which is highly significant. Furthermore, no single VAR model is

Granger Causal in both ways, which is a good sign since this would imply misspecification of the VAR model. However, the Granger Causality statistics do not specify the polarity of the effect of one variable on another. Therefore, Impulse Response Functions are constructed in the following section to thoroughly understand the significant relationships in the Granger causality tables.

5.4 Impulse Response Function

This study has constructed impulse response functions of the significant Granger causalities in table 8, 9 and 10, that graphically show the causal effect of one variable on the other.

First of all, this study looks at the significant effect of FTSE 100 tweet volume on the FTSE 100 trading volume, as this is the only Granger causality that is in line with the expectations of this study. Figure 5 indicates that a positive shock in tweet volume creates a positive shock in the trading volume of the FTSE 100. The grey area indicates the 95% confidence interval of the IRF, meaning that this study cannot say that the aforementioned effects hold 95% of the time as some parts of the grey area are negative.

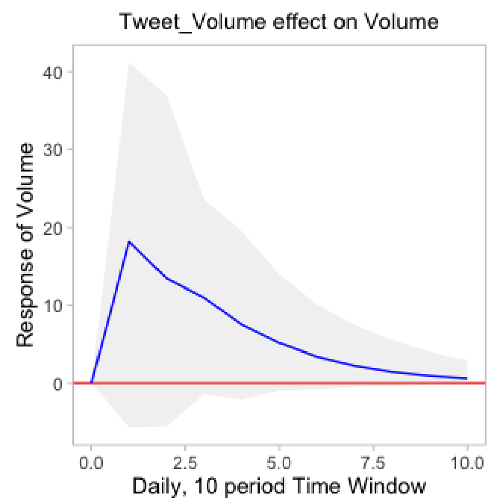

Figure 5: Impulse Response Function FTSE 100 Tweet volume on Trading volume, using a 10 day forecast

Furthermore, figure 6 shows the graphical representation of the significant Granger causalities in the DAX data and indicate the following effects. First of all, the IRF can be interpreted as a positive shock in the index returns causes a positive shock in the sentiment scores. Secondly, a positive shock in the index returns causes a negative shock in the tweet volume. Lastly, a positive shock in the trading volume of the index causes a negative effect on the sentiment scores. These effects are all consistent over the different time intervals, except for the trading volume, as this study cannot determine this relationship on 30- and 60-minute intervals. The IRFs shows a strong reaction after the shock, fading in the subsequent days and returning to zero around day 10.

Although these figures only illustrate the significant causalities in the daily DAX data set, figures B9 and B10 show that the direction of the effects is consistent regarding the other European Indices, i.e., FTSE 100 and CAC 40. For all indices at every interval, the Log Returns positively affect the Sentiment Score. Additionally, in both the DAX and the FTSE 100 data sets, Log Returns have a negative effect on the tweet volume, and Trading volume has a negative effect on Sentiment Score. However, the significance levels of the effects differ across European indices.

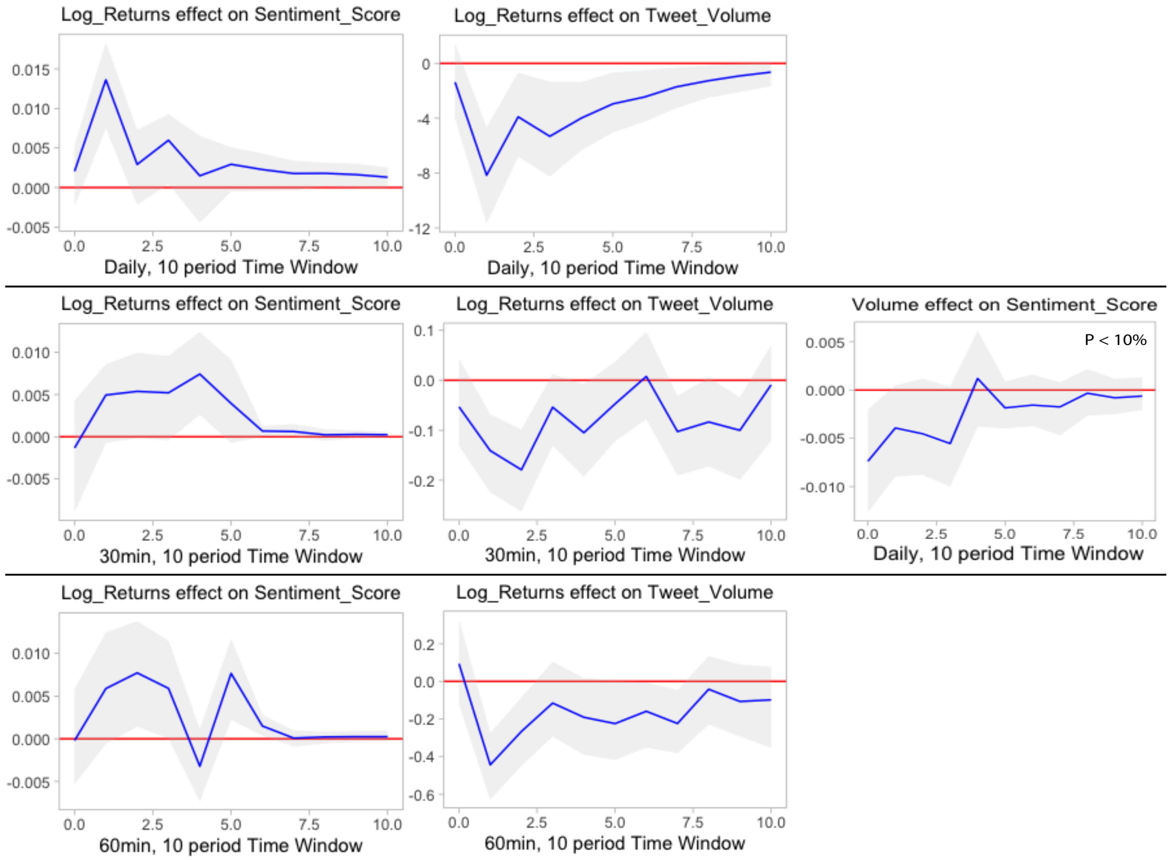

Figure 6: Impulse Response Function DAX index, using a 10 day forecast

6 Discussion

This study systematically analyses the effect of investor sentiment on Twitter, about European indices, on the returns and trading volume of the associated European indices. Discussing the findings of this study through answering each sub-question and concluding with the main research question. Furthermore, for the correct interpretation of the findings, the limitations, implications and future research are discussed.

6.0.1 Twitter Sentiment on the Returns of the European indices

The Granger causality tables show no significant causal effect of Twitter sentiment about the European indices on the returns of the European indices. This result is consistent regarding all the European indices, implying that there is no possibility of using investor sentiment about European indices to predict future European Index returns; therefore, the study cannot reject the null hypothesis regarding this sub-question. In contradiction to this study's expectation, the Granger causality test results show a significant causal effect of the returns of the European indices on Twitter sentiment about the European indices. The IRFs show that a positive change in a European index's returns positively affects investor sentiment about the European index. Even though the results contradict, they are not surprising, as the primary goal of investing is to create positive returns; thus, achieving this goal will create positive sentiment.

These results do not only contradict the expectations of this study, they are also not in line with the aforementioned studies (Bollen et al., 2011; Rao et al., 2012; Mittal and Goel, 2012). These studies found significant results of investor sentiment affecting the movement of stocks and indices, not the other way around. Some studies found no causality between the two variables, in either direction (Brown and Cliff, 2004; Zhang et al., 2018), but there are not many studies that found the causal effect of index returns on investor sentiment (Wang et al., 2006; Lo Giudice, 2015). The explanation may lie in the fact that almost no studies focus on predicting investor sentiment whilst using stock or index returns since predicting investor sentiment will not directly result in increased earnings. These contradicting findings could be the result of limitations in the data or methodology that is discussed in section 6.1.1.

6.0.2 Tweet Volume on the Returns of the European Indices

The Granger causality tables show no significant causal effect of the Tweet Volume, about the European indices, on the returns of the European indices. This result is consistent over all the European indices, implying that the involvement of investors regarding the European indices does not hold any predictability over the European index returns. Therefore, this study is not able to reject the null hypothesis regarding this sub-question, which contradicts, again, the expectation of this study and the results of similar studies (Rao and Srivastava, 2013). Similar to the effect of index returns on Twitter sentiment, this study has found the returns of European indices significantly affecting the tweet volume about the European Indices. However, this effect is only consistent for the DAX and the FTSE 100 index, making the generalization regarding all European indices less strong. The IRFs show a negative effect of returns on tweet volume, which means that investors are expressing less sentiment in terms of volume when the returns of the European index increase. This is in line with a minority of studies that find a negative relationship between twitter volume and stock returns (Leitch and Sherif, 2017).

6.0.3 Twitter Sentiment on Trading volume of the European indices

To answer this sub-question, this study has only been able to use the daily data set of the European indices, finding no significant effect of Twitter sentiment, about the European indices, on the trading volume of the European indices. Therefore, this study cannot reject the null hypothesis regarding this sub-question, implying that sentiment among investors does not affect their trading activity regarding the European indices. In contrast to the previous sections, this study did not find any significant relationships regarding these two variables. This finding is conflicting, as it is contradicting the majority of the aforementioned literature (Rao et al., 2012; Antweiler and Frank, 2004), but in line with the results of Brown and Cliff (2004). As noted at the beginning of this section, only daily data has been used; therefore, the study is not able to entirely exclude the effect of investor sentiment on the trading volume of European indices since shorter time intervals are more accurate in determining the short term effect of Twitter sentiment on the trading volume.

6.0.4 Twitter Volume on the Trading Volume of the European indices

Similar to section 6.0.3, only daily data sets are used to answer this sub-question. Table 8 shows that this study has been able to find a significant effect ($\alpha < 10\%$) of Twitter volume about the FTSE 100 index, affecting the trading volume of the FTSE 100. This implies that the number of tweets posted by investors could predict the trading activity of the investors interested in the FTSE 100. This statement needs to be carefully interpreted since it is only visible in the FTSE 100 index and not significant at 5%. However, because trading volume is assumed to relate to the volatility Antweiler and Frank (2004), it indicates that the volume of tweets posted by investors could be a variable that can predict future volatility of the indices in Europe. In order to validly conclude this, further research has to be done, including time intervals that are shorter than the daily intervals of this study. Furthermore, the study has not found significant additional effects of either of the variables on each other.

6.1 The Main Research Question

This study uses tweet sentiment and tweet volume as proxies to determine investor sentiment expressed on Twitter. Therefore, the first two answered sub-questions imply that there is no effect of social media sentiment on the returns of the European indices, and the second two answered sub-questions imply that there is no effect of Twitter sentiment on the trading volume of the European indices at a significance level of 5%. Therefore, concluding that there is no effect of social media sentiment, about the European indices, on the returns and trading volume of the European indices and accepting the null hypothesis of this study, meaning that it is not possible to study the sentiment of investors about European indices on Twitter to predict future returns or trading volume of European indices. The findings of this study are not able reject the traditional asset pricing theories, implicating that even though there might be anomalies, like size, market-to-book and momentum effects, investor sentiment is not likely to be one in the European index market. However, the study can reject the efficient market hypothesis and the random walk theory, as the various VAR models indicate autocorrelation.

The findings of this study should be interpreted with caution because of the important assumptions it used regarding the investor sentiment proxies. This study uses tweets posted by individuals regarding the European indices, assuming that the individuals who post the tweet are indeed investors involved in the European indices. Furthermore, this study assumes that the sentiment expressed by investors on Twitter at time point t reflects the actual sentiment of investors at that time point and does not reflect future or past sentiment of investors. At last, this study assumes that no sentiment expression by investors is equal to investors expressing neutral sentiment.

6.1.1 Limitations and Recommendations for Future Research

This study has several limitations in its data and methodology which will be addressed in order, for an accurate interpretation of the results.

1. The main limitation is that this study only used tweets directly related to the European indices. The European indices consist of 30+ individual companies; therefore, the stocks of the individual companies affect the price of the index. By measuring investor sentiment directly about the indices, the sentiment could be delayed, as investors would express sentiment regarding the individual companies in the index first. Furthermore, by not including the tweets about the individual companies, this study has not been able to measure the sentiment of all the investors that affect the returns and trading volume of the European indices. This limitation could have impacted the findings of this study, as the study finds a significant relationship, but only in index variables preceding investor sentiment. This implies that future research needs to focus on investor sentiment about the individual companies in the European indices to eliminate this delayed effect of investor sentiment.

2. This study used only English tweets to measure investor sentiment, deleting the non-English tweets from the original Twitter data set. This filtered out a large number of tweets, resulting in a data set that could only measure the sentiment of investors that posted in English; therefore, future research should find methods to incorporate the sentiment of non-English tweeting investors since these investors similarly affect the returns and trading volume of the European indices.

3. This study has not been able to retrieve daily interval data regarding the trading volume of the European indices. This poses a limitation as to the reactions of investors to publicly available information is considered very quick. Using only daily data might be insufficient to fully capture the effect of Twitter sentiment on the stock returns as the stock market reacts fast, and thus the effect might only be visible within minutes; therefore, future research should search for valid intraday data sets, including intraday trading volume.

4. This study has not taken retweets into the data set. Retweets are frequently used on Twitter, for example, by individuals who do not have the time or ambition to write a tweet but want to share the emotions or thoughts of a tweet posted by another user. However, these retweets could contain a certain amount of sentiment that an investor has regarding stock or index. Therefore, by not including these retweets in the data set, this study again fails to fully incorporate the sentiment that investors express on Twitter.

6.1.2 Practical Relevance

The primary findings of this study indicate that the investor sentiment expressed on Twitter does not affect the returns or trading volume regarding European indices. On the contrary, the study finds that the index returns have a significant effect on investor sentiment. The practical relevance is shown in the irrelevance of studying investor sentiment about European indices when the primary goal is to predict future movements or volatility of the European indices. This result is useful for both individual and institutional traders, as they often look for various trading strategies to beat the market, i.e., the European index market. These traders can ignore the usage of sentiment on Twitter in their trading strategies and search for other anomalies in the European index market.

6.2 Conclusion

This study concludes that traders sentiment, measured in strength and volume, does not affect European indices' returns or trading volume, implying that traders sentiment is not an anomaly that cannot be used for predictability purposes or reject the traditional asset pricing theories in the European index market. However, as this study is the first to study investor sentiment and its effect on aspects of the European indices, further research is needed that focuses on the limitations of this study to fully exclude the possibility of traders sentiment affecting returns and trading volume in the European stock market.

References

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294.

Antweiler, W. and Frank, M. Z. (2006). Do us stock markets typically overreact to corporate news stories? *Available at SSRN 878091*.

Autoriteit Financiële Markten (2020). Trend monitor 2021. Technical report.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The journal of Finance*, 61(4):1645–1680.

Baker, M., Wurgler, J., and Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of financial economics*, 104(2):272–287.

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., and Viratyosin, T. (2020). The unprecedented stock market reaction to covid-19. *The Review of Asset Pricing Studies*, 10(4):742–758.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1):3–18.

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

Brown, G. W. and Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of empirical finance*, 11(1):1–27.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1).

Chordia, T., Roll, R., and Subrahmanyam, A. (2001). Market liquidity and trading activity. *The journal of finance*, 56(2):501–530.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.

Dougherty, C. (2011). *Introduction to econometrics*. Oxford University Press.

Eugene, F. and French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.

Euronext (2018). Euronext Announces Volumes for September 2018. Technical report. retrieved from *https://www.euronext.com/en/about/media/euronext-press-releases/euronext-announces-volumes-september-2018*.

Fama, E. F. (1965a). The behavior of stock-market prices. *The journal of Business*, 38(1):34–105.

Fama, E. F. (1965b). Random walks in stock market prices. *Financial analysts journal*, 51(1):75–80.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds.

FinSMES (2019). How does forex trading work? *FinSMES, https://www.finsmes.com/2019/02/how-does-forex-trading-work.html*.

First Rate (2021). Historical intraday market price data. data retrieved from First Rate Data Base, *https://firstratedata.com/b/12/international-historical-index*.

Gallagher, L. A. and Taylor, M. P. (2002). The stock return–inflation puzzle revisited. *Economics Letters*, 75(2):147–156.

Gibbons, M. R. and Hess, P. (1981). Day of the week effects and asset returns. *Journal of business*, pages 579–596.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032.

Huang, C., Yang, X., Yang, X., and Sheng, H. (2014). An empirical study of the effect of investor sentiment on returns of different industries. *Mathematical Problems in Engineering*, 2014.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Hwang, J. (2020). What python package is best for getting data from twitter? comparing tweepy and twint.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of econometrics*, 52(3):389–402.

Jones, C. M., Kaul, G., and Lipson, M. L. (1994). Transactions, volume, and volatility. *The Review of Financial Studies*, 7(4):631–651.

Kahneman, D. and Tversky, A. (1982). The psychology of preferences. *Scientific American*, 246(1):160–173.

Kavussanos, M. G. and Dockery, E. (2001). A multivariate test for stock market efficiency: the case of ase. *Applied Financial Economics*, 11(5):573–579.

Lee, W. Y., Jiang, C. X., and Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of banking & Finance*, 26(12):2277–2299.

Leitch, D. and Sherif, M. (2017). Twitter mood, ceo succession announcements and stock returns. *Journal of computational science*, 21:1–10.

Lemmon, M. and Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, 19(4):1499–1529.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The review of financial studies*, 1(1):41–66.

Lo Giudice, M. (2015). The predictive characteristic of the social sentiment on the stock market: Twitter and the stock trend. B.S. thesis, University of Twente.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224.

Lux, T. (1998). The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. *Journal of Economic Behavior & Organization*, 33(2):143–165.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics*, 29(3):305–325.

Markowitz, H. (1952). Portfolio selection*. *American Finance Association*.

Maryville-University (2021). The evolution of social media: How did it begin, and where could it go next? *Maryville University, https://online.maryville.edu/blog/evolution-social-media/*.

Mittal, A. and Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Standford University, http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf*, 15.

Phillips and Lorenz (2021). 'dumb money' is on gamestop, and it's beating wall street at its own game. *New York Times, https://www.nytimes.com/2021/01/27/business/gamestop-wall-street-bets.html?searchResultPosition=49*.

Qian, B. and Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, 26(1):25–33.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PloS one*, 10(9):e0138441.

Rao, T. and Srivastava, S. (2013). Modeling movements in oil, gold, forex and market indices using search volume index and twitter sentiments. pages 336–345.

Rao, T., Srivastava, S., et al. (2012). Analyzing stock market movements using twitter sentiment analysis.

Robinhood (2021). What you need to get started. Article retrieved from Robinhood, *https://robinhood.com/us/en/support/articles/what-you-need-to-get-started/*.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. In *Handbook of the fundamentals of financial decision making: Part I*, pages 11–30. World Scientific.

Schumaker, R. and Chen, H. (2006). Textual analysis of stock market prediction using financial news articles. *AMCIS 2006 Proceedings*, page 185.

Shiller, R. C. (2000a). Irrational exuberance. *Philosophy and Public Policy Quarterly*, 20(1):18–23.

Shiller, R. J. (2000b). Measuring bubble expectations and investor confidence. *The Journal of Psychology and Financial Markets*, 1(1):49–60.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of finance*, 52(1):35–55.

Sibley, S. E., Wang, Y., Xing, Y., and Zhang, X. (2016). The information content of the sentiment index. *Journal of Banking & Finance*, 62:164–179.

Talwar, M., Talwar, S., Kaur, P., Tripathy, N., and Dhir, A. (2021). Has financial attitude impacted the trading activity of retail investors during the covid-19 pandemic? *Journal of Retailing and Consumer Services*, 58:102341.

Treynor, J. L. (1961). Market value, time, and risk. *Time, and Risk (August 8, 1961)*.

Twitter (2021). Twitter data about european indices. The data has been retrieved from *Twitter.com* with the python package 'Twint'.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Wang, J. (1994). A model of competitive stock trading volume. *Journal of political Economy*, 102(1):127–168.

Wang, Y.-H., Keswani, A., and Taylor, S. J. (2006). The relationships between sentiment, returns and volatility. *International Journal of Forecasting*, 22(1):109–123.

Winck, B. (2020). The us tech sector is now worth more than the entire european stock market, bank of america says. article retrieved from *https://markets.businessinsider.com/news/stocks/us-tech-stocks-worth-more-european-stock-market-apple-microsoft-2020-8-1029545001*.

Yu, J. (2013). A sentiment-based explanation of the forward premium puzzle. *Journal of Monetary Economics*, 60(4):474–491.

Zacharias, C. and Poldi, F. (2021). Twint package. The package has been retrieved under guidance from the GitHub page, *https://github.com/twintproject/twint*.

Zhang, W., Wang, P., Li, X., and Shen, D. (2018). Twitter's daily happiness sentiment and international stock returns: evidence from linear and nonlinear causality tests. *Journal of Behavioral and Experimental Finance*, 18:50–53.

A Epilogue: Reflection

At the beginning of 2021, with the increasing popularity of the Reddit page "WallStreetBets", I became interested in the topic of my study. I started to investigate the possibilities of scraping investor sentiment directly from social media and using this investor sentiment in the field of finance. I do not hold shares or cryptocurrencies of any sort; however, strategies and innovations regarding investing/trading keep me interested since it is considered by a significant part of the material given in the courses of the finance track on the VU. At the start of this thesis, I had some experience programming in Python. Furthermore, I fancied the challenge of programming, strategies based on possible investors sentiment to bring my coding skills to the next level. I searched for various studies and articles to learn the process of scraping, cleaning and analyzing data from Reddit or Twitter to create a proxy for investors sentiment. I stumbled onto so many different methods that I could not see the wood for the trees anymore. This created a feeling of impotence, as many things in life can do when you dive deep into new material.

However, because I knew the subject's potential for my upcoming thesis, I decided to keep reading the information in the articles about investors sentiment and associated Python code. This created a clear basis for this study before I began, marking the possibilities and barriers of the methods early on. Once I started with the thesis in April, the purpose of my thesis was clear, and the subject was stated. As I found multiple studies regarding the U.S. stock market but limited evidence on the European stock market. Therefore, I decided to focus my thesis on the most significant European indices. From the start, my supervisor, Lowie, told me it would be a challenge and a large thesis, not entirely appropriate for a bachelor thesis. My first response was that this was the idea since I was eager to finally write something that could have significant meaning, opposed to the minor assignments of previous courses where I needed to externally motivate myself.

A couple weeks later, I began to understand his worry and notation as I came to realize the involved barriers and options I had to take into account. I think this was the first learning moment of my thesis. Creating realistic goals and determining all the accompanied aspects of a potential study beforehand. Until this thesis, I had little experience with a project of such size. This made me realize the importance of all the future factors that should be considered if you want to avoid barriers in the analyses. In the case of this study, this was incorporating tweets about individual companies in the European indices. Prior to the analyses, I thought I could easily include these tweets and create a general sentiment score of both index and individual company-related tweets. However, as I progressed further, I found it very difficult to determine a reasonable procedure to calculate a general sentiment score. I proceeded with the study by only including index related tweets. This significantly influenced the research as I am not fully able to conclude if this sentiment is delayed. This still leaves a part of the empirical gap regarding the European indices and frustrated me since I had not enough time to adjust the study methods but knew I could have done it differently. However, it created a useful educational experience as I should be more careful at the start of an academic project.

During my bachelor, I have stumbled against the same stone multiple times since I have the urge to create a project of significance but end up with a report that is too big and not carefully written. This thesis emphasizes this experience making me cautious and motivated for my following paper to change things in my approach. During this thesis, I have learned multiple things besides my above-mentioned pitfall. Since this thesis has enabled students to carefully investigate a topic associated with your thesis, I have grown in quantitative and programming skills. The programming skills were essential for scraping, cleaning and processing the data and made for many of the small hurdles in the method section of this study, simultaneously creating many successful moments when overcoming these hurdles. This has made me more excited about programming and incorporating this skill into other branches in the field of finance. The quantitative part made me realize how unlimited and interesting statistical analyses can be. I have followed multiple mandatory courses and even some additional courses about quantitative research; however, I have still spent many hours investigating the appropriate methods to produce valid results. I am sure that an experienced quantitative academic could still comment on the methods and assumptions I have used; nonetheless, I have taken almost every measure to understand every quantitative part of my study to validly apply this in the thesis.

Altogether I am pleased with my thesis and the experiences it has brought me. It is frustrating to see the limitations in my study, which could have been prevented beforehand; however, this gives me great motivation to continue this research during the summer to fill my research limitations and fully exclude (or find) the effect that multiple U.S. studies are finding. Concluding that even though I might have bitten off more than I could chew, it also resulted in a hunger for more, which I think is the purpose of writing a thesis. To both learn the standards of writing academic papers and see the potential and importance of academic papers. I guess every cloud has a silver lining.

B Statistics and Results

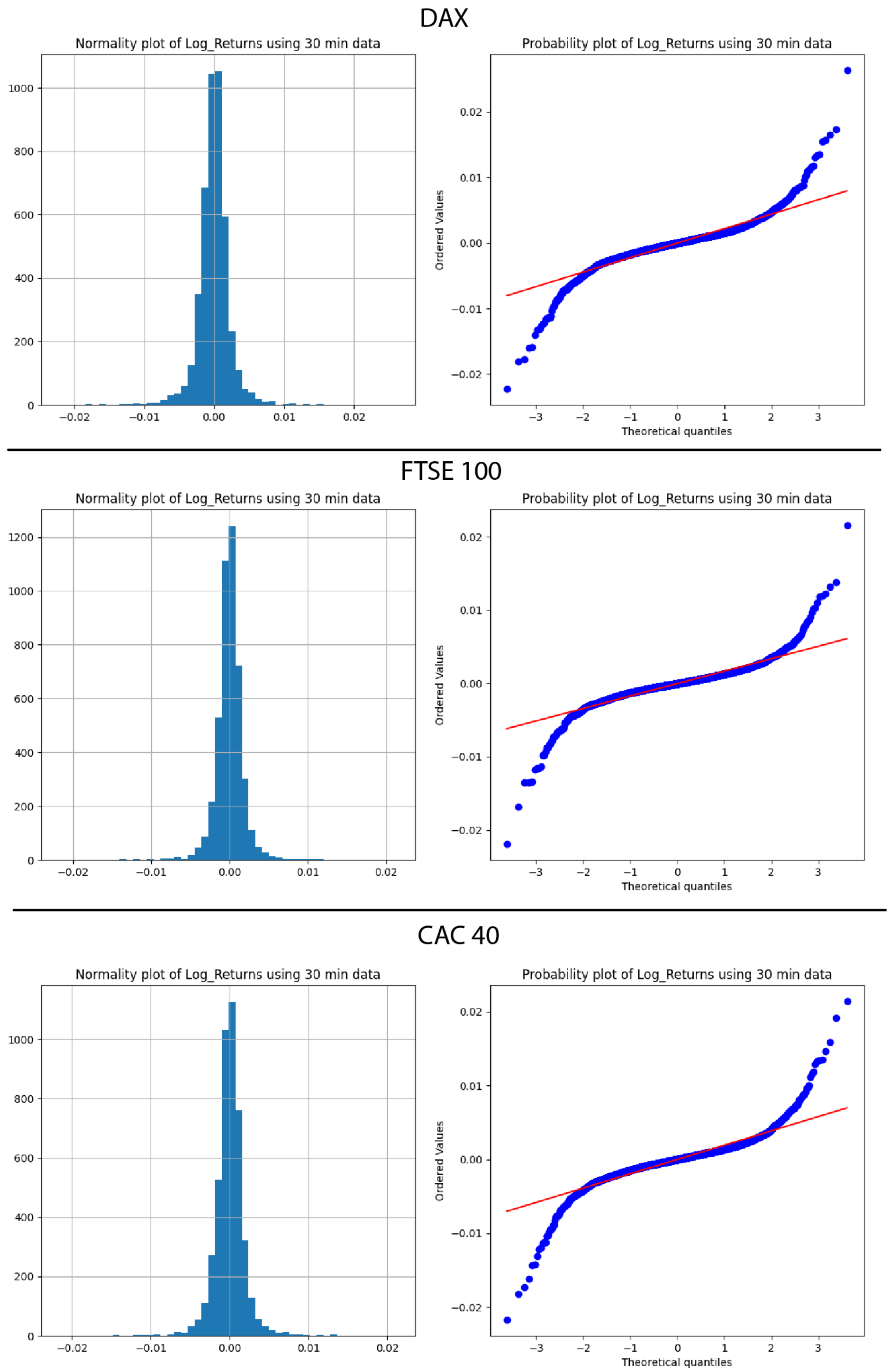

Figure 7: Distribution of the 30 minute DAX Log Returns (left panel) and the probability plot (right panel).

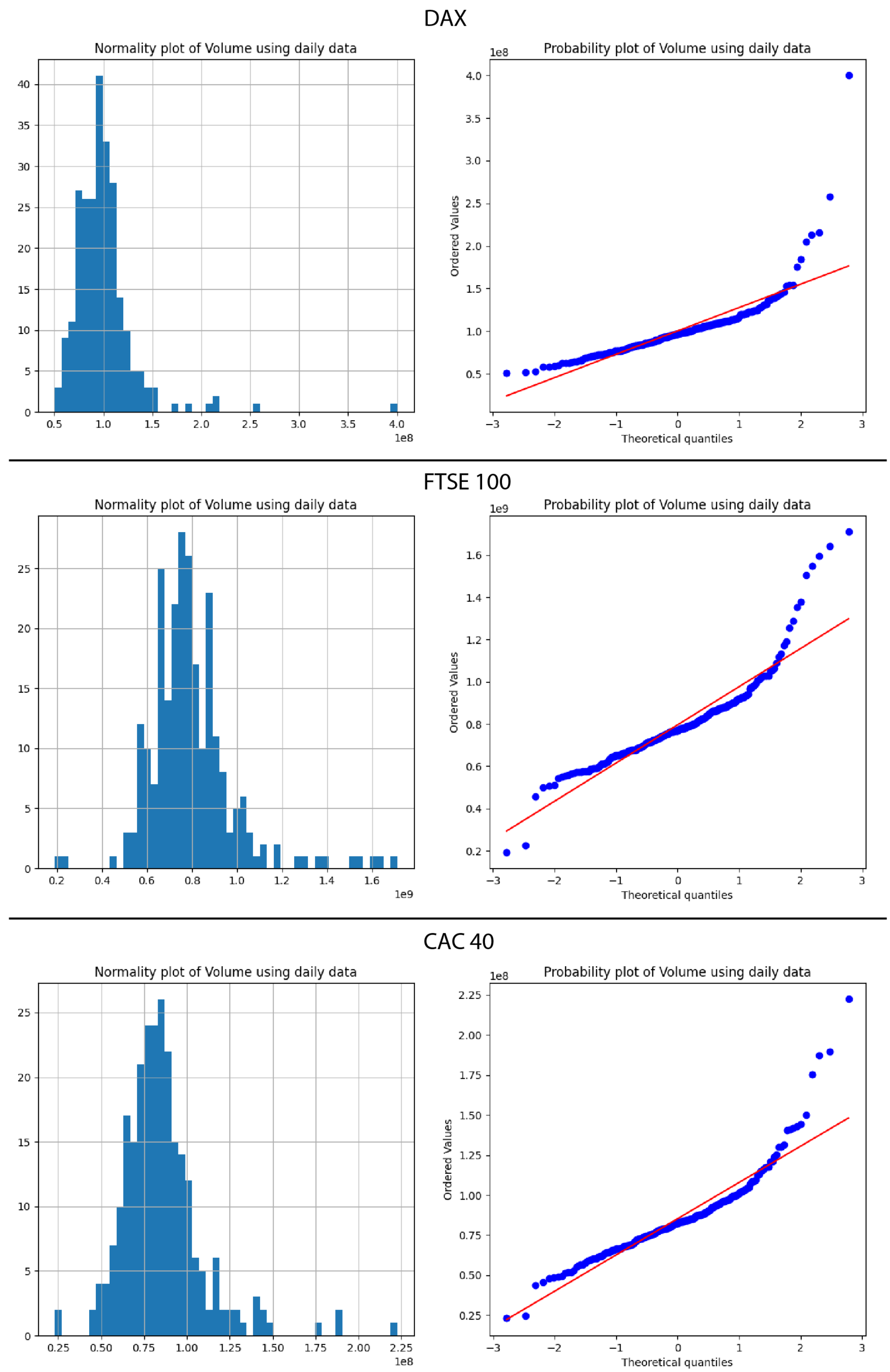

Figure 8: Distribution of the daily Trading Volumes of the European index (left) and the probability plot (right panel).

Table 11: Descriptive statistics of the DAX data set

DAX Daily Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Volume	Log_Returns
count	251	251	251	251	251	251	251	251
mean	0.05	133.40	12283.66	12346.62	12199.62	12269.59	100287529.1	-0.00078
std	0.04	31.18	656.25	651.98	664.90	664.22	32521795.86	0.0098
min	-0.07	62.00	10452.02	10586.80	10279.20	10381.51	50769600	-0.03537
25%	0.02	116.00	11992.39	12041.00	11883.12	11964.95	83117900	-0.00636
50%	0.05	129.00	12389.46	12444.35	12316.35	12384.49	97093400	-0.00043
75%	0.08	150.50	12690.56	12769.58	12616.56	12688.44	109705250	0.00588
max	0.27	257.00	13577.14	13596.89	13517.81	13559.60	400165400	0.02863

DAX 30-Minute Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Log_Returns
count	4510	4510	4510	4510	4510	4510	4510
mean	0.058	4.35	12279.14	12294.15	12262.78	12278.36	-0.00004
std	0.190	3.19	653.09	651.71	654.94	653.52	0.00236
min	-0.878	0.00	10302.13	10348.12	10279.20	10304.10	-0.02226
25%	-0.016	2.00	11962.69	11981.01	11944.75	11960.57	-0.00108
50%	0.013	4.00	12384.74	12399.10	12370.88	12384.58	0
75%	0.162	6.00	12712.26	12728.31	12700.78	12712.01	0.00102
max	0.939	26.00	13580.14	13596.89	13572.76	13579.55	0.02633

DAX 60-Minute Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Log_Returns
count	2255	2255	2255	2255	2255	2255	2255
mean	0.061	8.69	12279.45	12301.31	12255.14	12277.90	-0.00009
std	0.141	5.30	652.96	651.07	655.57	653.81	0.00337
min	-0.674	0.00	10302.13	10365.59	10279.20	10304.10	-0.02051
25%	-0.015	5.00	11965.03	11987.72	11939.04	11963.02	-0.00161
50%	0.057	8.00	12384.80	12404.03	12365.11	12384.49	0.00002
75%	0.144	11.00	12714.27	12735.09	12692.18	12713.52	0.0015
max	0.891	44.00	13580.14	13596.89	13572.10	13579.55	0.0258

Table 12: Descriptive statistics of the FTSE 100 data set

FTSE 100 Daily Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Volume	Log_Returns
count	253	253	253	253	253	253	253	253
mean	0.084	170.27	7366.68	7404.18	7326.62	7362.89	795958963.6	-0.001
std	0.043	38.30	304.88	298.74	311.38	306.83	192951444.4	0.008
min	-0.039	106.00	6584.70	6721.20	6536.50	6584.70	192808000	-0.032
25%	0.054	147.00	7105.30	7162.60	7062.10	7103.80	679057900	-0.005
50%	0.083	162.00	7418.30	7439.60	7358.00	7398.90	769512800	0.000
75%	0.114	186.00	7652.90	7685.70	7615.60	7651.30	872509100	0.004
max	0.218	487.00	7877.50	7903.50	7854.60	7877.50	1711560200	0.023

FTSE 100 30-Minute Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Log_Returns
count	4537.000	4537	4537	4537	4537	4537	4537
mean	0.085	5.95	7368.97	7376.17	7361.56	7368.87	-0.00003
std	0.183	3.67	301.61	300.40	302.94	301.72	0.00184
min	-0.796	0.00	6546.20	6573.56	6536.53	6546.21	-0.02191
25%	-0.005	3.00	7117.91	7128.10	7106.73	7117.93	-0.00083
50%	0.074	5.00	7397.94	7405.08	7388.46	7393.83	-0.00003
75%	0.188	8.00	7649.49	7655.46	7644.37	7649.73	0.00079
max	0.981	51.00	7894.28	7903.50	7893.04	7896.69	0.02158

FTSE 100 60-Minute Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Log_Returns
count	2269.000	2269	2269	2269	2269	2269	2269
mean	0.087	11.92	7369.28	7380.26	7357.79	7368.85	-0.0001
std	0.132	6.11	301.37	299.65	303.46	301.60	0.0026
min	-0.429	0.00	6546.20	6591.90	6536.53	6546.21	-0.0167
25%	0.004	8.00	7120.36	7132.15	7101.01	7119.63	-0.0013
50%	0.085	11.00	7402.47	7418.99	7385.73	7398.87	0.0000
75%	0.167	15.00	7647.98	7658.77	7640.94	7647.68	0.0012
max	0.713	88.00	7894.28	7903.50	7884.19	7894.27	0.0218

Table 13: Descriptive statistics of the CAC 40 data set

CAC 40 Daily Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Volume	Log_Returns
count	255	255	255	255	255	255	255	255
mean	0.049	53.07	5297.45	5321.81	5267.99	5293.62	85230644.71	-0.0005
std	0.040	9.81	214.66	211.83	221.12	219.09	24039911.30	0.0087
min	-0.057	28.00	4641.05	4664.20	4555.99	4598.61	23249900.00	-0.0337
25%	0.021	46.00	5148.71	5176.66	5115.63	5148.32	71796150.00	-0.0054
50%	0.050	54.00	5344.55	5366.56	5322.77	5344.26	82529900.00	-0.0001
75%	0.075	58.00	5472.83	5491.81	5448.60	5473.34	94334300.00	0.0050
max	0.164	87.00	5637.94	5657.44	5628.93	5640.10	222510800.00	0.0259

CAC 40 30-Minute Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Log_Returns
count	4570.000	4570	4570	4570	4570	4570	4570
mean	0.038	1.59	5298.42	5303.81	5292.62	5298.20	0.000
std	0.177	1.30	212.73	211.84	213.92	212.99	0.002
min	-0.873	0.00	4563.14	4584.13	4555.99	4564.07	-0.022
25%	0.000	1.00	5151.94	5159.43	5144.10	5152.24	-0.001
50%	0.000	1.00	5347.93	5352.44	5342.97	5347.83	0.000
75%	0.000	2.00	5470.51	5476.22	5465.39	5470.75	0.001
max	0.957	10.00	5655.62	5657.44	5651.83	5655.98	0.021

CAC 40 60-Minute Data Set

	Sentiment_Score	Tweet_Volume	Open	High	Low	Close	Log_Returns
count	2285.000	2285	2285	2285	2285	2285	2285
mean	0.046	3.18	5298.62	5306.57	5289.94	5298.20	0.000
std	0.172	2.09	212.41	211.16	214.23	212.91	0.003
min	-0.796	0.00	4563.14	4594.31	4555.99	4564.07	-0.018
25%	0.000	2.00	5151.75	5162.36	5140.77	5152.12	-0.001
50%	0.000	3.00	5348.49	5354.50	5341.06	5348.87	0.000
75%	0.113	4.00	5470.58	5478.00	5463.46	5470.77	0.001
max	0.957	13.00	5653.35	5657.44	5651.42	5653.16	0.022

Table 14: Normality test on the interested variables of the European indices

	DAX		**FTSE 100**		**CAC 40**	
			30 Minute			
	Stat	P-value	Stat	P-value	Stat	P-value
Sentiment_Score	302.708	0.000	226.414	0.000	936.476	0.000
Tweet_Volume	1280.978	0.000	1863.605	0.000	761.428	0.000
Log_Returns	922.612	0.000	1135.983	0.000	1033.176	0.000
			60 Minute			
Sentiment_Score	115.505	0.000	73.706	0.000	301.686	0.000
Tweet_Volume	643.426	0.000	1004.537	0.000	211.414	0.000
Log_Returns	346.059	0.000	376.426	0.000	357.931	0.000
			Daily			
Sentiment_Score	15.882	0.000	0.539	0.764	1.559	0.459
Tweet_Volume	30.299	0.000	201.967	0.000	4.880	0.087
Volume	260.523	0.000	99.281	0.000	113.452	0.000
Log_Returns	6.324	0.042	11.185	0.004	9.883	0.007

Note: The test statistics are computed using the Jarque-Bera test.

Table 15: ADF tests - DAX data

			Daily Interval			
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	245	-3.444	0.010	-3.457	-2.873	-2.573
Tweet_Volume	248	-5.920	0.000	-3.457	-2.873	-2.573
Close	245	-0.407	0.909	-3.457	-2.873	-2.573
Volume	245	-4.816	0.000	-3.457	-2.873	-2.573
Log_Returns	246	-8.333	0.000	-3.457	-2.873	-2.573
			30-Minute Interval			
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	4505	-27.542	0.000	-3.432	-2.862	-2.567
Tweet_Volume	4489	-8.582	0.000	-3.432	-2.862	-2.567
Close	4509	-0.414	0.908	-3.432	-2.862	-2.567
Log_Returns	4490	-15.514	0.000	-3.432	-2.862	-2.567
			60-Minute Interval			
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	2234	-7.645	0.000	-3.433	-2.863	-2.567
Tweet_Volume	2228	-4.683	0.000	-3.433	-2.863	-2.567
Close	2245	-0.245	0.933	-3.433	-2.863	-2.567
Log_Returns	2246	-17.057	0.000	-3.433	-2.863	-2.567

Table 16: ADF tests - FTSE 100 data

	Daily Interval					
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	251	-8.259	0.000	-3.457	-2.873	-2.573
Tweet_Volume	250	-5.544	0.000	-3.457	-2.873	-2.573
Close	252	-0.938	0.775	-3.457	-2.873	-2.573
Volume	251	-8.436	0.000	-3.457	-2.873	-2.573
Log_Returns	252	-17.391	0.000	-3.457	-2.873	-2.573
	30-Minute Interval					
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	4529	-21.320	0.000	-3.432	-2.862	-2.567
Tweet_Volume	4504	-7.673	0.000	-3.432	-2.862	-2.567
Close	4536	-0.862	0.800	-3.432	-2.862	-2.567
Log_Returns	4536	-66.435	0.000	-3.432	-2.862	-2.567
	60-Minute Interval					
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	2260	-13.435	0.000	-3.433	-2.863	-2.567
Tweet_Volume	2242	-4.992	0.000	-3.433	-2.863	-2.567
Close	2268	-0.892	0.791	-3.433	-2.863	-2.567
Log_Returns	2268	-46.358	0.000	-3.433	-2.863	-2.567

Table 17: ADF tests - CAC 40 data

	Daily Interval					
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	253	-7.532	0.000	-3.456	-2.873	-2.573
Tweet_Volume	250	-2.918	0.043	-3.457	-2.873	-2.573
Close	254	-0.635	0.863	-3.456	-2.873	-2.573
Volume	252	-6.478	0.000	-3.457	-2.873	-2.573
Log_Returns	254	-15.336	0.000	-3.456	-2.873	-2.573
	30-Minute Interval					
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	4541	-11.496	0.000	-3.432	-2.862	-2.567
Tweet_Volume	4537	-5.983	0.000	-3.432	-2.862	-2.567
Close	4569	-0.660	0.857	-3.432	-2.862	-2.567
Log_Returns	4568	-46.763	0.000	-3.432	-2.862	-2.567
	60-Minute Interval					
	Number of obs	t-test	p-value	1%	5%	10%
Sentiment_Score	2276	-14.042	0.000	-3.433	-2.863	-2.567
Tweet_Volume	2258	-3.191	0.021	-3.433	-2.863	-2.567
Close	2284	-0.659	0.857	-3.433	-2.863	-2.567
Log_Returns	2283	-34.527	0.000	-3.433	-2.863	-2.567

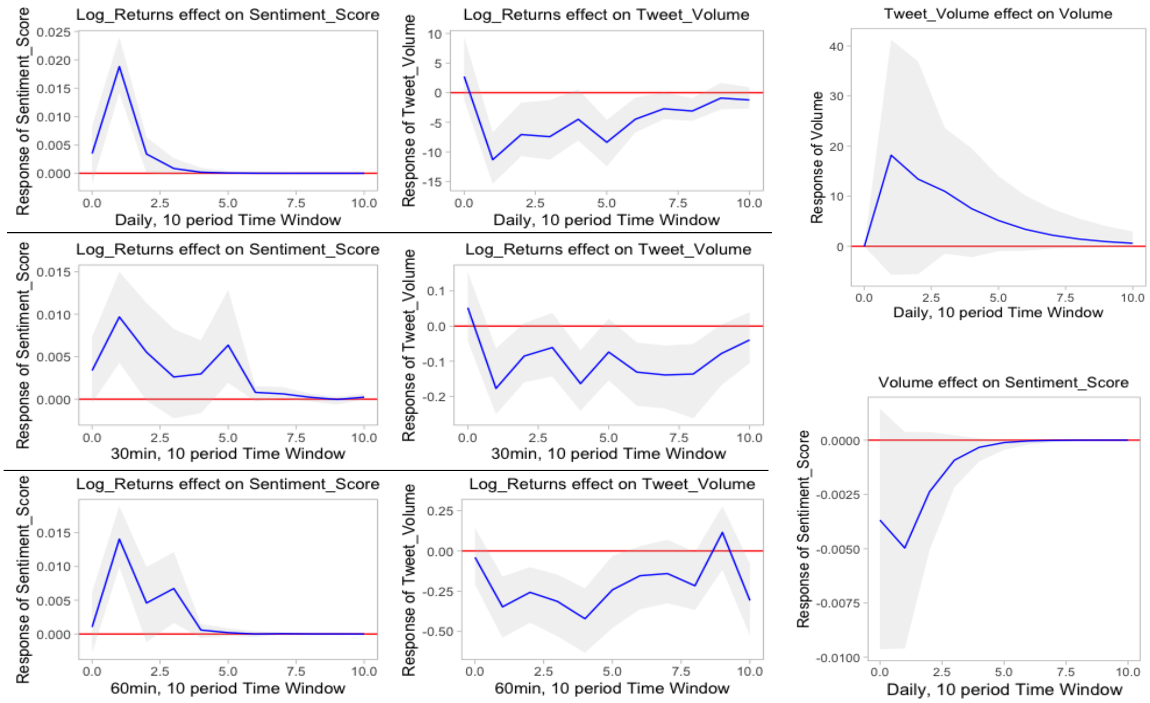

Figure 9: Impulse Response Function FTSE 100 index, using a 10 day forecast

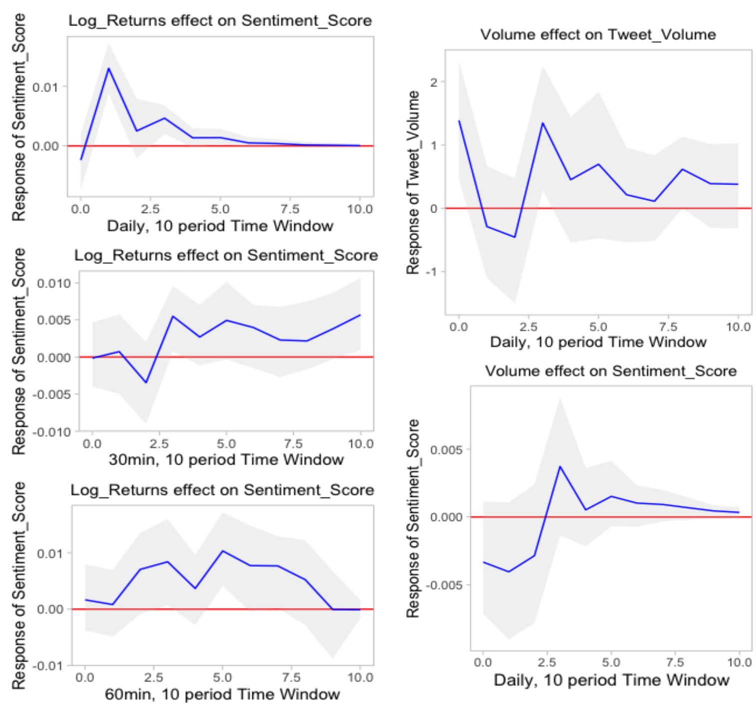

Figure 10: Impulse Response Function CAC 40 index, using a 10 day forecast

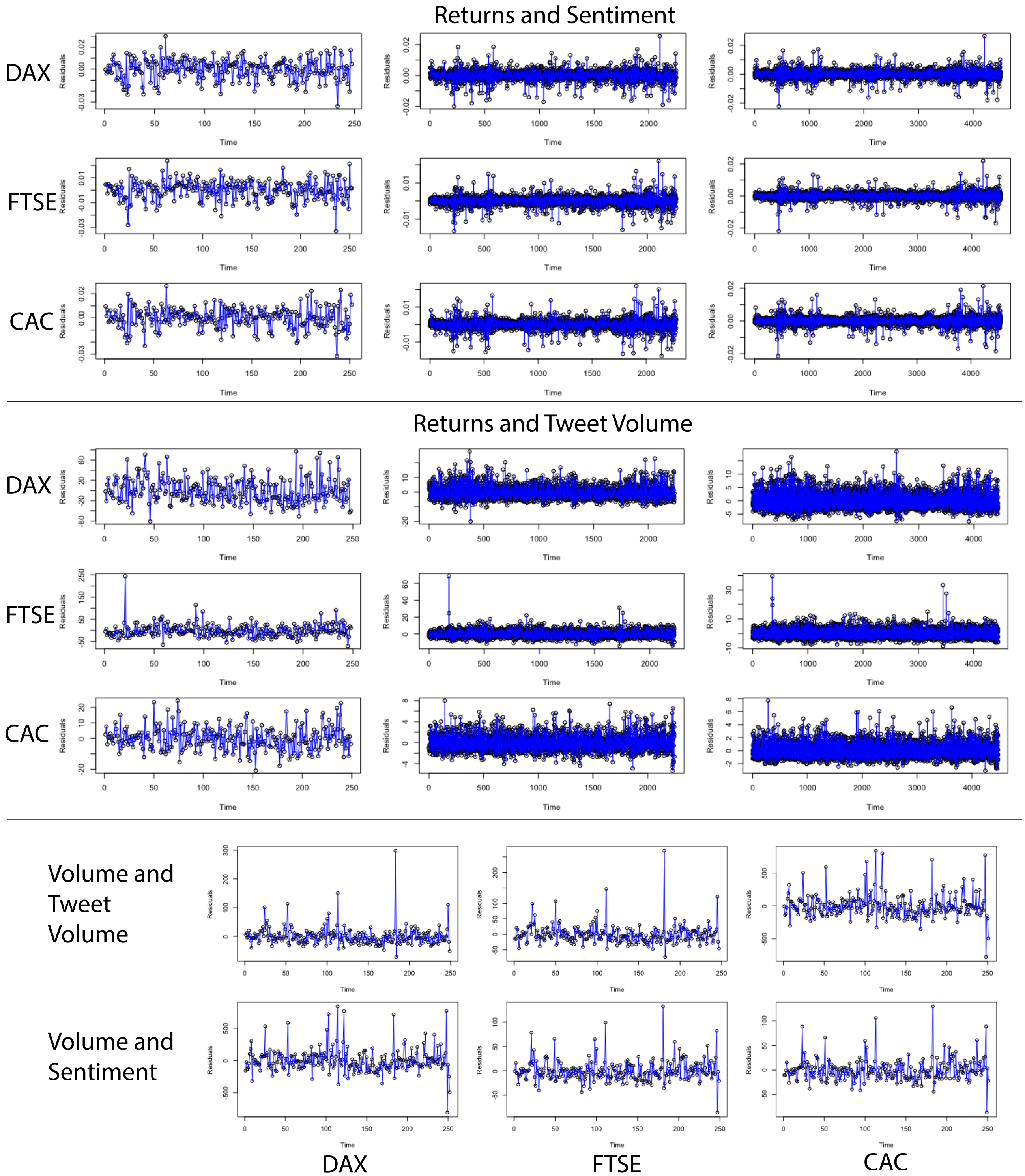

Figure 11: Residuals regarding all 24 VAR models

Table 18: VAR models: DAX Log Returns and Sentiment Score

	Daily variables:		*30Min variables:*		*60Min variables:*	
	(Returns)	(Sentiment)	(Returns)	(Sentiment)	(Returns)	(Sentiment)
Log_Returns.l1	−0.072	1.363***	0.009	2.109*	0.030	1.758**
	(0.065)	(0.247)	(0.015)	(1.192)	(0.021)	(0.877)
Sentiment_Score.l1	−0.004	0.157**	−0.0003	0.023	−0.001*	0.087***
	(0.017)	(0.064)	(0.0002)	(0.015)	(0.001)	(0.021)
Log_Returns.l2	0.147**	0.144	0.008	2.224*	−0.052**	2.094**
	(0.069)	(0.263)	(0.015)	(1.192)	(0.021)	(0.879)
Sentiment_Score.l2	−0.017	0.225***	0.00003	0.024	0.001**	0.018
	(0.016)	(0.063)	(0.0002)	(0.015)	(0.001)	(0.021)
Log_Returns.l3	0.074	0.048	0.004	2.097*	−0.009	1.556*
	(0.068)	(0.259)	(0.015)	(1.192)	(0.021)	(0.880)
Sentiment_Score.l3	0.011	0.116*	−0.0001	0.053***	−0.001*	0.034
	(0.017)	(0.063)	(0.0002)	(0.015)	(0.001)	(0.021)
Log_Returns.l4	−0.158**	−0.263	−0.027*	2.879**	−0.007	−1.125
	(0.067)	(0.258)	(0.015)	(1.192)	(0.021)	(0.879)
Sentiment_Score.l4	0.021	0.186***	0.0001	−0.012	−0.0001	0.033
	(0.016)	(0.060)	(0.0002)	(0.015)	(0.001)	(0.021)
Log_Returns.l5			0.006	1.484	−0.022	2.349***
			(0.015)	(1.193)	(0.021)	(0.878)
Sentiment_Score.l5			0.0002	0.044***	0.0002	0.038*
			(0.0002)	(0.015)	(0.001)	(0.021)
const	−0.002	0.017***	−0.00004	0.051***	−0.00005	0.049***
	(0.001)	(0.005)	(0.00004)	(0.003)	(0.0001)	(0.004)
Observations	246	246	4,505	4,505	2,250	2,250
R^2	0.055	0.343	0.002	0.010	0.009	0.023
Adjusted R^2	0.023	0.321	−0.0004	0.008	0.004	0.019
Residual Std. Error	0.010 (df = 237)	0.037 (df = 237)	0.002 (df = 4494)	0.189 (df = 4494)	0.003 (df = 2239)	0.140 (df = 2239)
F Statistic	1.736* (df = 8; 237)	15.461*** (df = 8; 237)	0.835 (df = 10; 4494)	4.709*** (df = 10; 4494)	1.977** (df = 10; 2239)	5.297*** (df = 10; 2239)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 19: VAR models: DAX Log Returns and Tweet Volume

	Daily variables:		*30Min variables:*		*60Min variables:*	
	(Returns)	(Tweet Volume)	(Returns)	(Tweet Volume)	(Returns)	(Tweet Volume)
Log_Returns.l1	−0.092	−770.993***	0.011	−55.004***	0.033	−137.261***
	(0.064)	(165.993)	(0.015)	(18.223)	(0.021)	(29.093)
Tweet_Volume.l1	−0.00003	0.321***	0.00001	0.213***	0.00002	0.196***
	(0.00002)	(0.064)	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l2	0.100	−210.409	0.013	−61.012***	−0.052**	−49.140*
	(0.066)	(171.687)	(0.015)	(18.239)	(0.021)	(29.257)
Tweet_Volume.l2	0.00004	0.206***	−0.00000	0.078***	0.00000	0.015
	(0.00003)	(0.066)	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l3	0.078	−145.937	0.008	−0.495	−0.009	−22.839
	(0.067)	(172.096)	(0.015)	(18.260)	(0.021)	(29.251)
Tweet_Volume.l3	0.00003	0.139**	0.00000	0.010	−0.00000	0.006
	(0.00002)	(0.063)	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l4			−0.031**	−31.689*	−0.008	−51.479*
			(0.015)	(18.232)	(0.021)	(29.241)
Tweet_Volume.l4			0.00003**	0.012	−0.00001	−0.015
			(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l5			0.005	−8.798	−0.019	−58.420**
			(0.015)	(18.241)	(0.021)	(29.258)
Tweet_Volume.l5			−0.00002	−0.021	0.00001	0.024
			(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l6			−0.012	9.883	−0.007	−35.858
			(0.015)	(18.231)	(0.022)	(29.413)
Tweet_Volume.l6			0.00001	0.018	−0.00001	0.024
			(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l7			−0.020	−42.984**	0.019	−55.283*
			(0.015)	(18.232)	(0.022)	(29.450)
Tweet_Volume.l7			0.00000	0.003	−0.00001	−0.039*

	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l8	0.008	−27.243	0.017	−3.858
	(0.015)	(18.244)	(0.022)	(29.474)
Tweet_Volume.l8	−0.00001	0.004	−0.00001	0.102***
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l9	0.004	−32.240*	−0.083***	−24.464
	(0.015)	(18.247)	(0.022)	(29.463)
Tweet_Volume.l9	−0.00001	−0.016	0.00000	0.242***
	(0.00001)	(0.015)	(0.00002)	(0.022)
Log_Returns.l10	−0.026*	9.973	0.032	6.475
	(0.015)	(18.261)	(0.022)	(29.463)
Tweet_Volume.l10	−0.00001	0.025	−0.00001	0.003
	(0.00001)	(0.015)	(0.00002)	(0.022)
Log_Returns.l11	−0.006	−27.432	−0.030	9.336
	(0.015)	(18.303)	(0.022)	(29.462)
Tweet_Volume.l11	0.00000	−0.008	−0.00000	−0.060***
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l12	0.020	−43.659**	0.009	−39.365
	(0.015)	(18.306)	(0.022)	(29.456)
Tweet_Volume.l12	0.00001	0.015	−0.00000	−0.007
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l13	−0.017	0.023	0.001	−43.907
	(0.015)	(18.337)	(0.022)	(29.457)
Tweet_Volume.l13	−0.00000	0.003	0.00000	0.007
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l14	−0.011	−15.348	0.008	4.768
	(0.015)	(18.334)	(0.022)	(29.542)
Tweet_Volume.l14	0.00001	−0.020	−0.00000	−0.018
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l15	0.040***	0.683	−0.001	19.740
	(0.015)	(18.342)	(0.022)	(29.550)
Tweet_Volume.l15	−0.00001	0.018	−0.00002	−0.010

	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l16	0.026*	−6.280	−0.007	−55.814*
	(0.015)	(18.349)	(0.022)	(29.558)
Tweet_Volume.l16	−0.00001	0.056***	0.00004**	−0.005
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l17	−0.046***	−3.075	−0.003	28.416
	(0.015)	(18.349)	(0.022)	(29.527)
Tweet_Volume.l17	0.00001	0.055***	−0.00001	0.047**
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l18	−0.026*	−35.628*	0.025	−64.353**
	(0.015)	(18.366)	(0.022)	(29.500)
Tweet_Volume.l18	−0.00002	0.136***	0.00001	0.184***
	(0.00001)	(0.015)	(0.00002)	(0.021)
Log_Returns.l19	−0.028*	4.574		
	(0.015)	(18.360)		
Tweet_Volume.l19	0.00000	0.043***		
	(0.00001)	(0.015)		
Log_Returns.l20	0.029*	33.874*		
	(0.015)	(18.344)		
Tweet_Volume.l20	0.00002	−0.006		
	(0.00001)	(0.015)		
Log_Returns.l21	−0.018	16.263		
	(0.015)	(18.347)		
Tweet_Volume.l21	−0.00001	−0.026*		
	(0.00001)	(0.015)		
Log_Returns.l22	0.001	−28.392		
	(0.015)	(18.337)		
Tweet_Volume.l22	0.00001	−0.014		
	(0.00001)	(0.015)		
Log_Returns.l23	−0.018	−10.256		
	(0.015)	(18.329)		
Tweet_Volume.l23	−0.00001	−0.010		

	(0.00001)	(0.015)
Log_Returns.l24	0.001	−37.452**
	(0.015)	(18.327)
Tweet_Volume.l24	0.00001	0.008
	(0.00001)	(0.015)
Log_Returns.l25	0.017	−28.587
	(0.015)	(18.334)
Tweet_Volume.l25	−0.00001	−0.031**
	(0.00001)	(0.015)
Log_Returns.l26	−0.001	8.996
	(0.015)	(18.342)
Tweet_Volume.l26	−0.00000	0.016
	(0.00001)	(0.015)
Log_Returns.l27	−0.004	1.977
	(0.015)	(18.361)
Tweet_Volume.l27	−0.00000	−0.013
	(0.00001)	(0.015)
Log_Returns.l28	0.003	7.577
	(0.015)	(18.374)
Tweet_Volume.l28	0.00001	0.017
	(0.00001)	(0.015)
Log_Returns.l29	0.011	10.604
	(0.015)	(18.373)
Tweet_Volume.l29	−0.00001	−0.008
	(0.00001)	(0.015)
Log_Returns.l30	0.003	−5.635
	(0.015)	(18.370)
Tweet_Volume.l30	0.00001	−0.016
	(0.00001)	(0.015)
Log_Returns.l31	−0.011	−11.527
	(0.015)	(18.368)
Tweet_Volume.l31	−0.00003**	0.004

			(0.00001)	(0.015)		
Log_Returns.l32			−0.006	−22.777		
			(0.015)	(18.377)		
Tweet_Volume.l32			−0.00000	−0.008		
			(0.00001)	(0.015)		
Log_Returns.l33			−0.003	18.117		
			(0.015)	(18.364)		
Tweet_Volume.l33			0.00004^{***}	−0.006		
			(0.00001)	(0.015)		
Log_Returns.l34			−0.013	−2.325		
			(0.015)	(18.353)		
Tweet_Volume.l34			0.00001	0.032^{**}		
			(0.00001)	(0.015)		
Log_Returns.l35			0.021	−27.650		
			(0.015)	(18.356)		
Tweet_Volume.l35			−0.00000	0.058^{***}		
			(0.00001)	(0.015)		
Log_Returns.l36			−0.002	−17.610		
			(0.015)	(18.356)		
Tweet_Volume.l36			−0.00001	0.114^{***}		
			(0.00001)	(0.015)		
const	-0.006^{*}	43.569^{***}	−0.0001	1.146^{***}	−0.0002	2.589^{***}
	(0.003)	(8.644)	(0.0002)	(0.187)	(0.0003)	(0.426)
Observations	247	247	4,474	4,474	2,237	2,237
R^2	0.050	0.378	0.019	0.214	0.019	0.262
Adjusted R^2	0.026	0.363	0.003	0.201	0.003	0.250
Residual Std. Error	0.010 (df = 240)	24.945 (df = 240)	0.002 (df = 4401)	2.854 (df = 4401)	0.003 (df = 2200)	4.598 (df = 2200)
F Statistic	2.096^{*} (df = 6; 240)	24.317^{***} (df = 6; 240)	1.197 (df = 72; 4401)	16.626^{***} (df = 72; 4401)	1.170 (df = 36; 2200)	21.652^{***} (df = 36; 2200)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 20: VAR models: Trading Volume and Tweet Volume

	Daily variables:			
	(Trading Volume)	(Sentiment)	(Trading Volume)	(Tweet Volume)
Volume.l1	0.223^{***}	-0.0001	0.057	-0.023
	(0.066)	(0.0001)	(0.065)	(0.046)
Sentiment_Score.l1	150.766^{***}	0.166^{***}		
	(51.670)	(0.064)		
Tweet_Volume.l1			0.136	0.377^{***}
			(0.090)	(0.063)
Volume.l2	0.069	-0.0001	0.105	-0.034
	(0.069)	(0.0001)	(0.066)	(0.047)
Sentiment_Score.l2	-3.663	0.237^{***}		
	(52.768)	(0.065)		
Volume.l3	0.075	-0.0001		
	(0.070)	(0.0001)		
Sentiment_Score.l3	-37.476	0.095		
	(53.383)	(0.066)		
Volume.l4	0.051	0.0002^{**}		
	(0.066)	(0.0001)		
Sentiment_Score.l4	-59.637	0.209^{***}		
	(52.494)	(0.065)		
Tweet_Volume.l2			-0.030	0.276^{***}
			(0.090)	(0.063)
const	55.821^{***}	0.022	66.971^{***}	51.911^{***}
	(11.195)	(0.014)	(13.093)	(9.194)
Observations	246	246	248	248
R^2	0.098	0.277	0.029	0.312
Adjusted R^2	0.067	0.252	0.013	0.300
Residual Std. Error	31.361 (df = 237)	0.039 (df = 237)	37.168 (df = 243)	26.098 (df = 243)
F Statistic	3.201^{***} (df = 8; 237)	11.328^{***} (df = 8; 237)	1.833 (df = 4; 243)	27.510^{***} (df = 4; 243)

Note: *p<0.1; **p<0.05; ***p<0.01