RADBOUD UNIVERSITY

Nijmegen School of Management

Master's Thesis

THE EFFECT OF SOCIAL MEDIA SENTIMENT ON DAILY STOCK RETURNS

Author: Ivo Kregting

Student ID: S4797094

Specialization: Financial Economics

Supervisor: Dr. D.J. Janssen

Date: August 15, 2021

Place: Nijmegen, the Netherlands

Abstract

In today's economic world, the influences of sentiment are increasing. Since investor's have adopted social media as source to spread information, thoughts and feelings, the general level of sentiment becomes measurable. This study investigates whether sentiment contains pricing power towards daily stock returns. A sample of Tweets, specifically related to Apple, Amazon, Google, Microsoft and Tesla was used to conduct and compare five different methods of sentiment analysis. It is found that social media sentiment had a significant impact on the same day's stock returns. Accordingly, the inclusion of sentiment variables in a traditional CAPM model significantly increased its explanatory power to determine stock returns. Since there is no consensus on how to appropriately measure sentiment on social media, this study compared five lexicon-based approaches in terms of performance. It is found that field-specific lexicons outperform general lexicons. This finding contributes to the process of creating consensus, and improvement of transparency and replicability in the field of sentiment analysis.

Keywords: Sentiment analysis, Social media, Stock returns, Twitter, Asset pricing

Table of Contents

1.Introduction

As a consequence of the recent 'Reddit Revolution', the financial world has been shocked by the tremendous power of social media to drive stock prices in a desired direction. A large community of investors used the Reddit-platform to organize a coordinated buying action of wisely picked stock targets. GameStop was picked as one of the main targets, because the investor's community discovered that large equity funds kept extremely large short positions in this stock. Forced by the Reddit buying regime, the stock price of GameStop soared more than 1000% in just two weeks leading to a bloodbath among the involved equity funds with large short positions (Smith, 2021). This cautionary tale raises numerous questions with regards to the significant role of social media in the current financial markets. It seems plausible to think that investment decision making is significantly influenced by thoughts and information spread via social media platforms. Therefore, it is reasonable to consider the thoughts of investors on social media as crucial predictor for stock returns.

Early research on the determination of stock prices stated that stock prices are driven by new information flowing into the markets. Fama, Fisher, Jensen, and Roll (1969) constructed the Efficient Market Hypothesis which assumed that stock prices fully reflect all available information in the market. Therefore, investors are seeking for new information to make proper investment decisions (Slovic, Fleissner, & Bauman, 1972). More recent studies argue that investment decision making is significantly affected by personal emotions and feelings (Baker & Wurgler, 2006; Dolan, 2002). Aggregated individual emotions reflect a general level of sentiment in the financial society. This general level of sentiment is found as an important measure of stock markets in general (Baker & Wurgler, 2007; Nofsinger, 2005).

The internet revolution, especially social media, provided a whole new source to measure the general level of sentiment among investors. Jong, Elfayoumy, and Schnusenberg (2017) found that 34% to 70% of all investors use social media to diffuse information, opinions and feelings. The most popular platform among the investment community is Twitter. Investor's use 'hashtags' and 'cashtags' ($AAPL) in their posted Tweets, to initiate that their message contains valuable investment related content. Prior research executed textual sentiment analysis to measure the general level of sentiment in an extracted sample of Tweets. Bollen, Mao, and Zeng (2011) found that different emotional states showed a significant positive relationship with daily stock returns. Changes in public mood, tracked from millions of Tweets, were Granger causative for movements in the Dow Jones Industrial Average. Additionally, Sul, Dennis, and Yuan (2016) suggest that sentiment in Tweets about a specific company acts as predictor for the next day, the next 10 days, and the next 20 days stock returns. Other studies focused on the intraday effects of sentiment on stock returns. Renault (2017) provided indications that sentiment around market opening hours predicts the final half-hour return on the same day. Broadstock and Zhang (2019) are the first to use a traditional CAPM model, to measure the effects of different emotions on stock returns. Their results are inconsistent in terms of varying positive and negative relationships between emotional states and stock returns. Motivated by prior empirical evidence that sentiment on Twitter shows significant relationships with stock returns, this study further contributes to the growing body of research related to the influence of social media on stock returns. Hence, this research will address the following research question:

To what extent does investor sentiment measured on social media affect daily stock returns?

This study uses a sample of company specific Tweets related to Apple, Amazon, Google, Microsoft and Tesla, posted by investors during a two-year period from 2018 to 2019. The gathered sample of Tweets is used to conduct an appropriate sentiment analysis, where five different lexicon-based methods are compared in terms of measurement performance. The constructed sentiment variables are used as predictor for daily stock returns. Furthermore, the pricing power of social media sentiment is measured using an augmented CAPM model on daily time intervals. |

This research extends the current literature on this topic in two different ways. First, this study attributes to the small-scale body of research using sentiment variables in traditional asset pricing models. This contributes to the current economic profession to further embrace sentiment as widely accepted influential factor with regards to stock returns. Secondly, this research contributes on improving replicability and transparency in this field of research. In the current literature there is no consensus on how to properly measure sentiment in social media content. Various studies use inappropriate general lexicons, or difficult, time-consuming and non-transparent machine learning algorithms. Oliveira, Cortez, and Areal (2016) and Renault (2017) constructed field-specific lexicons to improve replicability, transparency and offer an easier time-efficient method to measure sentiment in social media posts. Although, they found significant evidence that both field-lexicons outperform other methods, later research still uses inappropriate and inefficient methods to analyze sentiment. Therefore, this research uses five different lexicons, among which both field-specific lexicons of Oliveira et al. (2016) and Renault (2017), to further examine the differences in measurement performance.

The findings of this research draw attention to the influence of social media sentiment on daily stock returns. It is found that social media sentiment serves as significant predictor for stock returns on the same day. Furthermore, this research found that it makes sense to incorporate sentiment variables in Capital Asset Pricing Models. Those results might encourage the economic profession to further apply sentiment in asset pricing theory. Additionally, the results with regards to the conducted sentiment analysis provided further indications to assume that field-specific lexicons outperform general lexicons. This might encourage further research to use one of those lexicons in order to create consensus when measuring sentiment on social media. Besides, it might incentivize data-analytics software as R and Python to implement a lexicon specifically constructed to measure sentiment on social media. Moreover, this would further improve the accessibility, transparency and replicability of this new important field of research.

This study continues by providing an overview of the most relevant literature in this field of research. Based on this overview, four hypotheses are drawn. Chapter 3 will provide a detailed description of the data collection procedures. Chapter 4 provides a brief understanding of sentiment analysis in general, followed by the conducted sentiment analysis in this study. Chapter 5 presents the research method of the CAPM-analysis, followed by the corresponding results outlined in chapter 6. Lastly, chapter 7 discusses and interprets the results of this paper leading to an overall conclusion. Furthermore, the most important limitations, contributions and foundation for further research will be outlined in chapter 7 as well.

2. Literature review

Stock return prediction

Predicting stock prices is one of the most attractive but also challenging topics in financial research area. It has been found a challenging task as financial data is complex and stock prices are influenced by a lot of parameters (Wang & Lin, 2018). In general, it is even the question if stock markets are predictable? The origin of predicting stock prices can be found in the Efficient Market Hypothesis and random walk theory (Cootner, 1964; Fama et al., 1969). In accordance with the Efficient Market Hypothesis, stock prices fully reflect all available information in the market. Therefore, price changes are driven by ‘new information’ flowing into the markets.

Investors try their best to appropriately process new information and trade quickly to benefit from it. In the market efficiency framework of Grossman and Stilitz (1980) it is assumed that investors will be compensated for the marginal costs of monitoring information sources. However, the flow of ‘new information’ is random and therefore unpredictable. Since stock prices are driven by these random news flows, it seems reasonable to assume stock price changes must be random and unpredictable as well. This idea is defined as a ‘random walk’ which characterizes that price fluctuations are random deviations from previous prices (Bollen et al., 2011; Fama, 1991; Malkiel, 2003). Considering the assumption of a random walk theory, the accuracy of predicted stock prices cannot exceed a level of 50% (Qian & Rasheed, 2007; Walczak, 2001). However, the randomness of stock prices has been extensively investigated leading to numerous critics with regards to the Efficient Market Hypothesis and its associated random walk theory. A growing body of research found that stock prices contain several predictable components and do not follow a random pattern. Analyzing historical returns (Fama & French, 1988), assessing macroeconomic shocks (Gallagher & Taylor, 2002) and combining multiple classifiers (Qian & Rasheed, 2007), all provided evidence that stock markets are predictable to a certain extent.

Critiques on EMH followed by the advent of Behavioral Finance

The main foundation of the Efficient Market Hypothesis can be found in the assumptions of completely efficient markets. Efficient markets are only achievable when investors make decisions based on their rationale (Simon, 1979). This rationality enables investors to properly process information and make appropriate decisions, forcing stock prices to reflect its true fundamental value (Fama, 1991). However, the rationality of humans comes into question as psychological research states that emotions play an important role in decision making (Dolan, 2002). These findings presented in psychological research formed the foundation of the advent of behavioral finance. The behavioral finance framework is defined as an economic school that relinquish the assumption of efficient markets, in which investors make their decisions based on rational thinking (Ritter, 2003). Nofsinger (2005) defined the stock markets as a complex system of human interaction driven by what investors think instead of economic fundamentals. Psychologists argue that human thoughts are consequences of their feelings and emotions. A correlated aggregation of all these individual feelings reflects the general level of 'social mood' in society. Social mood can be defined as sentiment and occurs in three main states: pessimism, neutral or optimism. Based on this hypothesis, Nofsinger (2005) argued that sentiment determines investor decision making, and since the stock market reflects investment decisions, it is a direct estimation of sentiment. Since it is no longer a clue if investor sentiment affects the stock market, the question rises how to properly measure public sentiment.

Normally, the well-known financial data consist of numbers easy to quantify. However, as sentiment is defined as a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something, it is tough to measure and quantify its effects on the stock prices (Renault, 2017). In prior research many different methods have been used to first identify investor sentiment, measure it properly and obtain a quantifiable variable. Baker and Wurgler (2007) provided a theoretical review of proxies that explicitly measure sentiment. Subsequently, they constructed a sentiment index composed by the following five 'market data' proxies: dividend premiums, the equity shares in new issues, closed-end fund discounts, the number of first day IPO returns and trading volume. The empirical results showed that sentiment waves have clearly observable effects on individual stocks and the stock market in general. However, in terms of predicting stock prices they found a negative relationship between the sentiment index and returns, implying that high sentiments in previous months results in significantly lower returns next months.

Other research focused on traditional media sources to measure sentiment (Tetlock, 2007), or conducted surveys among investors (Baker & Cliff, 2005). Baker and Cliff (2005) initiated a positive relationship between mispricing in the market and sentiment, implying that returns over longer time horizons are negatively related to investor sentiment. This finding is theoretically substantiated as prices tend to revert to their fundament values over multiyear time horizons, while in the short run excessive optimism or pessimism drives pricing above their fundamental values. Overestimated prices will be corrected as prices revert to their intrinsic values, indicating that periods of high sentiment precede low returns (Baker & Cliff, 2005; Verma, Baklaci, & Soydemir, 2008). Overall, the first proxies to measure investor sentiment show similar results whereas sentiment is negatively correlated with stock returns over longer time horizons. In addition, the relationship is the opposite on shorter time horizons.

Social media and emotion

The current 'Social Revolution' where our environment in going through exploits new opportunities in measuring investor sentiment. Social media is currently integrated in many features of our daily lives (Ellison, 2007). It has grown to a general platform for sharing opinions, emotions, information and other thoughts about any particular topic. Also, investors have adopted the use of social media platform in their decision making (Oh & Sheng, 2011). The investor related content provided on social media varies from thoughts, feelings and opinions shared by amateur investors, to high quality analyses elaborated by professional finance analysts.

Twitter is determined as one of the most popular social media platforms amongst investors. The platform allows users to post messages with a maximum of 140 characters, so called tweets (Sul et al., 2016). This seems short but aggregating millions of messages containing information, feelings and opinions of investors may provide an applicable expression of general investors sentiment (Bollen et al., 2011; Sul et al., 2016). Extracting millions of investor related tweets seems a new reasonable approach to measure and quantify investor sentiment. Inspired by this hypothesis, Bollen et al. (2011) extracted ten million tweets and examined six different emotional conditions (alert, calm, happy, kind, vital and sure) in each tweet. They used a Granger-causality analysis to identify correlations between each mood state and the Dow Jones returns. Especially the emotional condition of 'calmness' showed a significantly positively correlation with the Dow Jones Index several days later. Accordingly, they used a SFONN non-linear model to predict daily directional movements of the Dow Jones Index based on the 'calm' emotional state, providing an accuracy of 86.7%. Zhang, Fuehres, and Gloor (2011) found similar results for the mood-conditions 'fear' and 'hope' in correlation with movements in the S&P500, Dow Jones and NASDAQ.

However, both studies provided a significantly correlation between emotional states measured in tweets and movements in stock markets, they are limited as both tweet-samples consists of randomly gathered tweets. Since these samples may contain many tweets without any relation towards the stock markets, it remains questionable if stock-specific content on Twitter is associated with stock returns (Sprenger, Tumasjan, Sandner, & Welpe, 2014). Secondly, both studies only identified the effects of separated emotional states rather than a general measurement of sentiment. Many researches respond to both limitations by gathering tweets, specifically related to a certain stock, and replacing emotional states by a general sentiment index. Oh and Sheng (2011) measured sentiment as a daily positive, neutral or negative index when dividing 'bullish' classified messages by 'bearish' classified messages. All messages are aggregated on daily basis leading to a general perception of positive (>0), negative (<0) or neutral (0) sentiment. The 5-day moving average of this sentiment index showed significant positive correlation with directional movements of stock prices.

Sprenger et al. (2014) and Smailovic, Grcar, Lavrac, & Znidarsic, (2013) used a similar sentiment index to quantify investor sentiment. They both found that a sentiment index contains predictive value for stock returns several days later. Following these studies many others investigated the effects of sentiment on daily stock returns. Risius, Akolk, and Beck (2015) investigated the connection between stock movements and emotions based on 5,5 million tweets related to 33 specific companies. Subsequently, they conducted a time frame of three months including a lagged fixed-effects panel regression. The results of their study provide three key aspects; emotional conditions show stronger correlation with company specific stock prices than the average market sentiment factors, negative emotions have more predictive power than positive emotions and finally the strength of emotions accounts for price movements. Sul, Dennis, and Yuan (2014) use a Cumulative Abnormal Returns (CAR) model to examine the return predictability They investigated the effect of positive and negative sentiment on the S&P500 using three periods: the same day, next day and 10 days later. The coefficients of the same day and 10-day returns are significantly positive and almost equal, indicating that sentiment has an equal impact on stock returns the same day as 10 days ahead. Overall, it seems reasonable to assume that general sentiment or different emotional states are associated with stock returns.

Lagged effects of sentiment

Reviewing the aforementioned studies provides insights that sentiment is associated with stock returns on the same day and several days later (Bollen et al., 2011; Smailovic et al., 2013; Sprenger et al., 2014; Sul et al., 2014;). An explanation of this lagged effect can be found in the cognitive biases of investors and the speed of information diffusion via Twitter. Following an alternative view on the Efficient Market Hypothesis, the Gradual Information Flow (GIF) model assumes that investors have cognitive biases, limiting their abilities to properly act on all available information (Hong & Stein, 2007). Therefore, investors's will overlook relevant 'new information' leading to a slower incorporation of new information in stock prices. In line with this hypothesis, GIF states that the speed of new information flowing through the investors community captures how quickly a stock price will incorporate new information.

Sul et al. (2016) investigated this assumption related to the diffusion of information, and thus sentiment, via Twitter. Information diffusion via Twitter relates on the number of followers and retweets. They assumed that information about stocks which spreads quickly through a platform as Twitter is incorporated quickly into the prices as well, for instance on the same day. Secondly, sentiment which flows slower trough social media takes longer to be incorporated into stock prices, leading to more predictive power over future days. To measure this assumption Sul et al. (2016) used a sample of tweets posted by users with few followers[1]. This described sample is used to measure sentiment, which is included as sentiment variable in a Cumulative Abnormal Return model. The results showed significant positive coefficients for each time window; next day, next to 10th-day and next to 20th-day. Since the coefficients increased over the defined time-windows, they indicate that sentiment based on information flowing slowly through Twitter provides the strongest effects on stock returns further ahead. Thus, based on the studies of Smailovic et al. (2013), Sprenger et al. (2014), and

[1] 'A few followers' is defined as user accounts with less than 171 followers. The number of 171 followers is chosen as this was the median of followers in their initial sample (Sul et al., 2016).

Sul et al. (2016), lagged sentiment effects show significant positive correlations with daily stock returns.

Contrary to previous findings of a positive relationship between daily lagged-sentiment and stock returns, Broadstock and Zhang (2019) found some significant negative coefficients of lagged-sentiment as predictor of stock prices. However, these effects are measured on a 30-minute interval timeframe and thereby classified as incidentally significant noise. The coefficients are inconsistent and there is no specific reasonable explanation given by Broadstock and Zhang (2019). They assume those negative coefficients of lagged-sentiment as potential signals of temporary mispricing. Although Broadstock and Zhang (2019) classified those negative coefficients as 'incidentally significant noise', it could imply signs of a relationship identified in early research. As previously defined Baker and Cliff (2005) and Verma et al. (2008) found a significant relationship where investors overreact to sentiment in the short run causing a reverting effect in the long run. This relationship was measured with monthly time-intervals over multiyear time horizons and therefore not linkable to the negative significant coefficients of Broadstock and Zhang (2019).

However, with the advent of social media, sentiment will flow faster through the investment community, causing a potential overreaction followed by reverting or correction effect on short time horizons. The study of Mo, Liu, and Yang (2016) found this kind of correction effect on daily time basis. They measured sentiment using news articles published on the Internet and found that the coefficients of 2-5 sentiment lags were all negative. At lag-5 the results were significant ($p < 0.05$) for all four index samples, respectively, SPY, DJIA, QQQ and IWV. Interpreting this result indicates a correction of investors overreaction after five trading days. The theory of over-and underreaction of investors with regards to market parameters is generally accepted in economic profession. Therefore, it seems reasonable to investigate if lagged sentiment effects are positively related to stock returns, indicating a slower incorporation of sentiment over time, or negatively related to stock returns, indicating a correction of investors overreaction.

Sentiment analysis with regards to social media

This paragraph elaborates on the fierce debate in the current literature with regards to the methodology behind sentiment analyses in finance. Textual sentiment analysis is a necessary process to convert a qualitative variable, e.g., a newspaper, a message or a tweet, into a quantitative sentiment variable (Renault, 2017). In the field of financial research two main approaches are used to conduct sentiment analysis: machine learning methods and lexicon-based methods. In short, a lexicon-based method uses a dictionary to classify sentiment in text, while machine learning methods train an algorithm to classify sentiment in texts[2].

[2] A detailed analysis of both methods will be outlined in chapter 4 'Sentiment Analysis'.

In the current literature there is no consensus on how to use both methods appropriately when analysing social media messages. Before the advent of social media many studies measured sentiment in formal written finance papers or news articles published by the traditional media. The Harvard-IV lexicon and the Loughran and MacDonald (2011) lexicon were the commonly used lexicons to measure sentiment in formal financial news articles. (Da, Engelberg, & Gao, 2011; Schumaker & Chen, 2009; Tetlock, Saar-Tsechansky, & Macskassy, 2008). Both lexicons did their job properly, although the performance significantly decreased when measuring sentiment in user-generated content written on informal social media platforms (Loughran & MacDonalds, 2016; Nardo, Petracco-Giudici, & Naltsidis, 2016; Renault, 2017). Consequently, many studies started to developed machine learning algorithms to quantify user-generated messages written on social media platforms (Antweiler & Frank, 2004; Smailovic et al., 2013; Sprenger et al., 2014). The machine learning algorithms were assumed to better perform when measuring sentiment in informal, short texts on social media.

In addition, other research exposed the weaknesses of the existing lexicons when measuring investor-specific content on social media. For example, the Harvard-IV and Loughran & MacDonald (2011) lexicons did not contain words and content used by investors on social media (Nardo et al., 2016). Oliveira, Cortez, and Areal (2016) and Renault (2017) responded by constructing an investor-specific lexicon consisting of only terminology used by investors on social media. Both provided evidence that an investors-specific lexicon outperforms traditional popular lexicons, e.g., Harvard-IV, Loughran & MacDonald, The General Inquirer, SentiStrength and others, in terms of classification accuracy. (Oliveira et al., 2016; Renault, 2017). Subsequently, Renault (2017) even identified an equal classification accuracy between his investor specific lexicon and a supervised machine learning algorithm. Consequently, a logical inference would be that these lexicons become widely accepted as proper field specific lexicons to conduct an appropriate sentiment analysis on social media. However, later comparable studies still tend to use other general lexicon-based methods; NRC-lexicon (Broadstock & Zhang, 2019), SentiMo and Vader Sentiment Analyzer (Wang & Lin, 2018), Loughran & MacDonalds lexicon (Affuso & Lahtinen, 2019). Besides, other studies hold on to machine learning algorithms when measuring sentiment on social media (Batra & Daudpota, 2018; McGurk, Nowak, & Hall, 2019; Tan & Tas, 2021). Overall, the current debate between machine-learning algorithms and different lexicon-based methods remains discussable. Chapter 4 provides a more detailed review of the used methods to analyse sentiment.

Hypotheses

Reviewing the existing literature implies that sentiment can be assumed as a significant factor influencing stock returns. Especially sentiment measured on social media platforms seems to influence daily stock returns (Smailovic et al., 2014; Sprenger et al., 2014; Sul et al., 2016). Additionally, the influence of social media sentiment on daily stock returns is still not widely accepted in economics. To strengthen this body of research the following hypothesis is formed:

Hypothesis 1: There is a positive relationship between social media sentiment and daily stock returns

One of the reasons that the influence of social media sentiment on daily stock returns are not widely accepted, relates to the implications when measuring sentiment. As previously described in section *Sentiment Analysis,* there is no consensus in the current literature with regards to an appropriate method when analysing sentiment. Although Renault (2017) and Oliveira et al. (2016) constructed field-specific lexicons, hereafter other studies still tend to use more general lexicons or machine learning methods. Hence, the second hypothesis is formulated as follows:

Hypothesis 2: Field-specific lexicons will provide a better performance when measuring sentiment on social media, compared to general lexicons.

Previous research with regards to the effects of sentiment on stock returns is mainly focused on the directional effects using predictive regressions (Corea, 2016; Renault, 2017), causality analysis (Bollen et al., 2011; Smailovic et al., 2013; Sprenger et al., 2014), Cumulative Abnormal Return models (Sul et al., 2016), or prediction models to measure the prediction accuracy of stock returns based on sentiment (Batra & Daudpota, 2018; Kordonis, Symeonidis, & Arampatzis, 2016; Meesad, 2014; Nguyen, Shiar, & Velcin, 2015). However, none of these used traditional asset pricing models to measure the pricing power of sentiment. Hence, the third hypothesis is formulated:

Hypothesis 3: The pricing power of a Capital Asset Pricing Model (CAPM) improves with the inclusion of sentiment variables on a daily time frame.

Note that Broadstock and Zhang (2019) used a CAPM model to measure the pricing power of sentiment. However, their analysis is substantially different compared to this study. First, they used the CAPM model to measure intraday day returns on a 1-minute, 5-minute and 30-minute interval whereas this paper focuses on daily returns. This seems more reasonable as the original CAPM equation is constructed to compute daily expected returns. Furthermore, the study of Broadstock and Zhang (2019) focused on different emotional states while this study will focus on general sentiment levels.

Lastly, the conflicting theories and results with regards to lagged sentiment effects gave rise to the fourth hypothesis. Smailovic et al. (2013) and Sprenger et al. (2014) found that lagged sentiment on daily basis shows a positive relationship with today's stock returns. This effect indicates that the incorporation of sentiment in stock returns takes some days. On the other hand, Broadstock and Zhang (2019), however on an intraday timeframe, and Mo et al. (2016) found indications that lagged sentiment is negatively related to today's stock returns. Therefore, the fourth hypothesis is formulated as follows:

Hypothesis 4: There is a positive association between lagged sentiment effects and stock returns on daily basis.

3. Data

This chapter will provide an explanation of the data used in this study. Section 3.1 describes the data collection procedures, the sample and its criteria with regards to the Twitter-data. The Twitter-data is gathered using Kaggle and the Twitter-developers application. Furthermore, section 3.2 will describe the data collection procedures, the sample and its criteria with regards to the stock financial data. The described data-sample is used for both the sentiment analysis outlined in chapter 4 as the CAPM-analysis defined in chapter 5. Data analytic procedures are conducting using R.

3.1 Twitter data

To answer the research question, a quantitative research method will be adopted. This research method contains a sentiment analysis using stock sentiment data retrieved from Twitter. This data source is chosen instead of other social media platforms (StockTwits, Reddit, Tumblr), as it is widely accepted and used in the investor community. Twitter is defined as a social media platform which allows users to post short messages with a maximum of 140 characters (Sul et al., 2016). Those messages, called 'Tweets', are used in this research to perform sentiment analysis. All tweets are accessible via the Twitter API Data stream services, available for researchers and developers. The API allow developers to retrieve 10 million tweets each month. However, retrieving historical tweets, which go further back in time then seven days, is very limited by the Twitter-API in terms of requests and number of tweets. To ensure the number of observations is sufficient to conduct a meaningful regression analysis, a tweet-sample retrieved from Kaggle[3] is used for the main analyses. The data-sample retrieved using the Twitter API services is used to perform robustness checks.

3.1.1 Data sample retrieved from Kaggle

The Kaggle data sample contains a set of more than 3 million unique tweets related to five different companies: Apple, Amazon, Google, Microsoft, and Tesla. Thereby, 5 of the 6th biggest US companies in terms of market cap are involved in this sample. The dataset is used in a paper published in the 2020 IEEE International Conference on Big Data under the 6th Special Session on Data Mining.

The company specific tweets are fetched using a parsing script based on Selenium in Python. [4] The script enables the researcher to retrieve company-specific tweets, because the tweets are matched with a stock ticker while these cashtags are used as search query. A cashtag is defined as a convenient way of tagging stock-related tweets using a dollar-sign ($) in front of the related ticker. An induvial investor tweeting about Apple will use the tag '$AAPL' in his tweet. This sample only focuses on cashtags, because it is assumed that online investors use this tag to initiate that their message contains relevant investment information about the specific company or index (Bartov, Faurel, & Mohanram

[3] Kaggle is a platform developed by and for Data-scientists to attract, train, nurture and challenge each other to solve data science, machine learning and predictive analytical problems. Many data-scientist share their data-samples via this platform: https://www.kaggle.com/getting-started/44916

[4] Detailed information about the used script is provided following this link: https://github.com/omer-metin/TweetCollector

2018; Broadstock & Zhang, 2019, Sul et al., 2016). The initial sample contains the following information per tweet: post date, the text body of the tweet, tweet-id, the username of the writer and the number of retweets, comments and likes of a tweet.[5]

The initial sample includes more than three million tweets over a time span of 5 years, however in this research the time-interval is reduced to a period from 01-01-2018 till 31-12-2019, leading to a horizon of exactly two years. The rationale for this timespan is to be in the upper regions of timespans used in prior research. Smailovic et al. (2013) uses a timeframe of nine months, Bollen et al. (2011) uses a timespan of eleven months, Sul et al. (2016) use a sample period of two years and Meesad (2014) a sample period of one year. Secondly, this time-interval is the most recent period from the sample leading to better generalizability to the present. Finally, a reduction of the initial sample is efficient for a better performance of the R data-analytics to conduct a sentiment analysis.

The abovementioned criteria result in a final sample of 1.6 million unique tweets distributed over the five specific companies. To ensure the reliability and usability of the sample in relation to the research purposes, the sample is inspected on multiple criteria. First, the uniqueness of the tweets is determined based on the removal of duplicates. Although, all tweets are labeled with a unique tweet-id, there are still duplicates within the initial sample. Those duplicates are removed on daily basis to correct for noise by advertisement accounts and users who share the exact same information multiple times. After removing duplicates, the samples per company reduce to the tweet counts shown in table 5. Secondly, after the removal of duplicates the tweets are checked on relatedness to the company specific ticker. This check is based on counting the number of company specific ticker in the total sample. All company specific samples showed similar results where the company-ticker was counted the most. Besides, the word-count approximately matched the number of tweets implying that each tweet specifically relates to the company. The descriptive statistics of the tweet-sample obtained from Kaggle are reported in table 5 in Appendix A.

3.1.2 Data sample retrieved from Twitter using API key

For robustness checks a second data sample of tweets is gathered via the Twitter Search API Key. The Search API allow Twitter-user with a specific developers account to request tweets based on a specific search query. The interest of this study is to measure sentiment related to a specific company or index. Therefore, the predefined query to extract tweets from Twitter was a company or index specific ticker in cashtag format ($AAPL). The R twitter-package facilitates the Twitter API Data stream to fetch tweets posted by users of the platform. For example, to gather Apple specific tweets the following function in R was used:

AppleTweets <- searchTwitter('$AAPL', n = 150000, lang = 'en')

[5] The data sample is available using the following link: https://www.kaggle.com/omermetinn/tweets-about-the-top-companies-from-2015-to-2020

This function opens access to your Twitter API service based on the application of your API Keys embedded in R. The function searches twitter for all $AAPL mentioned in tweets, the number *n* is set at a maximum number of 150,000 tweets per seven days and language to English[6]. These requests return a list of tweet-texts and their related metadata as username, created time, retweet count, favorite count etc. (Kordonis et al., 2016).

The tweet sample is gathered during the period from 08-04-2021 till 26-06-2021. The tweets are related to the following specific companies or indices: Apple, Tesla, Amazon, SPY (S&P500 Index) and QQQ (Nasdaq). To obtain enough Tweets the GOOGL and MSFT sample are replaced by two index samples. Additionally, this allows to identify relationships between company specific and general market sentiment. The sample is inspected based on the same criteria as the Kaggle-sample to ensure reliability and usability for this research purposes. The reduction in tweets after removing duplicates was significantly higher compared to the Kaggle-sample indicating the process was priorly done. Secondly, counting the company or index specific ticker in the total sample showed a gap between the number of tweets compared to the number of counted tickers indicating an error in fetching the tweets. To control this potential error all gathered tweets where the company or index specific ticker was missed are removed from the sample. The descriptive statistic with regards to the second data sample are disclosed in Appendix A Table 6.

3.2 Stock financial data

The stock financial data is extracted in two different ways based on the origin of the tweet data samples. The stock prices regarding the Kaggle data sample of five US companies are retrieved using the YahooFinance application in R. The data was extracted in the period from 1-1-2018 till 31-12-2019. The initial data sample contains the following attributes:

- **Date: The date of the stock market.**
- Open: The stock opening price during the trading date.
- High: The stock highest price during the trading date.
- Low: The stock lowest price during the trading date.
- **Close: The stock closing price during the trading date.**
- Adj. Close: The adjusted stock closing price during the trading date.
- **Volume: The trading volume of stock during the trading date.**

[6] The Twitter API services do not provide tweets further back than seven days in time. Inspecting the search requests showed, none of the company/index related cashtags reached the maximum of 150,000 tweets in seven days.

The bold attributes are retained in the stock market sample, because they are used to compute variables. Non trading days are automatically excluded by the YahooFinance application in R leading to 502 observations in terms of stock market data.

4. Sentiment analysis

Chapter four of this study will outline the sentiment analysis. A sentiment analysis is necessary to quantify the unstructured textual data present in the Twitter-data samples. Section 4.1 will briefly elaborate on the pros and cons whether to use a machine learning or lexicon-based method to conduct an appropriate sentiment analysis. Section 4.2 provides an overview of the used lexicon methods. Section 4.3 specifies the used methodology in this analysis containing textual preprocessing, the measured variables and the model specifications. Lastly, the results of this sentiment analysis are presented in section 4.4. All analysis and procedures are conducted using the data-analytics software R.

4.1 Comparing machine learning and lexicon-based approaches

The general purpose of a sentiment analysis is to convert unstructured qualitative information obtained from Twitter-messages, into organized quantitative variables usable for further analysis. In general: '*Sentiment analysis or opinion mining describes various computational techniques focused to discover ,extract and distil the human emotions, feelings, or opinions from textual information within the web content towards the certain entities'* (Bukovina, 2016; Fang & Zhan, 2015; Godsay, 2015). To conduct an appropriate sentiment analysis regarding this research there are two common options: (i) lexicon-based methods and (ii) machine learning methods (Meesad, 2014; Renault, 2017; Smailovic et al., 2013). As previously mentioned in the literature review, there is a debate going on between both methods in terms of performance, replicability and additional reliability. This section will therefore elaborate on the pros and cons of both methods.

<u>Machine learning method</u>

The purpose of a machine learning algorithm in general is to predict variable Y (dependent) based on a specified dataset of features X. In terms of sentiment analysis, the dependent variable Y occurs in two occasions: Y1 = positive sentiment and Y2 = negative sentiment. All the features of X are determined as a vector of words (Meesad, 2014; Renault, 2017; Sprenger et al., 2014). Before the machine can classify sentiment, it needs a dataset of manually labeled positive and negative classified documents[7]. This dataset is defined as a training dataset with pre-classified documents. The pre-classification process is mostly done manually by the researchers who classifies documents as positive or negative. After this training set is obtained, it is fit in a classifier algorithm. One of the most used algorithms to classify a dataset is the Support Vector Machine. Based on a labeled set of training data, the SVM algorithm will build a model which presents the data-examples as points in space separated by a hyperplane. The hyperplane maximizes the margin between two classes: in this case positive or negative sentiment.

[7] In this research document are defined as single tweet-messages. Those documents consist of many n-gram.

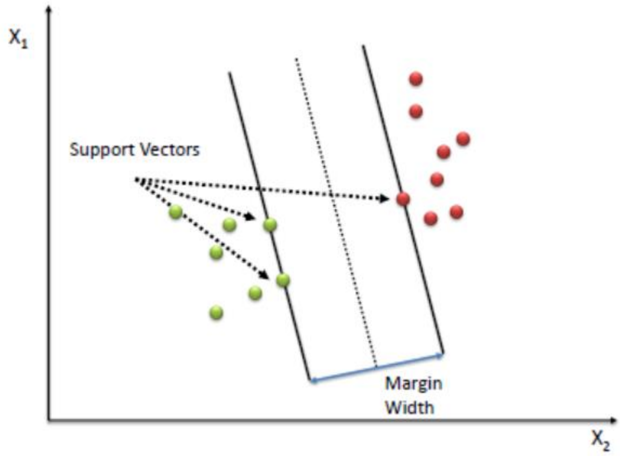

Figure 1 - Visualization of the SVM.

Figure 1 shows a visualization of the SVM concept. The defined hyperplane is presented in the middle-dotted line with a margin width around it. To define this optimal hyperplane (dotted line) the algorithm will be trained to maximize the width of the margin. In figure 1, the green (positive) and red (negative) dots represent supporting word-vectors. These are sentiment-words manually labeled with a positive or negative sign. The dots are touching the 'margin width' on both sides which shows the margin is maximized. When the SVM algorithm is trained it can separate the data linearly. The ideal situation is to obtain a hyperplane which completely divides the vectors (sentiment-words) into a non-overlapping classification. However, the model is not always providing a perfect separation when the dataset is large. In this case, the SVM is trained to obtain a hyperplane which minimizes misclassifications and maximizes the margin width.

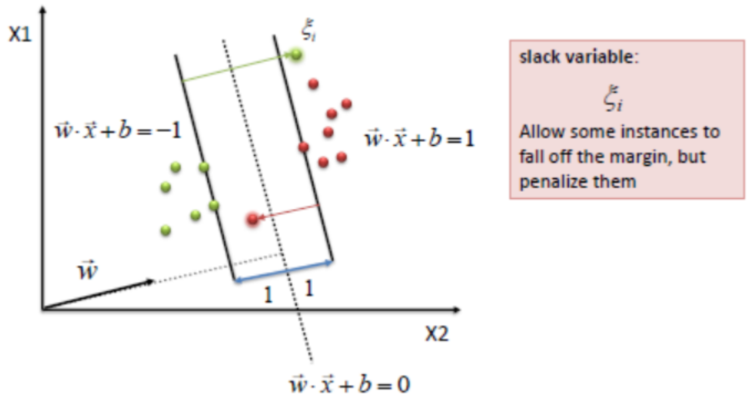

Figure 2 - Visualization of misclassification and the 'slack variable'

Figure 2 visualizes a misclassification in the SVM. The red-dot falls within the margin, however the 'slack variable' or misclassification variable penalizes the misclassifications. The SVM tries to maximize the margin with a slack variable at zero. [8]

Machine learning algorithms show some advantages compared to lexicon-based approaches. First, a machine learning algorithm can handle large sets of data appropriately. Twitter data sets can consist of million textual documents. A machine learning approach will automatically extract a large set of features it is memory efficient; it can handle large feature spaces and fairly robust to overfitting (Batra & Daudpota, 2017; Joachims, 1998; Sebastini, 2002; Smailovic et al., 2013).

However, these advantages are offset by the following limitations of a machine learning based approach. A machine learning algorithm needs a large manually classified dataset of documents to train the algorithm. Since the documents are manually labeled by humans, in prior studies mostly by the authors or external experts, the objectivity of the training set is questionable (Meesad, 2014; Renault, 2017). Due to this subjectivity, the machine learning algorithm could be biased. Furthermore, the training dataset needs to be large enough as the accuracy of the machine learning algorithm relies on the construction of this dataset. For example, Antweiler and Frank (2004) used a training dataset of 1,000 manually labeled messaged when fitting the algorithm. Such a low number of messages to train an algorithm raises concerns about the reliability (Renault, 2017). On the other hand, the process of manually label a bigger dataset is time inefficient (Meesad, 2014). The main issue with machine learning algorithms is their replicability across other studies, because the machine learning algorithms are built using a specific training set in combination with specific metrics and parameters, to fit the dataset used in a study.

Lexicon based method

The lexicon-based method is determined as sentiment analysis procedures focusing on analyzing 'sentiment words'. These words can be determined by obtaining a manually constructed 'bag of words' (Schumaker & Chen, 2009), or using a predefined dictionary with words assigned to a positive or negative sentiment type. In the simplest form, sentiment variables will be constructed by counting the number of negative and positive words in a document, based on a predetermined lexicon. For example, in a two-word investors specific lexicon the word 'buy' is classified as positive, while another word 'sell' is classified as negative. Using this two-word lexicon to assign a sentiment score to the following random tweet: 'I will buy 100 stocks of $AAPL tonight', results in a sentiment score of +1.

To start with a lexicon-based sentiment analysis, a dictionary including words of 'tone' and sentiment (positive/negative) is required. Lexicons are created in three different ways. First, pure experts can

[8] Source for this explanation of a SVM machine learning approach: https://www.saedsayad.com/support_vector_machine.htm

create a list of positive and negative words based on their knowledge. Secondly, a list of non-classified[9] documents can be obtained. The single words can be extracted from the document and fit in a vector of words to generate a list of words based on the non-classified words. Hereafter, all the retrieved words will be manually classified by an expert or research author. Both methods are sensitive to subjectivity as all words are manually labeled as positive, neutral, or negative, by one an expert or research author. The final method to create a lexicon is extracting a list of pre-classified documents. For example, the platform StockTwits allows investors to sign their messages as ‘bullish’ or ‘bearish’ or in other words ‘positive’ or ‘negative’. Extracting those pre-classified messages by investors serves as a dataset in which all documents are divided in a positive or negative class. For each word, the frequency in each class is computed. Based on a process of term-frequency weighting each word is classified as positive or negative (Oliveira et al., 2016; Renault, 2017).

The originally limitations with regards to lexicon-based methods are based on the lack of field-specific lexicons. For example, using the Harvard-IV-lexicon, constructed in the field of psychology, would be inaccurate when classifying financial documents (Loughran & MacDonalds, 2011). Secondly, many lexicons contain equally weighted words per class supposing that each word has the same explanatory power leading to potentially biased outcomes (Jegadeesh & Wu, 2013). For example, the commonly used NRC-lexicon weights a positive word with +1 and negative word with -1 leading to a weighting scheme where words in both classes obtains the same power (Mohammad, Kiritchenk, & Zhu, 2013). Besides, a lexicon-based approach has two main advantages compared to machine learning algorithms. First, a lexicon-based approach is easier to implement and time-efficient, because there is no phase of training an algorithm. Secondly, all lexicons are publicly available which ensures replicability, transparency and comparison of results with other studies (Loughran & McDonalds, 2016; Mukthar, Khan, & Chiragh, 2018; Oliveira et al., 2016; Renault, 2017).

Reviewing the comparison between the pros and cons of both methods, this study will use a lexicon-based approach to conduct sentiment analysis. An appropriate reason for this decision is that Renault (2017) and Oliveira et al. (2016) provided a solution with regards to the described limitations, since they constructed a field-specific lexicon with varying scaled weights per word. Additionally, the pros of easier implementation, time-efficiency, transparency and replicability with regards to a lexicon-based approach outweigh the pros of machine learning algorithms.

4.2 Overview lexicons

This paragraph will provide an understanding of the lexicon-based methods used in this study to analyze sentiment in Twitter-messages. In this study five different lexicon-based methods are used and compared to make a statement with regards to *Hypothesis 2.* Section 4.1 already explained why a

[9] Classified in this research means that a textual document is classified in terms of sentiment, thus positive, neutral or negative.

lexicon-based method is chosen instead of a machine-learning approach. However, as briefly described in the literature review there are many different lexicons. This section provides an understanding of the five chosen lexicons which are compared during this sentiment analysis.

Oliveira et al. (2016) lexicon

Oliveira et al., (2016) created an investor-based lexicon focusing on microblogging messages in which the sentiment tone is explicitly disclosed by the writer of the message, because Stocktwits allows investors to classify their message as 'bullish' or 'bearish'. Secondly, Twitter messages containing cashtags of stocks traded in the US stock markets were used for reasons of data availability. The obtained dataset was tested by two statistical measures and three other adaptations to calculate a weighted sentiment score for each word. Oliveira et al. (2016) provided evidence that the investor related lexicon significantly increased the accuracy of a sentiment analysis compared to other 'normal' lexicons. A detailed description on the procedures to develop this investor-based lexicon is provided in the study of Oliveira et. al. (2016). Renault (2017) used the same procedures to construct a field-specific investor lexicon. Since this approach is similar and improved, this study will provide a more detailed description in the following section.

Renault (2017) lexicon

Renault (2017) embroidered the findings of Oliveira and developed an investor-based lexicon as well, using 375.000 bullish and bearish messages as dataset to classify sentiment. Renault (2017) chooses for a conservative approach in terms of natural language processing (stemming, lemmatization, removing numbers etc.), because social media users use many variations of words to express themselves in various feelings. Where common dictionary lookups remove social media specific content as cashtags ($AAPL), numbers, links and mentions of users, Renault incorporates this value-weighted content in his lexicon. Numbers, cashtags, links and mentioned users are replaced by the respectively common words "numbertag", "cashtag", "linktag" and "usertag". Furthermore, Renault (2017) considers bigrams (two words) instead of just unigrams (one word), because they add valuable information improving the accuracy of a lexicon. Namely, the word "buy" is often used in the bigram "strong buy" adding extra sentiment value to the word "buy".

To build the investor-lexicon, all unigrams and bigrams which occurred at least 75 times in the sample of 750,000 messages were extracted using a bag of words approach. This selection leads to 19,665 specific n-grams used to calculate the sentiment weight of each n-gram. For the selected n-grams, the number of occurrences in the 375,000 bullish messages and 375,000 bearish messages were counted. This enables Renault (2017) to define the sentiment weight of each n-gram based on the following equation:

$$SW(w) = \frac{N_{mpos,w} + N_{mneg,w}}{N_{total,w}}$$

Where

$SW(w)$ = sentiment weight of each word n-gram *w*
$N_{mpos,w}$ = number of occurrences of n-gram *w* in bullish (pos) classified messages
$N_{mneg,w}$ = number of occurrences of n-gram *w* in bearish (neg) classified messages
N_{total} = sum of occurrences of n-gram *w* in bullish and bearish classified messages

An illustration of this formula: the n-gram "bear" occurred 5,669 times in the total message sample of 750,000 messages. This total amount is counted 4,163 times in the bearish classified messages and 1,506 in bullish classified messages, leading to a sentiment weight of (1,506-4,163)/5,669 = -0.4687 in the sentiment lexicon. Thereby, the weighted-scores are scaled between [-1, 1]. All n-grams are sorted by these sentiment weights to define the weighted field specific lexicon by selecting 4,000 positive terms (last quintile) and 4,000 negative terms (first quintile). Reducing the lexicon to the first and last quintile of the sentiment weights excludes all neutral sentiment scores between -0.20 and 0.20 (Renault, 2017).

<u>SentiWordNet</u>

SentiWordNet is the first general lexicon approach used in this study to measure sentiment on social media content. Prior research commonly used this lexicon before Oliveira et al. (2016) and Renault (2017) constructed a field specific lexicon (Meesad, 2014; Risius et al. 2015).

The study of Goncalves, Aurajo, Benevenuto, and Cha (2013) compared the coverage, accuracy, and prediction possibilities of six different sentiment analysis tools. They provided evidence that SentiWordNet outperformed the other tools in terms of coverage, while accuracy and prediction performance was on average. Furthermore, the study showed a potential bias that some of the tools overestimate the positive scores when measuring polarity. SentiWordNet showed better performance in terms of avoiding the potential bias compared to other commonly used tools as SentiStrength and SenticNet. Therefore, the SentiWordNet tool is used to analyze the samples of tweets in this study.

<u>SentimentR</u>

SentimentR is a dictionary-based sentiment analysis tool developed by Rinker (2020). This tool is available as a package in R. The main advantage of SentimentR compared to other dictionary lookups packages in R as Afinn, Bing & NRC, is that SentimentR considers valence shifters (amplifiers, negators, adversative conjunctions and de-amplifiers). Therefore, this tool is called an augmented dictionary lookup tool. For example, the sentence "I do not like $AAPL" contains a negator "not" which flips the sign of the polarized word "like". The SentimentR tool measures this sentence correctly because it considers the valence shifter, while other Lexicons won't. A frequency analysis provided evidence that in 20% of the observed text files, a negator appears in combination with a polarized word. Thus, not accounting for valence shifter would significantly affect the accuracy of sentiment analysis.

To obtain a sentiment score for each Tweet, the sentiment at the sentence level is measured using the *sentiment* function in R. The function breaks each message into a bag of words and assign a sentiment-value to each word based on a sentiment dictionary (Jockers & Thalken, 2017). The assigned words will be weighted based on the number of words in a sentence. Furthermore, the valence shifters will be measured affecting the weighted values of words in the sentences leading to a final sentiment score of each sentence.

NRC Sentiment

Since sentiment in general is driven by the emotions of online investors, this research will also consider an emotional based sentiment lexicon. Following Broadstock and Zhang (2019) the NRC Word-Emotion Association list of words is used to extract emotion types of certain tweets. The NRC-lexicon is developed by the Canadian Research Council and contains words associated with eight different types of emotions k: anticipation, anger, disgust, fear, joy, surprise, and trust. The NRC-Lexicon is used to obtain association scores (0 for not associated and 1 for associated) for each tweet per emotion type. For example, a tweet could have a score of 2 for joy, 1 for anticipation and zero for all the other emotion types. The emotion scores are useful in the sentiment analysis to get insights in different emotion types affecting sentiment or the stock price predictability in general.

4.3 Methodology sentiment analysis

4.3.1 Textual preprocessing of Twitter-messages

After both tweet samples are gathered as previously defined in chapter 3, the unnecessary metadata is excluded leading to a sample of all tweet texts matched with posting date. Before the sentiment variables can be constructed using the lexicon methods, the Twitter-messages require sufficient textual preprocessing (Batra & Daudpota, 2018; Broadstock & Zhang, 2019; Renault 2017).

Tweets are messages posted in a causal format written by the users of the platform. Those users have created their own language and not every user has the same pattern to post a tweet (Batra & Daudpota, 2018; Meesad, 2014; Smailovic et al., 2013). The tweets posted by users contains text written in informal manner with slangs, abbreviations, emoticons, URLs, punctuation, numbers, special symbols, Hashtags and Cashtags (Batra & Daudpota, 2018; Ramachandran & Parvathi, 2019). Tweets in an informal manner without any preparing or preprocessing are defined as raw tweets. For example, a raw tweet text from the Apple-sample:

$AAPL still trading at the 50 sma $172 is clearing the 4-day (weekly) cluster after the gap down. Nest target $175-$177-$179.

As previously defined in section 4.2 he five lexicons are all unique in terms of classifying sentiment in textual data. Therefore, the textual pre-processing procedures are different for the general lexicons[10] compared to the investor's specific lexicons of Oliveira et al. (2016) and Renault (2017). The following pre-processing procedures are performed and appropriate with regards to the SentimentR, NRC and SentiWordNet approaches.

1) *Text – Data cleaning*

   The first step in pre-processing the raw tweets is to focus purely on the body of the text. Following Batra & Daudpota (2018), Kordonis et al. (2016), Meesad (2014) all meaningless characters and symbols are removed from the texts. For example, tweet texts contain many symbols which are commonly used in the twitter-language to mention (@) other users, to add a tag to your tweet (#) or assign your tweet as financial related ($). Those symbols, URLs, links, RT-signs, digits, and punctuations are removed from the texts. These symbols and characters are classified as noise in sentiment analysis (Kordonis et al., 2016). Following Wong, Rovalino, and Akyildirim (2019), all non-ASC11 characters[11] are removed as well. The final step in the text data cleaning process is to delete empty text columns and add a document-id to each column considering each text row as a specific document.

2) *Tokenization*

   After the text is cleaned from meaningless characters and symbols, the process of tokenization can be performed. Tokenization is the process of breaking down the texts to lists of words per document. The smaller units of texts are called tokens, which can be either characters, words, or sub-words. The words counted per document (text of a tweet) are stored in a Vector called Corpus. A Corpus is a collection of all the documents containing tokenized texts based on word tokenization. The tokens stored in Corpus are used to create a vocabulary to perform further analysis. During the process of tokenization English stopwords are removed. The R – Stopwords dictionary contains a list of stopwords that are meaningless and inappropriate in sentiment analysis.[12] For example, a Corpus created for the Apple sample counts the tokens per document:

[10] Hereafter, the general lexicons refer to the SentimentR, NRC and SentiWordNet lexicons.
[11] ASCII stands for American Standard Code for Information interchange based on the English alphabet consisting of all possibilities on a standard English keyboard.
https://www.dynadot.com/community/help/question/what-is-ascii
[12] Information regarding the process of tokenization is retrieved from:
https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/

	buy	sell	stock	earnings	Apple
Doc1	1	0	0	0	1
Doc2	1	0	0	0	0
Doc3	0	1	1	1	1

Figure 3 illustrates the process of tokenization.

3) *Prepare the Corpus*

The final step in textual preprocessing is to remove white spaces from the Corpus, stem words and change all letters to lower case. Word-stemming reduces words to their root form, for example the stem of "calculation", "calculators" is "calcut". Reducing words to its original stem creates unification among the documents. Furthermore, extra white spaces in the document will be removed and all words are changed to lower case. Finally, all prior cleaning procedures are performed again in the Corpus to be sure all preprocessing requirements are met.

As previously defined in section 4.2, the lexicons of Renault (2017) and Oliveira et al. (2016) are constructed using pre-classified documents retrieved from the investor-specific microblogging platform StockTwits. This implies both lexicons consist of n-grams which are frequently used by investors who use social media to spread their thoughts. Additionally, the general lexicons do not consist of these investors specific n-grams, because they are not specifically developed to classify investor sentiment in social media texts. To illustrate these differences figure 5 shows the sentiment scores assigned to the n-gram 'bull' for each lexicon. In investors language the word 'bull' is defined as an investor who thinks a stock will rise in price, while a 'bull' in general language is defined as an animal. Consequently, the general lexicons misclassify the n-gram 'bull' because they assign a negative sentiment score while it is assumed as a positive sign among investors. This misclassification could result in biased results when measuring investor sentiment. Subsequently, the lexicons of Oliveria et al. (2016) and Renault (2017) also incorporate many variations on the n-gram 'bull' as 'bullish, bull flag, bull trap, bullish sign, bullish engulfing'.

Lexicon	N-gram	Sentiment-score
Renault	Bull	0.587
Oliveira	Bull	0.383
R	Bull	-1
NRC	NA	NA
SWN	Bull	-0.375

Figure 4 illustrates the assigned sentiment-scores to the n-gram "bull" for each Lexicon.

Furthermore, the removal of meaningless characters as pre-processing done by the general lexicons is not supported by Oliveira et al. (2016) and Renault (2017), because these characters contain valuable sentiment information. Therefore, during the textual preprocessing for the investor-based lexicons, these characters are replaced by generalized n-grams implemented in the lexicons. For example, the company specific $TSLA is replaced by 'cashtag' (Renault) or 'tkr' (Oliveira), a number by 'numbertag' (Renault) or 'num' (Oliveira) and so on. The other procedures as tokenization and creating a corpus are also not executed for these lexicons. Table 7 in Appendix A shows a raw tweet after pre-processing for each lexicon to illustrate the differences.

4.3.2 Sentiment variables

After the tweet-texts are pre-processed a sentiment-score is computed based on the sentiment weighs assigned to each n-gram. The sentiment score per tweet is computed by matching all n-grams per tweet with the n-grams documented in a certain lexicon. The sum of all weighted n-grams constitutes the aggregated sentiment score per tweet. In formula:

$$SS_{sit} = \sum(N_{ls} * W_l) \quad (1)$$

Where

SS_{sit} = aggregated sentiment-score per tweet *s* for company *i* on day *t*.

N_{ls} = number of n-grams in lexicon *l* matched per tweet *s*

W_l = weights assigned to each match n-gram in lexicon *l*

This approach to measure weighted sentiment per document is called a simple relative word count term-frequency approach. This approach is commonly used in prior research when computing weighted sentiment based on Twitter-messages. (Renault, 2017; Smailovic et al., 2013).

The aggregated sentiment scores per tweet are aggregated per trading day to obtain an overall sentiment-variable per trading day. Following prior research, the aggregated sentiment-scores are divided by the number of messages posted on the same trading day (Renault, 2017; Smailovic et al., 2013; Mo et al., 2016, Sul et al., 2016). Accordingly, the sentiment variables are defined as the aggregated sentiment scores on daily basis divided by the number of messages posted on the same day using the following formulas:

$$SREN\_mv_{ti} = \sum SS_{sit} / \sum MV_{ti} \quad (2)$$

$$SOL\_mv_{ti} = \sum SS_{sit} / \sum MV_{ti} \quad (3)$$

$$SR\_mv_{ti} = \sum SS_{sit} / \sum MV_{ti} \quad (4)$$

$$SNRC\_mv_{ti} = \sum SS_{sit} / \sum MV_{ti} \quad (5)$$

$$SSWN\_mv_{ti} = \sum SS_{sit} / \sum MV_{ti} \quad (6)$$

Where

$SREN\_mv_{ti}$ = sentiment variable SREN for company *i* on day *t*

SS_{sit} = aggregated sentiment-score per tweet *s* for company *i* on day *t*.

MV_{ti} = aggregated message volume for company *i* on day *t*.

The name of each variable corresponds to the used lexicon measuring sentiment divided by the number of messages on a certain day: Renault (*SREN_mv),* Oliveira (*SOL_mv),* SentimentR(*SR_mv),* NRC *(SNRC_mv)* and SentiWordNet (*SSWN_mv).*

As previously mentioned, the fundamentals of each lexicon are different, leading to various scales in terms of sentiment scores. Table 8 in Appendix A provides an overview of the differences between each lexicon in terms of n-grams, purpose, scores-scale, mean-score, development year and source. During the tweets pre-processing phase the initial sample is reduced by removing duplicates as previously defined. Additionally, the number of tweets is further decreased through the process of computing sentiment-scores per tweet. If none of the n-grams in a tweet document is assigned to the n-grams in a lexicon, the tweet obtains a zero-score leading to removal of the sample. Table 9 in Appendix A shows the number of tweets retained after the process of assigning sentiment scores to each tweet. The number of tweets retained per sentiment variable could be related to the number of n-grams existing in the lexicon.

4.3.3 Dependent variable

To examine the explanatory power of each lexicon in relation to daily stock returns, the following dependent variable is created. The daily return of a certain stock (i) on day(t) is computed following based on the daily close price (Smailovic et al., 2013; Sul et al., 2016):

$$R_{it} = 100 * (C_{it} - C_{it-1})/C_{it-1} \quad (7)$$

Where

R_{it} is defined as the daily stock return for firm *i* on day *t*. C_{it} is defined as the close price of company I on day *t*. Finally, the C_{it-1} is defined as the close price of company *i* on day t-1.
The daily return using the close price of today and yesterday is used to explore the predictive relationship where daily sentiment could predict daily stock returns.

4.3.4 Control variables

In prior research the incorporation of control variables in the models is questionable because company specific sentiment captures many other variables. Accordingly, the use of control variables is dependent on the models used to measure the effect of sentiment on stock returns. Sul et al. (2016) using an event-study focusing on cumulative abnormal returns caused by sentiment effects. They control for earnings surprise, past returns, abnormal returns on the prior trading and the effects of upgrades/downgrades by financial analysts. Following Smailovic et al. (2013) this study assumes that earnings surprises, upgrades, and downgrades by financial analysts are incorporated in the sentiment-variables itself. Therefore, these controls are not used in this study.

In addition, the effect of message volume will be captured in a message volume control variable. The message volume on social media platforms is determined as the number of messages posted on a certain platform (Twitter) about an index or specific stock (Alostad & Davulcu, 2016; Antweiler &

Frank, 2004; Sprenger et al., 2014). Dewally (2003) presented results that most investment related messages represent buy signals leading to the assumptions that an increase in message volume associates with an increase of bullish sentiment. Additionally, Sabherwal, Sarkar, and Zhang (2008) and Wysocki (1998) find evidence that high message volume on day (t) leads to significantly positive returns on the next day. The message volume control variable *(MV)* is measured by aggregating all messages related to company *i* on day *t.*

Trading Volume *(TV)* is used in this research due to its relationship with stock returns, sentiment and message volume. In general, trading volume captures liquidity in the market. Baker and Stein (2004) explained that in practice short selling is costlier than buying stocks and closing those positions to take profit. Furthermore, they assume that irrational investors are more likely to act when they are in positive mood instead of negative mood. Thus, the effect of positive sentiment on trading volume outweighs the effect of negative sentiment. They found that high trading volume is caused by sentiment driven noise traders who drive stock prices over their fundamental value. Additionally, trading volume is also determined as predictor for stock returns. Due to these findings, Baker and Wurgler (2007), and Gao and Liu (2020) used trading volume as proxy to identify sentiment. On the other hand, Pathirawasam (2011) identified a negative relationship between trading volume and stock returns. The reason for this negative relationship was found in investor's misspecification of news events. Other studies found a feedback relation where trading volume causes stock returns, and vice versa (Khan & Rizwan, 2008). Alostad and Avalcu (2016) found evidence that outbreaking Tweet volumes causes significant boosts in trading volume and stock returns. To control for abovementioned effects the aggregated daily trading volume for company *i* on day *t* is used as control variable.

Furthermore, to control for the market-sentiment effect on the individual stocks a market sentiment variable is included as control variable. As a sample of tweets containing information about the general market sentiment is not available, the AAII bull-bear market indicator[13] will be used to measure market sentiment among investors. The American Association of Individual Investors Index (AAII) is used as a proxy for general market sentiment in prior research (Sayim, Morris, & Rahman, 2013). The AAII index has been conducted around a weekly survey where investors express themselves about their thoughts about which direction the market will follow. AAII processes these surveys and label them as bullish, neutral, or bearish. The results form an index of the bullish, neutral, and bearish spread among investors. In this research the market sentiment variable *(MS)* is computed by dividing the percentage of bulls by the percentage of bears in the index. A market sentiment-score of 1 implies an equal spread between bulls and bears in the market. The AAII index gives only weekly results, thus each trading day in the same week is labeled with the same market sentiment variable.

[13] The AAII index is downloaded from: https://www.aaii.com/sentimentsurvey. The sentiment votes are weekly obtained by a survey among investors.

4.3.5 Regression models

The relationship between stock returns and sentiment is examined using simple ordinary least squares regression (OLS). The constructed sentiment variables are inserted as independent predictors of the dependent variable stock returns (R_{ij}). Regressing each sentiment variable separately on the daily stock return will examine which sentiment variable fits the best as predictor for stock returns.

Before the OLS analysis will be constructed, the sample and variables need to satisfy the fundamental OLS assumptions. First, the data needs to be stationary meaning the data do not suffer from trends and seasonality. To check if the variables are stationary an augmented Dickey Fuller test will be executed. All variables showed p-values < 0.05 meaning the data is assumed to be stationary. Accordingly, the data is tested on autocorrelation with graphical and statistical measures. The autocorrelation graph shows that the sentiment variables experience autocorrelation over time. Therefore, a Durbin Watson test is executed to determine autocorrelation in the data. The p-value of the Durbin Watson tests > 0.05 implying the data does not suffer from autocorrelation.

Subsequently, the residuals are graphically and statistically checked on homoscedasticity. The graphical check is done by plotting the error terms against the fitted values. These plots already suggest that the residuals suffer from heteroscedasticity. To be sure a statistical Breusch-Pagan test is executed. This test provides significant evidence that the data suffers from heteroscedasticity, leading to the situation where the variance of the residuals is not constant over a range of measured values. This could imply biased standard errors (Schwert & Seguin, 1990). To solve the problem of heteroscedasticity, robustness standard errors are used in this analysis following Broadstock and Zhang (2019).

Furthermore, the outliers in this sample are graphically checked by the residuals versus leverage plot to identify influential cases. This plot function in R is part of the autoplot-function which identifies automatically the three most influential points in the dataset. Influential points are outliers which have a disproportional influence on the regression analysis. After recognizing potential influential points or outliers, it is necessary to identify their nature and why they exist in the sample (Dhakal, 2017). Inspecting the nature of the influential points reveals that they often relate to sentiment peaks caused by company specific news events as already showed in prior research (Smailovic et al., 2013). Therefore, it is assumed that these influential points cause valuable information in the sample whereby these points are not caused by bad data, measurements error and non-validity (Dhakal, 2017). However, outliers in terms of daily message volume (*MV*) are considered as potential influential points caused by bad data. A boxplot analysis shows potential outliers on the min and max side of the distribution, however the outliers on the max side are related to company specific news events or extreme price actions (Smailovic et al., 2013; Alostad & Davalcu, 2016). On the other hand, the outliers below the determined minimum value of the distribution determined by the boxplot indicate potential data errors. Therefore, these rows are removed from the sample.

Finally, the variables are checked for being normally distributed based on the Q-Q plots. These plots indicate that most variables are normally distributed because the deviations from the straight line are minimal. However, numerical tests in terms of a skewness and kurtosis-test are executed for each variable. The acceptable ranges for skewness and kurtosis are respectively between -2 and +2 and -7 and +7 (Byrne, 2010; Hair, Black, Babin, & Anderson, 2010). The variables message volume (*MV*) and trading volume (*TV*) are consistently not in the defined acceptable ranges, which is corrected by taking the natural logs of these variables. The log transformation follows the study of Antweiler and Frank (2004) who mention that taking the logarithm of these variables allows to control for scaling.

To test hypothesis 1 and hypothesis 2 an OLS-regression will be performed as previously defined. The basic regression specification is formulated as follows:

$$R_{it} = \beta 0 + \beta 1 SREN\_mv_{it} + \epsilon_{it} \quad (8)$$

Where R_{it} represent the stock return of company *i* on day *t*. β0 is defined as the constant of the regression equation and $\beta 1SREN\_mv_{it}$ is defined as the independent sentiment variable related to company *i* on day *t*. Finally, ϵ_{it} represents the error term in this regression equation. To determine the accuracy and predictive power of each sentiment lexicon, the SREN_mv will be replaced by the other sentiment variables separately. Thus, five single regressions will be executed for all five stock samples resulting in 25 single regressions.

Sul et al. (2016) and Tetlock et al. (2008) state that the performance of a lexicon is measurable using an OLS. If a certain lexicon does not accurately measure sentiment, it is less likely to find a significant relationship between sentiment variables and stock returns. Secondly, these simple regressions will examine if the lexicons suffer from the sentiment bias. The sentiment bias states that lexicons tend to overestimate positive sentiment as they assign more words to positive values compared to negative values. The existence of a sentiment bias leads to a poor performance of the sentiment analysis (Han, Zhang, Zhang, Yang, & Zou, 2018; Goncalves et al., 2013).

The second measure is to evaluate the correlation among the different sentiment lexicons. Renault (2017) assumes that high correlations between significant sentiment variables could confirm the accuracy of their explanatory power. On the other hand, a low correlation between two sentiment lexicons implies many classification errors by one of them. Pearson’s correlation matrix measures the correlations between the five sentiment variables. To identify biases in each sentiment lexicon, the control variables are included in Pearson’s correlation matrix as well. A high correlation between message volume and sentiment variables could indicate the potential sentiment bias, because if the lexicon is biased it will potentially over or underestimate the sentiment-scores on days with high message volumes. Furthermore, it makes sense to identify if the stock-specific sentiment variables are correlated with the general market sentiment variable (*MS*).

To further examine the explanatory power of each sentiment-variable a multiple regression is executed. This multiple regression contains the input of each sentiment-variable augmented with the control variables. The regression specification is formulated the following:

$$R_{it} = \beta 0 + \beta 1SREN\_mv_{it} + \beta 2SOL\_mv_{it} + \beta 3SR\_mv_{it} + \beta 4SNRC\_mv_{it} + \beta 5SSWN\_mv_{it} + \beta 6TV_{it} + \beta 7MV_{it} + \beta 8MS_{it} + \epsilon_{it} \quad (9)$$

Where R_{it} represent the stock return of company *i* on day *t*. β0 is defined as the constant of the regression equation and β1 – β8 are the main independent variables and the control variables to company *i* on day *t*. Finally, ϵ_{it} represents the error term in this regression equation.

The purpose of this multiple regression model, including all sentiment variables augmented with the control variables, is to examine the possible complementary power of combining multiple sentiment variables. It seems reasonable to suggest that some sentiment-variables might be complementary as their units of measurement are completed different. The second reason for executing a multiple regression including all created variables, is to identify which sentiment-variable emerges as most significant predictor in relation to stock returns. In the Pearson's Correlation matrix, it is not observable if a high correlation assumes that both sentiment-variables are accurate or inaccurate. Therefore, the significance levels in the multiple regression offer a solution to obtain a better interpretation of the mutual correlations. Additionally, it seems obvious to expect problems with multicollinearity as all sentiment variables are included in the model. But the purpose of this multiple regression is not on constructing an acceptable model to measure the relationship between sentiment and stock returns. The results of the multiple regression are only used to obtain an understanding of the explanatory of each sentiment-variable and their interrelation. Besides, it provides an explicit overview in terms of descriptive statistics.

4.4 Results sentiment analysis

This section presents the results with regards to the sentiment analysis. Section 4.4.1 elaborates on the simple OLS regression model. Section 4.4.2 outlines the descriptive statistics, correlation matrices and the results of the multiple OLS regression. Finally, section 4.4.3 describes the executed robustness checks.

4.4.1 Simple linear regressions

In this paragraph, the main results with regards to hypothesis 1 and hypothesis 2 are outlined. Table 1 represents the results of the simple OLS-regressions measuring the relationship between the single sentiment variables and the daily stock returns. The performance of the sentiment variables in the single regressions are evaluated based on the significance levels and the adjusted R-squared. Table 1 reports a positive effect and significant effect for each sentiment variable separately for each sample. The coefficients for both SREN_mv and SOL_mv are equally significant at the highest level ($p < 0.001$) for each sample, while the other sentiment variables show significance on various lower levels. The significance is an important measure in this analysis as it states the level in which the observed results are not caused by randomness. The substantial higher significance level for SREN_mv and SOL_mv over the entire sample, provides a first indication that a field specific investor lexicon shows the best measurement performance.

The adjusted R-squared measures the explanatory power of a model implying how well a regression model fits the observed data. In these simple OLS regressions, a higher R-squared indicates that a sentiment-variables contains of more explanatory power when predicting stock returns. On average the adjusted R-squared is the highest for the SREN_mv variable followed by SOL_mv, respectively 13.89% and 10.09%. The other sentiment variables show an adjusted R-squared varying from 2.28% till 5.04% on average. This observation indicates that both investor-related lexicons provide the best fit when measuring sentiment in tweets. In general, the explanatory power of these simple regression models is low, but with stock returns as dependent variable it seems logical as there are countless factors influencing stock returns. Intuitively, an adjusted R-squared of 13.89% on average, implies that sentiment considers 13.89% of these factors. The first hypothesis predicts a positive relationship between sentiment and daily stock returns. Since the coefficients of all sentiment variables are positive and significant at the 5% level, the first hypothesis can be accepted. This indicates that sentiment measured on social media contains predictive value with regards to stock returns. The second hypothesis predicts that field-specific lexicons will provide a better performance when measuring sentiment on social media, compared to general lexicons. Both SREN_mv and SOL_mv report better performance compared to the general lexicons when comparing the adjusted R-squared and significance levels. However, the general lexicons report significant positive coefficients too. Therefore, accepting or rejecting hypothesis two will be done after assessing the results presented in section 4.2.2.

Table 1: Simple OLS regression results on daily returns - MV

Variable	TSLA sample	AAPL sample	AMZN sample	MSFT sample	GOOGL sample	Full sample (mean)
Intercept	0.5488***(3.708)	-0.4071***(-4.975)	0.0798(1.035)	-0.7462***(-5.657)	-0.2133*(-2.060)	
SREN_mv	18.94***(9.054)	8.588***(11.80)	8.579***(10.924)	5.895***(7.671)	1.4418***(3.704)	8.689(8.73)
Adjusted R^2	0.1534	0.2187	0.1933	0.1041	0.0248	0.1389
Variable	TSLA sample	AAPL sample	AMZN sample	MSFT sample	GOOGL sample	Full sample (mean)
Intercept	-3.287***(-3.954)	-3.472***(-9.613)	-3.119***(-9.366)	-1.705***(-6.409)	-0.9069***(-4.299)	
SOL_mv	1.258***(4.157)	1.056***(10.12)	1.151***(9.955)	0.6628***(7.129)	0.3615***(4.876)	0,8979(7.8926)
Adjusted R^2	0.0316	0.1704	0.1657	0.0909	0.0437	0.1005
Variable	TSLA sample	AAPL sample	AMZN sample	MSFT sample	GOOGL sample	Full sample (mean)
Intercept	-1.834***(-4.956)	-0.5720*(-2.543)	-1.335***(-4.818)	-0.8109**(-3.058)	-0.4573*(-2.401)	
SR_mv	64.77***(5.737)	6.845**(3.239)	26.16***(5.443)	13.73***(3.681)	8.904**(2.943)	24,08(4.258)
Adjusted R^2	0.0602	0.0189	0.0548	0.0246	0.0151	0.0347
Variable	TSLA sample	AAPL sample	AMZN sample	MSFT sample	GOOGL sample	Full sample (mean)
Intercept	-2.502***(-4.369)	-0.5242*(-2.154)	-1.533***(-4.285)	-0.6213*(-2.032)	-0.4270(-1.888)	
SNRC_mv	14.27***(4.728)	1.5514**(2.759)	6.478***(4.705)	2.6303*(2.531)	1.892*(2.277)	5,364(3.398)
Adjusted R^2	0.0411	0.0132	0.0410	0.0107	0.0083	0.0228
Variable	TSLA sample	AAPL sample	AMZN sample	MSFT sample	GOOGL sample	Full sample (mean)
Intercept	-2.578***(-4.275)	-2.491***(-7.701)	-1.065***(-3.955)	-0.2391(-1.412)	-0.4828**(-3.18)	
SSWN_mv	18.08***(4.604)	15.16***(8.262)	8.075***(4.568)	2.965*(2.402)	4.670***(4.07)	9,79(4.7804)
Adjusted R^2	0.0389	0.1198	0.0386	0.0095	0.0302	0.0504

Table 1: presents the OLS regressions on R_{it} for each stock specific sample. The simple regressions are executed one by one inserting each of the five sentiment variables separately in each model. The control variables are not included as these simple regressions serve as prove to measure the performance of the selected sentiment-lexicons. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

4.4.2 Multiple linear regression including descriptive statistics and correlations

As previously described in section 4.3.5., a second approach is conducted to identify measurement performance of each lexicon. This section will elaborate on the descriptive statistics of the sentiment variables, the correlations presented in Pearson's Correlation Matrix and the results of the executed multiple regression. This regression includes all the variables as defined previously. Table 10 in Appendix B presents the descriptive statistics of all the dependent, independent and control variable for all the samples in period 2018 – 2019. For each company the average daily returns (R_{it}) are positive, which reveals that the stock prices increased during the sample period. Accordingly, all sentiment variables show an average positive sentiment score as well which could imply that they all suffer from the previously defined sentiment bias. Only the SREN_mv variable presents an average negative mean for the TSLA sample. Furthermore, the SREN_mv variable is the only sentiment variable which consistently measures negative sentiment scores as minimum value leading to an approximately equal distribution of positive and negative sentiment days. For example, the SREN_mv variable reports a minimum score of -0.3231 and a maximum of 0.338 in the AAPL sample. In addition, the other sentiment variables show a skewed deviation in positive and negative sentiment. This is remarkable because it would make sense that each stock-sample includes negative sentiment trading days since all samples show daily returns of -5% or lower. Especially the SOL_mv variable reports an overestimation of positive sentiment as none of the trading days obtained an average negative sentiment score in the TSLA, AMZN and AAPL samples. This is also remarkable because the SOL_mv variable showed approximately the same performance results as SREN_mv in the single regressions. Therefore, the expectation was that this variable should present a distribution in positive and negative sentiment days comparable to SREN_mv. The partition in positive and negative sentiment is for the other variables also skewed to the positive side, whereby in the most cases the minimum value is also positive[14]. The log_MV variable's mean is deviated from 5.046 till 6.90. Furthermore, it is worth mentioning that the market sentiment was 1.257 on average during this period which implies a positive sentiment as 1.257 > 1 in the bullish/bearish ratio as previously defined in section 4.3.4.

Hereafter, table 11 in Appendix B reports the Pearson's correlation matrix for all variables included in the executed multiple regression. The correlation matrix shows significant correlations among the sentiment variables which obviously makes sense in this analysis. The SREN variable is set as benchmark since its explanation power is the highest as determined in the single regressions. As expected, SOL_mv shows the most correlation with SREN_mv varying from 0.57 to 0.77 per sample. The highest correlations between SREN_mv and SOL_mv seems to confirm that quantifying sentiment of messages on social media using field-specific lexicons outperforms the use of general lexicons to measure sentiment. Accordingly, the other sentiment variables report correlations between a range of 0.55 and -0.43 with the SREN_mv variable which are classified as low ($r < 0.4$) or moderate ($0.4 < r > 0.8$) (Shi & Konrad, 2009). Furthermore, it is worth mentioning that the SNRC_mv and SR_mv variables show correlations between 0.8 and 0.98 indicating that their sentiment measures are approximately similar. The other correlations between the sentiment variables and the control variables are mostly below 0.4 and thereby classified as low.

The results of the multiple regression are disclosed in table 12 of Appendix B. The executed multiple regression confirms the results from the single regression as the coefficient of SREN_mv stays

[14] To prevent confusion; this type of distribution does not elaborate on the normal distribution of the variables. The variables are normally distributed as tested in section 4.3.5. However, the distribution is relatively skewed to positive sentiment trading-days.

positive and significant ($p < 0.001$) for all samples except from the GOOGL sample. Secondly the Sol_mv variable shows a negative relation in the TSLA sample and positive for the others at minimum significance of $p < 0.05$ for all samples. As expected, the other sentiment variables report inconsistent results over the multiple samples indicating their lack of explanatory power compared to SREN_mv and SOL_mv. The log_TV and log_MV variables show varying results in terms of positive or negative coefficients and are mainly insignificant. The market sentiment (MS) variable shows a consistent negative relationship with the stock specific daily returns, but this result is only significant for the AAPL sample. Regarding the results of this Sentiment Analysis, it can be concluded that there is a significant relationship between sentiment variables and daily stock returns, indicating that sentiment variables can serve as predictor for daily stock returns. Therefore, it makes sense to incorporate sentiment variables in an augmented CAPM model to determine the pricing power of sentiment. Secondly, hypothesis 2 can be accepted since SREN_mv and SOL_mv show the best performance in measuring sentiment compared to the other three general lexicons.

4.4.3 Robustness checks

To assure the reliability of the sentiment analysis results some additional tests are executed. Motivated by the observation that many sentiment variables show signs of the positive sentiment bias and the study of Broadstock and Zhang (2019), a robustness check is executed by altering the measurement of sentiment variables. In this different approach the aggregated daily sentiment scores are not divided by the number of messages on the same day. This approach generates an absolute sentiment value for each trading day instead of a weighted by number of messages value. Comparing both measurements provide insights on which one is the most efficient when measuring sentiment correctly. Secondly, the altered measurements provide a better understanding of the correlation between a sentiment variable and message volume. To assess the effect of altering the measurement of sentiment variables, the same simple and multiple regression will be executed changing SREN_mv into SREN. Intuitively, SREN is defined as the sum of all sentiment scores on a single trading day (d) : $\sum SS_{t,d}$, instead of $SREN\_mv_d = \sum SS_{t,d} / \sum MV_d$ which divides the sum of all sentiment scores by the sum of all messages on the same trading day.

Table 13 in Appendix C reports the single regression results for the sentiment variables computed as absolute aggregated daily sentiment scores. First, the coefficients and corresponding significance levels are quite similar comparing these results to the prior simple regressions. However, it is worth mentioning that the R-squared of each sentiment variable drops significantly except from the SREN variables where the R-squared even increases from 13.89% till 17.5% on average. The most significant drop in explanatory power is visible in the SOL variable where the R-squared drops from 10% to 2.45%. It seems that the performance of SREN increases when the aggregated sentiment scores are not divided by the number of messages on a day, while the SOL performance decreases significantly.

This observation is supported by the correlation matrix presented in table 14 in Appendix C. In this altered approach of measuring sentiment, the correlation between SREN and SOL varies from -0.24 till 0.27 in the AAPL, AMZN and TSLA sample. However, for the GOOGL and MSFT sample the correlation stays at the same levels around 0.57 and 0.82. Since the correlations in these two samples are significantly higher in general compared to the AAPL, AMZN and TSLA sample, it could imply that the average number of tweets for these companies is not sufficient for conducting a reliable sentiment analysis. This observation is supported by the relatively low R-squared of both samples in each executed regression. Subsequently, the correlations between the daily message volume (log_MV) and all the sentiment variables except from SREN are fluctuating between 0.75 and 0.90. This high positive correlation seems to confirm that these lexicons are positively biased as daily sentiment scores tend to rise simultaneously with message volume. In contrary, SREN is negatively correlated with the log_MV variable implying that SREN is not affected by the sentiment bias. The results for the MSFT and GOOGL sample are again not in line with the others. This robustness check implies that for each sentiment variable a different approach is necessary to strengthen their performance. The SOL-lexicon is only usable when measuring sentiment based on the average relative score scale to correct for the sentiment bias, whereby the mean-score serves as the dividing line in distributing positive and negative sentiment.[15]In addition, the performance of the SREN-lexicon increases when measuring sentiment with absolute aggregated daily sentiment scores (SREN), instead of dividing these absolute scores by the daily message volume (SREN_mv).

Another robustness check is executed by altering the Kaggle-samples to the samples retrieved using the Twitter API Key. This sample is described in section 3.1.2. The same single regressions are executed for these sample to determine the reliability of the sentiment lexicons. The results presented in table 16 Appendix C support the results of the main analysis where the SREN_mv and SOL_mv variables outperform the other sentiment variables in terms of explanatory power and significance. Although the number of observations drops from 500 in the main analysis to approximately 50 in this robustness check sample, both variables remain highly significant ($p < 0.001$) and the adjusted R-squared is respectively 28.34% and 32.7% on average. The other general lexicon variables report for 3 out of 5 samples insignificant results. This indicates an inconsistent performance throughout the sample. Secondly, the adjusted R-squared of the general lexicons shows dubious results as some samples e.g., AAPL and AMZN report a negative adjusted R-squared for SR_mv and SSWN_mv.

[15] For example, the descriptive statistics presented in Table 7 Appendix B reports the following distribution for SOL_mv in the TSLA sample: minimum: 1.070, mean: 2.702 and 4.087. Normally, this indicates that none of the trading days is assigned as a negative sentiment day. Therefore, the distribution is used relatively with 2.702 as a relative dividing line between negative and positive scores.

Whereas in this robustness check the performance and explanatory power of both field-specific lexicons remains solid, the general lexicons show substantial inconsistency in significance per sample. Therefore, this robustness test strengthens the findings that field specific lexicons outperform general lexicons when measuring sentiment in tweets.

The multiple regression is also executed for the robustness test sample to identify the performance and correlations among sentiment variables and message volume. Pearson's correlation matrix shown in table 18 Appendix C confirms the consistent correlation between the SREN_mv and SOL_mv variables as all correlation coefficients vary between 0.77 and 0.88. The correlations with other variables are low and moderate. To check if the sentiment bias occurs in these robustness samples, the analysis where the sentiment variables are not divided by the daily message volume, is executed. The results presented in table 18 Appendix C show that the SREN variable outperforms the other variables in terms of significance and explanatory power. Secondly, the explanatory power is on average slightly lower when comparing with SREN_mv. As expected, the performance of SOL decreases significantly comparing with the SOL_mv. The performance of the other variables remains drastically low with even negative R-squares in some samples. The correlation matrix in table 19 reports the correlations in this analysis. The correlations between the log_MV variable and all the sentiment variables, except from the SREN variable, are consistently classified as moderate ($r > 0.4$) or mostly high ($r > 0.8$) supporting the findings with regards to the sentiment bias in the main analysis. In addition, the correlation between SREN and the message volume variable varies per sample as the correlation is -0.11 in the TSLA sample, 0.84 for AAPL and 0.61 for AMZN. Subsequently, it is remarkable that the SREN variable is highly negatively correlated with daily message volume in both index samples, with a correlation of -0.75 in the SPY sample and -0.79 in the QQQ sample. This indication is supported by the negatively skewed distribution of sentiment variable SREN_mv for both index samples as presented in table 15. An overall negative sentiment existing in the index samples while the company specific sentiment is positive, designates a contrary relationship between overall market sentiment and stock specific sentiment. This finding is supported by the results of the other sentiment lexicons, as the mean of daily sentiment scores is consistently lower for both SPY and QQQ samples in all sentiment variables. This interesting finding will be further outlined in the discussion section as eligible for further research.

The final robustness check is executed to assure the relatedness of the company specific sentiment variables towards its company returns. For example, TSLA measured sentiment variables are used as predictor for AAPL's stock return. This approach identifies if company-specific sentiment is not randomly measured and is indeed 'company-related'. Since the SREN and SOL_mv variables are classified as best performers, this test is only executed for those variables. The results of both relatedness tests are disclosed in table 21 and table 22 in Appendix C. Focusing on the explanatory power of each randomly company specific sentiment, implies that the adjusted R-squared is significantly lower compared to the results in table 16. This indicates that sentiment is not randomly measured on social media, thus for example, the AAPL tagged tweet-sample really represents investor's sentiment related to AAPL. On the other hand, it is noteworthy to mention that some non-company specific sentiment variables are highly significant ($p < 0.001$). Especially the consistent highly significant ($p<0.001$) coefficient with regards to the AAPL and AMZN relation is noticeable. This implies that AMZN specific sentiment can be used as predictor for AAPL's stock returns, and vice versa. Since both stocks are traded in the same market (S&P500) and acting in the same technology industry, this interrelation in sentiment is not classified as measurement error.

Overall, the robustness checks prove that hypothesis 1 can be accepted as a significant positive relationship between sentiment variables as predictor for daily stock returns is measured in the robustness tests as well. Note that a suitable lexicon is necessary, because the general lexicons showed insignificant results as well. Besides, the robustness tests support the acceptation of hypothesis two since both SREN_mv and SOL_mv outperform the general lexicon measurements. The performed robustness check focused on an altered measurement of each sentiment variable provided a useful finding. It seems proper to use the SREN variable rather than the SREN_mv variable since both significance and explanatory power increased when applying this altered measurement. Thus, in the remainder of this study the SREN measurement will be used instead of SREN_mv.

5. Research method CAPM-analysis

This chapter will outline the research method with regards to the CAPM-analysis. This analysis is conducted to measure the pricing power of sentiment in an asset pricing model. Regarding the prior sentiment analysis, only the SREN and SOL_mv will be considered as sentiment variables. The gathered data used in this analysis is previously described in chapter 3. The dependent, independent and control variables used in this analysis are previously defined in the section 4.3. This section only elaborates on the research method with regards to the CAPM-models and distributed lag model.

5.1 Simple CAPM

The original simple CAPM model was developed by Sharpe (1964) and Lintner (1965) as first step in asset pricing theory. 'The CAPM explains the trade-off between assets' returns and their risks, measuring the risk of an asset as the covariance of its returns with returns on the overall market' (Rossi, 2016). The general CAPM relationship developed by Sharpe (1964) and Lintner (1965) is formulated in the following way:

$$E(r_i) = r_f + \beta_i(E(r_m) - r_f) \text{ (10)}$$

Where $E(r_i)$ is the expected return on an individual asset i, r_f is the risk-free rate in the market. The risk-free rate is defined as a return which "an investor can expect to earn on an investment that carries zero risk. In practice, the risk-free rate is commonly considered to equal to the interest paid on a 3-month government Treasury bill, generally the safest investment an investor can make." (Corporate Finance Institute)[16]

B_i measures the sensitivity of the individual assets returns to the market return r_m. Subsequently, the difference between the expected market return $E(r_m)$ and the risk-free rate in the market r_f is called the risk premium (Rodriguez, 2016; Sharpe, 1964).

A simple CAPM equation is used in this analysis as basis model. The dependent variable in this CAPM is the log-difference return (RC_{it}) of an individual stock defined as:

$$RC_{ti} = 100 * (ln(C_{ti}) - ln(C_{t-1,i})) \text{ (11)}$$

Where C_{ti} is defined as the closing price of an individual stock on a certain day t. The risk-free rate is retrieved using the YahooFinance application in R searching on ticker '^IRX' which associates with the 13-week treasury bill recorded by the U.S. Department of the Treasury. As the individual asset returns and the market returns are determined as log-difference returns, the daily risk-free rates are also defined as logarithmical rates using the following modification:

$$Rf_{it} = ln(1 + CIRX_{it}) /365 \text{ (12)}$$

[16] Source: Corporate Finance Institute

This modification is in line with prior research to obtain equally weighted daily log-returns and daily log risk-free rates (Broadstock & Zhang, 2019; Yobero, 2018) The number of observations in the upcoming CAPM models are reduced by 6 for each sample due to non-availability of risk-free data on six trading days.

Following Broadstock and Zhang (2019) and Yobero (2018), the alpha and beta of the simple CAPM model are determined by regressing the log returns of the specific company on the market risk premium, using the following regression equation:

$$RC_{it} = \alpha + \beta(Rm - Rf) \text{ (13)}$$

Where RC_{it} is defined as the log difference returns calculated using equation 13. Alpha and beta will be derived using this formula and the risk premium is defined as the difference between the log returns in the market minus the log risk free rates (Rm – Rf). When the alphas and betas are estimated the expected returns of each stock can be calculated.

5.2 Sentiment augmented CAPM

To evaluate the pricing power of sentiment in a Capital Asset Pricing Model a sentiment augmented CAPM model is constructed using a simple CAPM augmented with sentiment and control variables. The log-difference company specific and market return, the risk-free rate and the determination of the alpha and betas are used in the same order as previously described in section 5.1. Subsequently, the results obtained from the sentiment analysis have determined to use SREN and SOL_mv as sentiment variables.

Finally, adding the control variables as described in section 4.3.4., leads to the following augmented CAPM equation:

$$RC_{it} = \alpha + \beta(Rm - Rf) + \beta 1SREN + \beta 2SOL\_mv + \beta 3log\ \_TV + \beta 4\, log_{MV} + \beta 5MS + \epsilon_i \text{ (14)}$$

The model is checked on the same fundamental assumptions as previously described in section 4.3.5. The outliers with regards to message-volume were already removed and the same HAC-standard robust errors are used to control for heteroskedasticity and autocorrelation. The purpose of comparing a simple CAPM-model with a sentiment augmented CAPM addresses to hypothesis 3.

To evaluate the explanatory power and prediction accuracy of both CAPM models three methods are used. First, the explanatory power of both models will be compared based on the difference in adjusted R-squared. A substantial increase in this metric implies that augmented a simple CAPM model with sentiment variables improves its explanatory power. Secondly, the prediction accuracy of both is analysed executing an Analysis of Variance (ANOVA) test. The assumption of an ANOVA test is to identify if adding more independent variables is justifiable to fit the data appropriately. Intuitively, comparing a simple model with just one independent variable with a complex model using ten independent variables, the complex model needs to fit the data much better to justify its complexity.

Therefore, the Analysis of Variance (ANOVA) test implies if the more complex augmented CAPM model shows a significantly better of data compared to the simple CAPM.

The third approach to measures the accuracy difference of both models when comparing predicted values with the actual log returns. This prediction accuracy test is executed in two different formats. First, an in-sample test is constructed splitting the initial sample in a training set and test set. The split ratio is determined at 80:20 since this ratio is commonly used among data-scientist[17]. A training set of 80% of the initial sample is used to conduct both CAPM models. Subsequently, these computed models are used to predict the 20% test set of each company specific sample. Secondly, the CAPM specifications are both used to execute a true out-of-sample test. Contrary to sample-splitting, a true out-of-sample test is executed with a completely new dataset containing new independent observations. It is assumed that using completely new observations improves efficiency and provides the only real validation of a model (Anscombe, 1967; Fang, Jacobsen, & Qin, 2014)[18]. The predicted log-returns using the simple – and augmented CAPM models are examined in terms of accuracy with the actual values. The accuracy function generates multiple measurements as the Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE). All five measurements of accuracy are computed for all samples. The MAE and RMSE measurements are currently the most widely used metrics in economic profession. The Mean Absolute Error (MAE) computes the average magnitude of all absolute errors between predicted and actual observation. The Root Mean Squared Error (RMSE) measures the same average of errors but uses the squared root average to put heavier weights on larger errors. When comparing the accuracy, the focus will be on the MAE and RMSE metrics.

5.3 Distributed lag model

To further evaluate the effect of sentiment on stock returns a distributed lag model is constructed. Referring to the current literature it is assumed that lagged-sentiment effects contain explanatory power with regards to current stock prices (Smailovic et al., 2013; Sprenger et al., 2014; Sul et al., 2016). However, there is no clear consensus if lagged-sentiment effects are positively and negatively related to daily stock returns. Therefore hypothesis 4 was formulated. A distributed lag-model will identify the relationship between stock returns and lagged sentiment effects. Intuitively, a lagged sentiment effects determines if the sentiment measured on yesterday (t-1) influencing the stock returns of today. The dependent variable in this model is previously described in section 4.3.3. and defined as R_{it}. Based on the results of the executed sentiment analysis, only sentiment variable SREN and SOL_mv are used in this distributed lag model. The distributed lag model regression specification is formulated as follows:

[17] This split ratio of 80:20 is determined based on the information with regards to the following source: https://www.journaldev.com/45019/split-data-into-training-testing-sets

[18] Note that this true out of sample test is only executed with regards to the AAPL sample.

$$R_{it} = \alpha + \beta 1 SREN_{i,t} + \beta 2 SREN_{i,t-1} \dots + \beta 11 SREN_{i,t-10} + \varepsilon_{it} \quad (15)$$

Where R_{it} is defined as the daily stock return for stock i at day t. Alpha represents the constant of the regression specification. $\beta 1 - \beta 11$ are the estimations of the sentiment variable on the current day t till a lag of t_{-10}. Intuitively the sentiment variable at lag-10 measures the effect of sentiment 10 days prior to current day t.

The use of ten lags is argued be comparing the Akaike Information Criteria (AIC) values for selecting optimal lags, because using too few lags can lead to significant serial correlation. In three of the five stock samples the optimal lag number is established on ten lags. To ensure the same number of lags for all samples, the regression specification is executed with ten lags included for all samples. The current day effect of sentiment at day t is incorporated to examine whether this expected positive significant coefficient stays positive, or switches to negative when regarding lagged sentiment.

6. Results CAPM analysis

This section contains the results of the CAPM-analysis as outlined in chapter 5. Section 6.1 elaborates on the results of the simple CAPM analysis compared to the augmented CAPM analysis. Section 6.2 outlines the results with regards to the distributed lag model.

6.1 Simple CAPM and augmented CAPM

In this paragraph the main results with regards to hypothesis three are outlined. The third hypothesis states that the pricing power of a Capital Asset Pricing Model (CAPM) improves with the inclusion of sentiment variables on a daily time frame. Table 2 presents the results of estimating alpha and beta for each stock to derive the simple CAPM equations. The alpha of each equation is on average 0.004. The Beta of each equation varies in a range between 1.289 till 1.509 assuming that the risk of each stock is theoretically higher than the SPY. Therefore, investors demand compensation with a return that justifies the risk of the asset. The Betas are all significant ($p < 0.001$) as expected, because it is assumed that stock returns are significantly affected by the risk premium in the market. Subsequently, the adjusted R-squared varies from 54.97% till 71.42% for the AAPL, AMZN, GOOGL and MSFT sample. The adjusted R-squared for the simple CAPM-model is with 13% notably lower for the TSLA sample.

Table 2: Simple CAPM Alpha and Beta determination

Sample	Alpha (α)	Beta (β)	Adjusted R-squared
TSLA	0.0041**(2.819)	1.289***(8.632)	0.13
AAPL	0.0046***(8.478)	1.359***(24.43)	0.5497
AMZN	0.0048***(8.139)	1.509***(24.94)	0.5599
MSFT	0.0048***(12.64)	1.369***(35.08)	0.7142
GOOGL	0.0039***(7.992)	1.331***(26.43)	0.5859

Table 2 reports the results of computing the alpha and beta for each simple CAPM model. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

The next part in this analysis elaborates on the augmented CAPM model results as defined in equation 14. Furthermore, the results will be compared with the simple CAPM results as previously defined. The descriptive statistics of this model are disclosed in table 22 Appendix D. As previously mentioned in section 5.1, the number of observations is reduced by six for each sample due to non-availability of the 90-day treasury bill data. The log-returns (RC_{it}) of each stock are normally distributed around a mean of approximately zero. The other variables have already been disclosed in the sentiment analysis. Pearson's correlation matrix regarding the augmented CAPM model is disclosed in table 23 Appendix D. The correlation between the log-returns (RC_{it}) and the risk premium (Rm-Rf) is the highest fluctuating between 0.74 and 0.85. The TSLA sample reports a significantly lower correlation of 0.36 between both variables which is related to the much lower adjusted R-squared in the simple CAPM model. The correlation between the log returns and both sentiment variables is varying between 0.18 and 0.54. Since two sentiment variables are included in the same CAPM model it makes sense to check for multicollinearity. Multicollinearity occurs when independent variables are highly correlated, which implies that one independent variable can be predicted by the model itself. Therefore, the estimated coefficients and its significance level will be affected. (Mansfield & Helms, 1982). To determine if multicollinearity is a problem in the model, the variance inflation factor (VIF) is calculated for all explanatory variables. As disclosed in table 24 Appendix D all VIF values are amply within the critical range of 10. Therefore, both sentiment variables are included in the model as they are complementary instead of correlated.

The output of the augmented CAPM-model is represented in table 3. As expected, the risk premium and the SREN variables show a positive and significant ($p < 0.001$) relationship with the log returns. The SOL_mv variable shows alternately positive and negative coefficients, whereby only the relationship in the AAPL sample is significant. This lack of significance is unexpected, because this variable showed equal significance as SREN. The log_TV coefficient is positive and significant for all samples disregarding GOOGL. This positive relationship indicates that stock returns will increase as trading volume increases. The log_MV and MS coefficients are inconsistent and not significant. The explanatory power of the augmented CAPM model is generally higher compared with the simple CAPM model, because all adjusted R-squares increased. The biggest improvement in explanatory power is reported in the TSLA sample as the R-squared increased from 13% to 33.72%.

Table 3: Results Augmented CAPM model

Variable	TSLA	AAPL	AMZN	MSFT	GOOGL
α	-0.281***(-4.579	-0.066*(-1.976)	-0.066*(-2.437)	-0.04(-1.602)	-0.011(-0.275)
β(Rm-Rf)	1.127***(8.551)	1.129***(18.634)	1.36***(17.282)	1.328***(27.04)	1.301***(25.78)
SREN	0.0002***(11.758)	0.0001***(7.443)	0.0001***(5.486)	0.0001***(3.875)	0.0001***(2.906)
SOL_mv	-0.002(-0.662)	0.002**(2.571)	-0.0004(-0.392)	0.001(1.180)	-0.0001(-0.145)
log_MV	-0.0001(-0.041)	0.001(0.474)	-0.001(-0.46)	-0.002(-1.174)	-0.003(-1.547)
Log_TV	0.017***(4.324)	0.003*(1.744)	0.005**(2.766)	0.003**(2.200)	0.002(0.762)
MS	-0.001(-0.242)	-0.004***(-6.411)	-0.0002(-0.297)	0.0002(0.404)	-0.0003(-0.434)
Adjusted R-squared	0.3372	0.674	0.645	0.744	0.6218

Table 3 reports the results of the augmented CAPM model. The coefficients are reported for each variable, followed by the significance level (*** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.1$) with corresponding t-statistic between the brackets.

To ensure the statistical validity of this assumption an Analysis of Variance (ANOVA) test is executed. The ANOVA test identifies whether the more complex augmented CAPM-model fits the data significantly better than the simple CAPM model. As reported in table 25 in Appendix D the augmented CAPM model shows a lower variance leading to a significantly improved fit, compared to the simple CAPM model.

To further exploit the differences in explanatory power, both CAPM models are fitted to predict the company specific log-returns using a sample-splitting approach of the initial sample and a true out-of-sample database. First, the results of the sample splitting tests are disclosed in table 26. The accuracy metrics show different results for each company specific sample. For AAPL and MSFT the RMSE and MAE metrics are slightly lower for the augmented CAPM model indicating better accuracy in predicting log-returns. On the other hand, the results for TSLA and AMZN show the opposite as both metrics are slightly higher for the augmented CAPM model, implying inconsistency between the stock specific samples. The GOOGL sample reports significantly better performance of the simple CAPM model as the average errors using the augmented CAPM model are substantially higher. The other metrics, ME, MPE and MAPE, show inconsistent results between the different samples. Due to the ambiguous results in terms of prediction accuracy it is hard to draw conclusions with regards to the in-sample accuracy test. Accordingly, the results of the true-out of sample prediction are disclosed in table 27. Except from the MAPE metric, all other metrics provide a consistent higher error in the augmented CAPM model. This indicates that a simple CAPM model provides better accuracy when executing a true-out of sample prediction.

Overall, the reported results showed that the explanatory power of a CAPM model augmented with sentiment variables increases in comparison with a simple CAPM model. These results are supported by the executed ANOVA test results because the augmented CAPM model justifies a significantly improved fit. On the other hand, the prediction accuracy results measured with the five defines metrics show inconsistent results. Regarding the consistent results of the adjusted R-squared comparison and the ANOVA test, hypothesis three can be accepted. However, one should note the inconsistency observed in prediction accuracy. This could have implications for the acceptance of hypothesis three.

6.1.1 Additional tests

This section provides additional test results focusing on the effect of changing sentiment variables in the augmented CAPM model. First, the augmented CAPM analysis is executed replacing the SREN variable with the SREN_mv variable. Prior robustness checks executed in the sentiment analysis revealed that an altered measurement of sentiment (SREN) showed better performance in comparison to the SREN_mv variable. To ensure the reliability of this finding it make sense to test it again in this augmented CAPM model. Table 28 shows the results of this robustness check. The results indicate that an augmented CAPM model fitting SREN_mv instead of SREN leads to less explanatory power denoted by the adjusted R-squared. Therefore, the use of SREN instead of SREN_mv in an augmented CAPM model is approved.

Another additional test is executed to determine the effect of including both sentiment variables together in one model. This additional robustness test includes both sentiment variables separately in the augmented CAPM model. Table 29 reports the results regarding the inclusion of SREN as only independent sentiment variable. As expected, the results show highly significant positive coefficients for each sample. Focusing on the adjusted R-squared, the explanatory power of this SREN-only CAPM model is identical to prior results of the augmented CAPM including both sentiment variables. This indicates that the inclusion of both sentiment variables in the same model does not add any value. Although it is not harmful in terms of multicollinearity to include both sentiment variables, tit is not conducive either. Table 30 report the results of the CAPM-model using SOL_mv as only sentiment variable. The explanatory power of this model is consistently lower compared to the SREN- CAPM model. Regarding these additional tests, it can be concludes that combining two related sentiment variable is not beneficial in terms of explanatory power. Secondly, it reveals that the field-specific lexicon of Renault (2017) shows slightly better measurement performance compared to the lexicon constructed by Oliveira et al. (2016)

6.2 Lagged effects

This section elaborates on the main results with regards to the fourth hypothesis. Hypothesis four states: there is a positive relationship between lagged sentiment effects and stock returns on daily basis. Table 4 reports the results of the distributed lag models with SREN and SOL_mv as explanatory lagged variables predicting R_{it}. The results confirm the positive and significant relationship between the sentiment-variables and stock returns without lag. This positive relationship changes to negative for all significant lagged SREN variables. The AAPL, TSLA and MSFT show negative and significant ($p < 0.05$) coefficients at lag-1. Intuitively a negative significant coefficient at lag-1 implies that yesterday's (t-1) sentiment shows a negative relationship with today's stock returns. Furthermore, the other significant coefficients diffused over the lags and samples show all negative coefficients. The presence of significant lagged effects shows that sentiment on days prior to day t still effect the current stock return on day t. Moreover, it is remarkable that the relationship changes to negative as shown by the negative significant coefficients. Especially lag-1 and lag-2 show these results varying per sample. The coefficients of all lagged sentiment effects are substantially lower than the positive coefficient at today or in other words lag zero. For example, Table 4 reports a positive significant coefficient of 0.0185 at lag zero for the TSLA sample, compared to a negative significant coefficient of -0.0067 at lag-1. This implies that the positive effect of current days sentiment on today's stock return outweighs the lagged effect in terms of impact. Accordingly, these effects hold throughout all samples.

Besides the identified effects at lag-1 and lag-2, other significant negative coefficients are reported around lag-8. AMZN reports a highly significant negative coefficient ($p < 0.001$), followed by a negative significant coefficient as well for AAPL ($p < 0.10$) and MSFT ($p < 0.05$) using the SREN variable. Regarding the AMZN sample, the impact and significance of this effect at lag- is higher compared to lag-1 and lag-2 for both the SREN and SOL_mv measurements. Since this effect at lag-8 is measured within 3 out of 5 samples it seems to make sense. However, there is no clear explanation for this effect at lag-8. The results in terms of significance are approximately the same comparing the use of SREN and SOL_mv in this distributed lag model. Overall, the distributed lag model shows consistent results when analysing the daily lagged effects. Expect from the GOOGL sample, all samples show negative coefficients at lag-1 and lag-2. Hereafter, the coefficient sign shows inconsistency, but all significant results throughout the whole sample report negative coefficients. Therefore, the fourth hypothesis can be rejected, implying that there is no positive relationship between lagged sentiment effects and stock returns on daily basis.

Table 4: Lagged effects of the SREN and SOL_mv variables

SREN	AAPL	TSLA	AMZN	MSFT	GOOGL
Intercept	-0.017(-0.231)	0.3711*(2.355)	0.096(1.298)	-0.137(-0.752)	0.0063(-0.428)
Lag 0	0.0117***(14.618)	0.0185***(12.754)	- 0.0205***(12.412)	0.0256***(8.224)	0.0129***(6.473)
Lag 1	-0.0023*(-2.573)	-0.0067***(-4.200)	-0.0019(-1.092)	-0.0071*(-2.105)	0.0028(1.308)
Lag 2	-0.0014(-1.603)	-0.0039*(-2.414)	-0.0032+(-1.756)	-0.0027(-0.813)	-0.0036+(-1.678)
Lag 3	-0.00048(-0.533)	0.0012(0.770)	-0.0016(-0.859)	0.0015(-1.290)	-0.0029(-1.368)
Lag 4	0.00044(0.496)	-0.0030+(-1.885)	-0.0009(-0.479)	0.0015(0.454)	-0.0025(-1.186)
Lag 5	-0.0017*(1.993)	-0.0011(-0.656)	-0.0028(-1.446)	-0.0024(-0.716)	0.0022(-1.020)
Lag 6	0.00044(0.494)	0.0011(0.688)	0.0014(0.733)	0.0012(0.354)	0.0009(0.417)
Lag 7	0.00052(0.584)	0.00054(0.341)	-0(-0.007)	-0.0001(-0.039)	-0.0004(-0.177)
Lag 8	-0.0015+(-1.718)	-0.00036(-0.227)	-0.0064***(-3.410)	-0.0069*(-2.050)	-0.0012(-0.547)
Lag 9	-0.0005(-0.553)	-0.0015(-0.927)	-0.0039*(-2.067)	0.0013(0.380)	0(0.044)
Lag 10	-0.0006(-0.744)	0.0023(1.589)	-0.0018(-1.037)	0.0028(0.901)	0.0018(-0.857)
Adjusted R^2	0.3282	0.2545	0.2758	0.12	0.078

SOL_mv	AAPL	TSLA	AMZN	MSFT	GOOGL
Intercept	-1.045(-1.939)	-0.8835(-0.664)	-0.7205(-1.362)	-0.3858(-0.860)	-0.2795(-0.803)
Lag 0	1.566***(12.56)	1.832***(5.002)	1.599***(11.619(	0.8609***(8.169)	0.4824***(5.446)
Lag 1	-0.301*(-2.213)	0.4825(-1.233)	-0.1839(-1.238)	-0.1743(-1.543)	0.055(0.606)
Lag 2	-0.514***(-3.773)	1.063**(-2.721)	-0.2599+(-1.735)	-0.1686(-1.474)	-0.162(0.606)
Lag 3	-0.172(-1.258)	-0.0023(-0.006)	-0.3044*(-2.023)	-0.0456(-0.397)	-0.1625+(-1.778)
Lag 4	0.151(1.106)	-0.658+(-1.691)	0.045(0.301)	-0.0617(-0.534)	-10.35(-1.13)
Lag 5	-0.2341+(-1.713)	0.1742(0.447)	0.049(0.332)	-0.166(-1.438)	0.033(0.364)
Lag 6	-0.0046(-0.034)	0.2551(0.656)	-0.2858+(-1.908)	0.1673(1.447)	-0.041(0.431)
Lag 7	0.129(0.945)	0.2196(0.564)	0.1816(1.208)	-0.157(-1.361)	0.093(1.006)
Lag 8	-0.2488+(-1.827)	0.069(0.179)	-0.5608***(-3.755)	-0.223+(-1.938)	-0.041(-0.442)
Lag 9	-0.1116(-0.820)	-0.2287(-0.588)	0.2732+(1.846)	0.0589(0.517)	-0.1155(-1.268)
Lag 10	0.085(0.685)	0.2505(0.691)	-0.2635+(-1.911)	0.094(0.862)	-0.091(-1.001_
Adjusted R_2	0.2638	0.053	0.2402	0.1231	0.013(0.148)

Table 4 reports the results of the lagged variables analysis. The coefficients are reported for each time lag per variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

7. Discussion

Investor's sentiment measured in social media plays a significant role as driver of stock prices. As illustrated by the Reddit-community during the GameStop buying regime, it is even possible to overrule all fundamentals of the market by steering stock prices via social media. A growing body of research investigated the influence of sentiment as predictive variable for stock returns. As prior research mainly focuses on the directional effects of sentiment variables on stock returns, this study focuses on the effects of implementing sentiment variables in a traditional CAPM asset pricing model. Since there is no consensus on which method gives the best performance in measuring sentiment, this study aims to test five lexicon-based approaches to obtain the best method. To provide an answer on the research question, the lexicon-based methods which incorporates the most prediction power are implemented in an augmented CAPM model. This chapter discusses the findings of the results chapter while considering the formulated hypotheses. Subsequently, the findings will be compared with prior research. Hereafter, the limitation of this research will be outlined followed by recommendations for further research. The final section will provide a brief conclusion in combination with an elaboration on the contribution of this research.

7.1 Interpretation of the results

The results of the sentiment analysis show that in general, sentiment measured on social media contains predictive power with regards to daily stock returns. In this analysis the measurement of sentiment is done using five different lexicon-based approaches. Each sentiment variable is included as predictor in a simple OLS regression to predict stock returns in the same manner. Although each sentiment variable shows a positive significant relationship with daily stock returns, there are substantial differences in terms of explanatory power between each sentiment variable. Therefore, it seems reasonable to use an appropriate field-specific lexicon when measuring sentiment on social media. This statement obtains further strength when considering the multiple regression and additional robustness test results. So, when reviewing the results of the sentiment analysis, it can be concluded that social-media sentiment is usable as significant predictor of stock returns. In practice, this relationship implies that sentiment of the investor community diffused on social media causes stock prices to change. Thus, for example if the investor community spreads positive facts, thoughts and opinions about Apple, the price of Apple-stocks will rise during the same day to some extent. This effect is theoretically supported by the idea that investors make decisions based on their emotions, feelings which are influenced by the content spread on social media. The second key finding of analysis one relies on the approach to measure sentiment. To prevent measurements errors and attain more explanatory power, it is significantly proved that field-specific investor lexicons outperform widely accepted general lexicons.

The second main analysis focused on the pricing power of sentiment in a Capital Asset Pricing Model. When looking at the results of the simple CAPM model in comparison with the augmented CAPM model, it can be concluded that the inclusion of sentiment variables in a simple CAPM model increases the explanatory power of the model. Those results are consistent throughout the full sample implying that sentiment contains significant pricing power. Regarding the TSLA sample, it is noteworthy that sentiment even outperforms the stocks risk premium, implied by an increased adjusted R-squared from 13% to 33% with the inclusion of sentiment. The results of a better performing CAPM when including sentiment variables is supported by the ANOVA test. On the other hand, the prediction accuracy of both showed equal performance. These mixed results give rise to remain cautious when concluding that a sentiment augmented CAPM outperforms a simple CAPM. However, since the first analysis showed significant results that sentiment serves as predictor for stock returns, it is intuitively acceptable to include sentiment variables in asset pricing models.

The final analysis addresses to the relationship between daily lagged sentiment effects and stock returns. In general, the results reported negative coefficients at one or two lags for each sample, but the results are not consistently significant. However, all observed significant coefficients reported negative signs, serving as an appropriate observation to reject the assumption of a positive relationship between lagged sentiment and stock returns. Since all significant observations report negative coefficients, it is arguable to rely on a negative relationship.

7.2 Findings in comparison with prior research

The findings of this study can be compared with prior research in several manners. As previously stated, it can be concluded that sentiment measured on social media shows a significant positive relationship with daily stock returns. This indicates that sentiment could serve as predictor for stock returns. When looking at prior research a growing body of research agrees on this finding. For example, Sul et al. (2014) found also highly significant results for social media sentiment as predictor for the same day return. Sprenger et al. (2014) found comparable results since ‘bullishness’ acted as highly significant predictor for daily stock returns. On the other hand, both Sprenger et al. (2014) and Sul et al. (2014) reported an R-squared around zero in most of their analysis while this study found an adjusted R-squared around 15-25% focusing on the best performers in terms of sentiment measurement. This contrast in explanatory power is likely caused by dissimilar methods to measure sentiment. Sul et al. (2014) and Sprenger et al. (2014) used respectively the Loughran and McDonalds (2011) lexicon and a machine learning algorithm. Other studies are more focused on directional accuracy when using sentiment as predictor for stock returns. Renault (2017) provided evidence that social media sentiment shows a positive relationship with intraday stock returns. These findings implies that sentiment acts as predictor for stock returns on multiple timeframes.

After Oliveira et al. (2016) and Renault (2017), this research is the first to examine the performance of field-specific lexicon in comparison to general lexicons. This study found that Oliveira’s and Renault’s lexicon significantly outperformed the NRC, SentiWordNet and SentimentR lexicons. These results are in line with Oliveira et al. (2016) and Renault (2017). Both studies showed that their constructed field-specific lexicons provided a more accurate measurement of social media sentiment compared to other different general lexicons.

This research is the first study in this field of research, which includes sentiment in a CAPM model on a daily timeframe. The results showed that sentiment increased the explanatory power of a simple CAPM model. Additionally, the prediction accuracy of both models showed similar results. Broadstock and Zhang (2019) used a CAPM model to identify the pricing power of emotion on intraday time-intervals. Their purpose was to measure the relationship of different emotional states on stock returns. In contrast with this study, they found in some occasions negative coefficients of positive emotional states, which suggest that sentiment could act as negative predictor for stock returns on short time frames. This study only found positive relationships for sentiment included in the augmented CAPM model.

Furthermore, this research has found that there is no positive association between daily lagged sentiment and today's stock returns. This is contrary with the results of Smailovic et al. (2013), which do find that 2 and 3 day lagged sentiment is positively related to current day stock returns. Sprenger et al. (2014) showed similar results in lags of one and two days. This study found a contrary relationship since all significant coefficients were negatively signed. Most of those significant results occurred for one- and two-day lags. Mo et al. (2016) found a similar negative relationship consistently observable at a lag of five days. They theorized this finding as a correction of a previous overreaction. One should note that Mo et al. (2016) used news sentiment instead of social media sentiment. Broadstock and Zhang identified inconsistency with regards to lagged sentiment on timeframes of 30-minutes.

7.3 Limitations and foundation for further research

When reviewing this research some limitations can be found. The first limitations can be attributed to the data sample. Due to the lack of data availability only five specific companies are used in the main analysis. These firms are all traded in the same S&P500 index, located in the US markets. Additionally, these five firms are all presented in the top seven of largest market caps across the world which could jeopardize the generalizability of the research results. Secondly, this study focuses on one specific period (2018-2019) in the markets. Since market conditions and investor behaviour on Twitter changes rapidly over time, the observed results may be time dependent (Liu et al., 2014; Sul et al., 2016). Besides, this study only considers Twitter data to measure investors sentiment on social media, because it provides a decent API key to gather data. Currently Twitter is not the only social media platform used by the online investors community to share their thoughts, feelings and opinions. Especially the platforms of StockTwits, Reddit and Robinhood are intensively used by investors. The significant role of Reddit as investors platform has shown previously during the GameStop-scene. Therefore, this study could be extended by using a combination of multiple social media platforms to retrieve messages to measure sentiment. Additionally, a wider and more diverse stock sample would improve the generalizability of these research results.

Another limitation can be attributed to the relationship between sentiment and daily stock returns. This research assumes that sentiment is a predictor for daily stock returns, therefore the sentiment variables are used as explanatory variables to determine stock returns. However, prior research identified a contemporaneous relationship between sentiment and stock returns (Mo et al., 2016; Smailovic et al., 2013; Sprenger et al.,2014). This contemporaneous relationship suggests that stock returns contain predictive information for sentiment as well, instead of just a one-sided relationship where sentiment only explains stock returns. To test this relationship a Granger Causality analysis is executed. A Granger causality test provides statistical evidence if it can be shown that values of variables X granger causes the future values of variable Y. This implies that the lagged values of variable X have a statistically significant relationship with variable Y (Granger, 1969). In this research the Granger causality test is applied in two directions with the following null-hypothesis in the model; (1) a certain sentiment variable (SREN_mv or SOL_mv) does not predict daily stock returns and (2) daily stock returns do not predict certain sentiment variables. The results of the Granger Causality test reported in table 31 show varying significance levels for each sample. Overall, it can be concluded that the tests executed with the sentiment variable granger causing the stock returns show higher significance compared to the opposite direction. This implies that the predictive power of sentiment variables on daily stock returns surpasses the opposite causal relationship where daily stock returns predict sentiment. However, zooming in on the AAPL sample provides evidence that the relationship between sentiment and stock returns could be two-sided and therefore contemporaneous.

The remarkable findings presented in robustness checks 4.1.3. could serve as a foundation for further research. These findings provided an indication that market-sentiment moves contrary in relation to stock specific sentiment, as on average the sentiment scores are consistently lower for each sentiment variable in both index samples compared to the stock specific samples. Subsequently, the SREN_mv variable even shows an average sentiment score of -0.10 for both index samples compared to means between 0.0334 and 0.0739 for the stock-specific samples. This is a relatively big difference between the general market sentiment in relation to company specific sentiment. An explanation of this phenomenon can be found in risk-taking behaviour of investors. It is commonly known that investing in single stocks involves more risk compared to investing in an index due to diversification. Risks are rewarded with returns in general, therefore investors tend to take more risk lured by higher returns (Egozcue, Garcia, Wong, & Zitikis, 2011). Especially when the markets are in a bull market, investors tend to take more risk leading to overconfidence and excessive trading behaviour (Trinugroho & Sembel, 2011). As the current markets move in a bull market, it makes sense to dedicate this difference in sentiment between indexes and stocks towards investor overconfidence. The idea is that investors think they can beat the market indices by investing in stocks, causing a negative sentiment with regards to indices and a positive sentiment with respect to specific stocks. This interesting intuition could serve as a foundation to further research using an appropriate sample with enough observations.

Another foundation of further research can be found in the remarkable observation that the TSLA sample showed a significantly lower explanatory power of the simple CAPM compared to the other samples. The adjusted R-squared was just 13% while the other samples had a minimum adjusted R-squared of 55%. Subsequently, when including a sentiment variable during the augmented CAPM analysis, the R-squared improves from 13% to 33%. This substantial improvement suggests that sentiment has more explanatory power than TSLA's risk premium. Therefore, it seems reasonable to suggest that some stocks are significantly driven by social media sentiment as main predictor. Referring to the GameStop buying regime driven by the online investors' community on Reddit, it makes sense to further investigate the role of social media sentiment as main predictor for stock returns.

7.4 Contribution and concluding remarks

This study contributes to the existing literature by providing a better understanding of the impact of social media sentiment on daily stock returns. The first analysis found that social media sentiment, when measured in an appropriate way, serves as a significant predictor of company specific stock returns. Furthermore, this study examined the impact of sentiment in a widely accepted Capital Asset Pricing Model and found that sentiment contains significant pricing power when included in this model. Both findings contribute to the growing body of research who consider the significant role of sentiment as relative new authority in investment decision making theory. Accordingly, the advent of social media as important platform in the investors community to spread opinions, thoughts, feelings and information, emerges as significant factor to diffuse and measure sentiment. Since Jong et al. (2017) in 2017 already found that 34%-70% of all investors uses social media to some extent in their investment decision making, it makes sense to assume that social media provides a valuable proxy to measure investor sentiment. Applying such sentiment in a widely accepted asset pricing model in economic theory, provides a steppingstone to consider social media sentiment as an influential factor in the current economic profession.

Another contribution of this study relates to the measurement of sentiment on social media platforms. One of the main issues with regards to implementing social media sentiment in economic theory, is a lack of consensus, transparency, replicability and skill, when measuring investor sentiment on social media. As previously defined, many researches use machine learning algorithm methods when measuring sentiment. This approach is time-inefficient, not transparent and therefore not replicable and difficult to compare. On the other hand, lexicon-based approaches solve these problems, but the vast majority of research still uses inappropriate general lexicons, leading to poor and biased performance when measuring sentiment on social media. Oliveira et al. (2016) and Renault (2017) provided field-specific lexicons serving an appropriate, transparent and easier way to measure sentiment on social media. However, both lexicons are still under the radar and not used in any later published study. This study compared the NRC, SentiWordNet, SentimentR, Oliveira et al. (2016) and Renault (2017) lexicons and found that both Oliveira et al. (2016) and Renault's (2017) lexicon, classified as field-specific lexicons, significantly outperform the general lexicons when measuring sentiment on social media. This key finding encourages to a general use of field-specific lexicons in this field of research. This contributes to the process of creating a consensus on how to appropriately measure sentiment on social media, which will improve the transparency, replicability and wider accessibility in this field of research. In the end, it would make sense to add one of those field specific lexicons as package in data analytics software programs e.g., R and Python. Currently only general lexicons as NRC, Afinn, Bing and SentiWordNet are provided as packages in the programs. Improving the measurability of sentiment on social media will directly contribute to a wider acceptation of social media sentiment as influential factor in economic profession.

8. References

Affuso, E., & Lathinen, K. D. (2019). Social media sentiment and market behavior. *Empir Econ, 57*, 105-127.

Alostad, H., & Davulcu, H. (2016). Directional Prediction of Stock Prices using Breaking News on Twitter. *Web Intelligence and Agent Systems: An International Journal* , 5, 1-5.

Anscombe, F. (1967). Topics in the investigation of linear relations fitted by the method of least squares. *Journal of the Royal Statistical Society. Series B, 29(1)*, 1-52.

Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 29(3), 1259-1294.

Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets, 7*, 271-299.

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

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, 21(2)*, 129-151.

Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter Help Predict Firm-Level Earnings and Stock Returns? *The Accounting Review*, 93(3), 25-57.

Batra, R., & Daudpota, S. (2018). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. *International Conference on Computing, Mathematics and Engineering Technologies.* Sukkur: Sukkur IBA University.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Comptutational Science*, 2(1), 1-8.

Broadstock, D. C., & Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters*, 30, 116-123.

Bukovina, J. (2016). Social media big data and capital markets - An overview. *Journal of Behavioral and Experimental Finance*, 11, 18-26.

Byrne, B. M. (2010). *Structural equation modelling with AMOS: Basic concepts, application and programming.* New York: Routledge.

Cootner, P. H. (1966). The Random Character of Stock Market Prices. *The Journal of Business, 39(4)*, 532-535.

Corea, F. (2016). Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector. *Big Data Research*, 4(1), 70-74.

Corporate Finance Institute. (2021). Risk-Free Rate. What is a Risk-Free Rate and why is it important? Retrieved on 26-06-2021, from: https://corporatefinanceinstitute.com/resources/knowledge/finance/risk-free-rate/.

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 1461-1499.

Dewally, M. (2003). Internet Investment Advice: Investing with a Rock of Salt. *Financial Analysts Journal, 59(4)*, 65-77.

Dhakal, C. P. (2017). Dealing with outliers and influential points while fitting regression. *Journal of Institute of Science and Tehcnoglogy, 22(1)*, 61-65.

Dolan, R. (2002). Emotion, Cognition and Behavior. *Science Compass review*, 298, 1190-1195.

Ellison, N. (2007). Social network sites: Definition, history and scholarship. *Journal of Computer-Mediated Communication, 13 (1)*, 210-230.

Fama, E. F. (1991). Efficient Capital Markets: ll. *The Journal of Finance, 46(5)*, 1575-1617.

Fama, E. F., & French, K. R. (1988). Permanent and Temporary Components of Stock Prices. *The Journal of Political Economy, 96(2)*, 246-273.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The Adjustment of Stock Prices to New information. *International Economic Review*, 10(1), 1-21.

Fang, J., Jacobsen, B., & Qin, Y. (2014). Predictability of the simple technical trading rules: An out-of-sample test. *Review of Financial Economics, 23*, 30-45.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *J. Big Data*, 2(5), 1-14.

Gallagher, L., & Taylor, M. (2002). Permanent and temporary components of stock prices. Evidence from asessing macroeconomic stocks. *Southern Eco J* , 69, 245-262.

Gao, B., & Liu, X. (2020). Intraday sentiment and market returns. *International Review of Economics and Finance, 69*, 48-62.

Godsay, M. (2015). The process of sentiment analysis: A study. *International Journal of Computing Applications*, 126(7), 26-30.

Goncalves, P., Araujo, M., Benevenuto, F., & Cha, M. (2013). Comparing and Combining Sentiment Analysis Methods. *COSN Social and Behavioral Sciences H.3.5*.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37* , 424-438.

Hair, J., Black, W., Babin, B. J., & Anderson, R. (2010). *Multivariate data analysis (7th ed.).* Upper Saddle River, New Jersey: Pearson Educational International.

Han, H., Zhang, Y., Zhang, J., Yang, J., & Zou, X. (2018). Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. *Journal PLoS ONE, 13(8)*.

Hong, H., & Stein, J. C. (2007). Disagreement and the Stock Market. *Journal of Economic Perspectives, 21(2)*, 109-128.

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics, 110(3)*, 712-729.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Proceedings of the ECML*, 137-142.

Jockers, M., & Thalken, R. (2017). *Text Analysis with R.* Springer.

Jong, P. D., Elfayoumy, S., & Schnusenberg, O. (2017). From returns to tweets and back: an investigation of the stock in the dow jones industrial average. *Finance*, 18(1), 55-64.

Khan, S. U., & Rizwan, F. (2008). Trading volume and Stock Returns: Evidence from Pakistan's Stock Market. *International Review of Business Research Papers, 4(2)*, 151-162.

Kordonis, J., Rampatzis, A., & Symeonidis, S. (2016). Stock Price Forecasting via Sentiment Analysis on Twitter. (p. p. 36). Patran, Greece: In Proceeding of the 20th Pan-Hellenic Conference on Informatics; Associations for Computing Machinery .

Liu, L., Wu, J., Li, P., & Li, Q. (2015). A social-media-based approach to predicting stock comovement. *Expert Systems with Applications* , 42(8), 3893-3901.

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance, 66(1)*, 35-65.

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research, 54(4)*, 1187-1230.

Manfield, E. R., & Helms, B. P. (1982). Detecting Multicollinearity. *The American Statistician, 36(3)*, 158-160.

McGurk, Z., Nowak, A., & Hall, J. C. (2019). Stock returns and Investor Sentiment: Textual Analysis and Social Media. *Economically Faculty Working Papers Series, 37*, 1-52.

Meesad, P. (2014). *Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets.* University of Bangkok: Faculty of Information Technology.

Mo, S. Y., Liu, A., & Yang, S. Y. (2016). *News Sentiment to Market Impact and its Feedback Effect.* USA: Steven Institute of Technology.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). *Nrc-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets.* Ottawa, Ontario, Canada: National Research Council Canada.

Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematics and Informatics, 35(8)*, 2173-2183.

Nardo, M., Petracco-Giudici, M., & Naltsidis, M. (2016). Walking down WallStreet with a tablet: A survey of stock market predictions using the web. *Journal of Economic Surveys, 30(2)*, 356-369.

Nguyen, T. H., Shiar, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications, 42(24)*, 9603-9611.

Nofsinger, J. (2005). Social Mood and Financial Economics. *Journal of Behavioral Finance, 6(3)*, 144-160.

Oh, C., & Sheng, O. R. (2011). *Investigating Predictive Power of Stock Microblogging Sentiment in Forecasting Future Stock Price Directional Movement.* Shanghai: ICIS.

Oliveira, N., Cortez, P., & Areal, N. (2016). Stock market sentiment acquistion using microblogging data and statistical. *Decision Support System*, 62-73.

Pathirawasam, C. (2011). The Relationship Between Trading Volume and Stock Returns. *Journal of Competitiveness, 3(3)*, 41-49.

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

Ramachandran, D., & Parvathi, R. (2019). Analysis of Twitter Specific Preprocessing Technique for Tweets. *International Conference on recent trends in advanced computing* (pp. 245-251). Chennai, India: Vellore Instiute of Technology .

Renault, T. (2017). *Intraday online investor sentiment and return pattern in the U.S. stock market.* Paris: IESEG School of Management & Universite Paris 1 Pantheon Sorbonne.

Rinker, T. (2020, 9 4). Package 'sentimentr'. *Retrieved from: https://github.com/trinker/sentimentr*.

Risius, M., Akolk, F., & Beck, R. (2015). Differential Emotions and the Stock Market - The Case of Company-specific Trading. *ECIS Completed Research Papers*, Paper 147.

Ritter, J. R. (2003). Behavioral Finance. *Pacific-Basin Finance Journal, 11(4)*, 429-437.

Rodriguez, F. J. (2016). *Capital asset pricing modle in domain frequency.* Espana: Classification Journal of Economic Literature, Universidad de Cantabria (UNICAN).

Rossi, M. (2016). The capital asset pricing model: A critical literature review. *Global Business and Economics Review, 18(5)*, 604-617.

Sayim, M., Morris, P. D., & Rahman, H. (2013). The effect of US individual investor sentiment on industry-specific stock returns and volatility. *Review of Behavioral Finance (5) 1*, 58-76.

Schumaker, R. P., & Chen, H. (2009). *Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System.* University of Arizona: Artificial Intelligence Lab, Department of Management Information Systems.

Schwert, G. W., & Sequin, P. J. (1990). Heteroskedasticity in Stock Returns. *The Journal of Finance 45(4)*, 1129-1155.

Sebastini, F. (2002). Machine learning in automated text categorization. *ACM Computing surveys (CSUR)*, 34(1), 1-47.

Sharpe, W. (1964). Capital Asset Prices; A Theory of market equilibirum Under Conditions of Risk. *The Journal of Finance, 19*, 425-442.

Shi, R., & Conrad, S. A. (2009). Correlation and regression analysis. *Annals of Allergy, Asthma & Immunology - Statistics for Clinicians, 103(4)*, 34-59.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American Economic Review, 69(4)*, 493-513.

Slovic, P., Fleissner, D., & Bauman, W. S. (1972). Analyzing the Use of Information in Investment Decision making: A Methodological Proposal. *The Journal of Business, 45(2)*, 283-301.

Smailovic, J., Grcar, M., Lavrac, N., & Znidarsic, M. (2013). Predictive Sentiment Analysis of Tweets: A Stock Market Application. *Jozef Stefan Institute*, (pp. 1-12). Ljubljana.

Smith, A. (2021). *The Reddit revolt: GameStop and the impact of social media on institutional investors.* Retrieved on 27-06-2021, from: https://www.thetradenews.com/the-reddit-revolt-gamestop-and-the-impact-of-social-media-on-institutional-investors/.

Sprenger, T. O., & Welpe, I. M. (2010). *Tweets and Trades: The Information Content of Stock Microblogs.* Munchen: Technische Universitat.

Sul, H. K., Dennis, A. R., & Yuan, L. (2016). Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns: Trading on Twitter. *Decisions Sciences*, 1-35.

Sul, H. K., Dennis, A. R., & Yuan, L. I. (2014). Trading on Twitter: The Financial Information of Emotion in Social Media. *47th Hawaii International Conference on System Sciences* (pp. 806-815). Waikoloa, HI, USA: IEEE.

Tan, S. D., & Tas, O. (2021). Social Media Sentiment in International Stock Returns and Trading Activity. *Journal of Behavioral Finance, 22(2)*, 221-234.

Tetlock, P. (2007). Giving content to investor sentiment: the role of media in the stock market. *The Journal of Finance, 62*, 1139-1168.

Tetlock, P. C., Saar-Techansky, M., & Macskassy, S. (2008). More than words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance, 64(3)*, 1437-1466.

Verma, R., Baklaci, H., & Soydemir, G. (2008). The impact of rational and irrational sentiment of individual and instituational investors on DJIA and S&P500 index returns . *Applied Financial Economics*, 1303-1317.

Wang, Z., Ho, S.-B., & Lin, Z. (2018). Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment. *IEEE International Conference on Data Mining Workshops* (pp. 1374-1380). Singapore: Singapore Management University.

Wong, S. S., Rovalinho, J., & Akyildirim, A. (2019). *Twitter Sentiment Analysis.* Retrieved on 11-04-2021, from: https://rstudio-pubs-static.s3.amazonaws.com/557519_5ed17ef6f70a465cb760cb682dd8a226.html.

Wysocki, P. (1998). *Cheap talk on the web: The determinants of postings on stock message boards.* Michigan: Research Support. University of Michigan Business School.

Yobero, C. (2018). *Capital Asset Pricing Model Using Linear Regression in R.* Retrieved on 20-05-2021, from: https://rpubs.com/cyobero/capm.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter ''I hope it is not as bad as I fear'. *Procedia - Social and Behavioral Sciences*, 26, 55-62.

9. Appendices

Appendix A

Table 5: Descriptive statistics of the tweet sample obtained from Kaggle

Sample	Tweet count before duplicates removal	Tweet count after removing duplicates	Sample-ticker count	Mean per trading day (502 trading days)
TSLA	723,771	706,621	708,012	814.58
AAPL	440,133	408,920	420,285	518.44
AMZN	267,143	260,258	266,291	222.83
MSFT	124,005	119,105	121,306	237.26
GOOGL	122,187	111,860	131,804	1441.77

Table 5 represents the descriptive statistics of the tweet sample fetched from Kaggle. It reports the number of tweets before and after removing duplicates, the counted company specific tickers in each sample and the average tweets per trading day.

Table 6: Descriptive statistics of the tweet-sample fetched via the Twitter API Key

Sample	Tweet count before duplicates removal	Tweet count after removing duplicates	Sample-ticker count	Tweet count after word-count error control	Sample ticker count error control
TSLA	740,502	420,774	316,380	258,959	269,876
AAPL	183,069	135,292	90,762	75,589	77,915
AMZN	148,546	100,701	67,628	58,364	59,778
SPY	291,674	206,319	151,422	123,525	130,295
QQQ	122,187	113,913	68,458	59,298	60,404

Table 6 represents the descriptive statistics of the tweet sample fetched using the Twitter API Key. It shows the process to obtain a reliable tweet-sample after removing duplicates and tweets not matched with the specific company ticker.

Table 7: Pre-processing of a raw tweet for each lexicon.

Raw tweet	@JOHN: shall we buy some $MSFT stocks? My advice buy 100 shares. #trading making $ https.//trade//JOHN
Renault	usertag shall we buy some cashtag stocks? my advice buy numbertag shares making $ linktag
Oliveira	shall we buy some tkr stocks my advice buy num shares trading making
R	shall we buy some stocks my advice buy shar trade make
NRC	shall we buy some stocks my advice buy shar trade make
SentiWordNet	shall we buy some stocks my advice buy shar trade make

Table 7 reports the differences in preprocessing a raw tweet before it is reasonable to measure sentiment based on a certain lexicon.

Table 8: Descriptive statistics of each lexicon

Lexicon	N-grams	Purpose	Score-scale	Mean -score	Year of development	Source
Renault	8,000	Measuring sentiment of investor based texts in social media	[-1,-0.2] [0.2, 1]	-0.0245	2017	Renault (2017) - http://www.thomas-renault.com/data.php
Oliveira	20,465	Measuring sentiment of investor based texts in social media	[-10.5, 10.88]	0.2686	2016	Oliveira et al., (2016) - https://github.com/nunomroliveira/stock_market_lexicon
R	11,710	Measuring sentiment in general, focusing on valence shifters	[-2, 1]	-0.2270	2018	Rinker (2018) - https://cran.r-project.org/web/packages/sentimentr/readme/README.html
NRC	5,468	Measuring sentiment in general, focusing on emotions	-1 or 1	-0.1854	2010	R-package - https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
SentiWordNet	20,093	Measuring sentiment in general	[-1, 1]	-0.0617	2019	https://github.com/aesuli/SentiWordNet

Table 8: reports all descriptive statistics to understand the differences per lexicon. It reports the number of n-grams used to classify sentiment, the lexicons purpose/speciality. Furthermore it represents the scale in which the sentiment-scores are assigned to the n-gram. The mean-score indicates the average sentiment-score in the lexicon. Finally, the source and year of development are included in the table.

Table 9: Percentage of tweets retained during the process of assigning sentiment-scores to tweets for each lexicon.

Sample	After duplicates removal	SREN	SOL	SR	NRC	SWN
Apple	408,902	316,743 (77.46%)	406,155 (99.33%)	305,351 (74.68%)	296,246 (72.45%)	356,413 (87.16%)
Amazon	260,258	211,226 (81.16%)	257,985 (99.13%)	186,261 (71.57%)	180,021 (69.17%)	239,864 (92.16%)
Google	111,860	88,755 (79.34%)	109,800 (98.16%)	80,034 (71.55%)	77,433 (69.22%)	99,520 (88.97%)
Microsoft	119,087	93,331 (78.37%)	116,948 (98.20%)	82,759 (69.49%)	79,284 (66.58%)	107,860 (90.57%)
Tesla	706,621	562,437 (79.60%)	704,524 (99.70%)	542,162 (76.73%)	529,040 (74.87%)	666,082 (94.26%)
Total	1,606,728	1,272,492 (79.20%)	1,595,412 (99.30%)	1,196,567 (74.47%)	1,162,024 (72.32%)	1,469,739 (91.47%)

Table 9 reports the numbers of tweets which retains after assigning sentiment-scores to each tweet. Some tweets contain no matching n-grams with a certain sentiment lexicon. Those tweets are threatened as noise and removed automatically during the process. The percentages in the brackets show the retaining rate of tweets after the process of creating sentiment variables.

Appendix B

Table 10: Descriptive statistics multiple regression

TSLA							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R_{it}	499	3.42	-13.90	-1.568	0.11	1.76	17.67
SREN_mv	499	0.071	-0.3012	-0.0711	-0.0231	0.0271	0.1699
SOL_mv	499	0.4978	1.070	2.402	2.702	3.069	4.087
SR_mv	499	0.0131	-0.0179	0.0209	0.03	0.0384	0.074
SNRC_mv	499	0.0496	-0.0212	0.1523	0.1831	0.2168	0.3272
SSWN_mv	499	0.0382	0.0388	0.1227	0.1488	0.1732	0.2934
log_MV	499	0.5671	5.39	6.61	6.90	7.19	8.72
log_TV	499	0.4540	16.33	17.17	17.50	17.75	18.94
MS	499	0.5225	0.4277	0.9076	1.257	1.575	3.84
AAPL							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R_{it}	495	1.7365	-9.96	-0.69	0.11	1.00	7.04
SREN_mv	495	0.0949	-0.3231	0.0059	0.0604	0.1235	0.338
SOL_mv	495	0.6817	1.085	2.972	3.393	3.831	5.412
SR_mv	495	0.0366	-0.0019	0.0724	0.099	0.1211	0.2742
SNRC_mv	495	0.1379	0.0541	0.3051	0.4102	0.4912	1.042
SSWN_mv	495	0.0399	0.0616	0.1456	0.1717	0.1978	0.2810
log_MV	495	0.3993	4.787	6.19	6.41	6.59	8.46
log_TV	495	0.3766	17.63	18.28	18.56	18.80	19.77
MS	495	0.5362	0.4277	0.9076	1.257	1.575	3.84
AMZN							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R_{it}	495	1.909	-7.82	-0.72	0.1032	1.08	9.44
SREN_mv	495	0.0983	-0.4710	-0.0571	0.0027	0.0666	0.3619
SOL_mv	495	0.6787	0.7561	2.370	2.8	3.2	4.989
SR_mv	495	0.0174	0.0023	0.0438	0.0549	0.0661	0.1192
SNRC_mv	495	0.0611	0.0242	0.2118	0.2526	0.2908	0.4818
SSWN_mv	495	0.0476	0.0264	0.1112	0.1447	0.17445	0.3100
log_MV	495	0.4344	4.094	5.814	6.020	6.255	7.587
log_TV	495	0.4364	13.69	14.95	15.28	15.57	16.52
MS	495	0.5368	0.4277	0.9076	1.257	1.575	3.84
MSFT							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R_{it}	499	1.542	-5.43	-0.6140	0.1326	0.9960	7.57
SREN_mv	499	0.0851	-0.1668	0.1014	0.1491	0.1987	0.5393
SOL_mv	499	0.7088	-0.1462	2.325	2.773	3.222	6.551
SR_mv	499	0.0183	0.0066	0.0566	0.0687	0.0810	0.1341
SNRC_mv	499	0.0661	0.0349	0.2470	0.2866	0.3305	0.5017
SSWN_mv	499	0.0557	-0.0389	0.0922	0.1253	0.1537	0.4204
log_MV	499	0.3442	4.477	5.021	5.227	5.38	6.65
log_TV	499	0.3549	16.01	16.83	17.08	17.29	18.53
MS	499	0.5351	0.4277	0.9076	1.257	1.575	3.84
GOOGL							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R_{it}	500	1.641	-7.50	-0.6699	0.0605	0.9777	9.62
SREN_mv	500	0.1864	-1.665	0.0942	0.1899	0.2971	0.6708
SOL_mv	500	0.9693	-2.329	2.203	2.676	3.288	4.691
SR_mv	500	0.0241	-0.0537	0.0441	0.0582	0.0749	0.1440
SNRC_mv	500	0.0881	-0.0801	0.2115	0.2577	0.3153	0.5501
SSWN_mv	500	0.0631	-0.1997	0.0759	0.1163	0.1586	0.2814
log_MV	500	0.5919	3.258	4.585	5.046	5.472	7.203
log_TV	500	0.3886	13.37	14.05	14.31	14.51	15.71
MS	500	0.5349	0.4277	0.9076	1.257	1.575	3.84

Table 10 reports the descriptive statistics of each variable. It reports the number of observations, standard deviations, the mean, the 25% quantile, 75% quantile, the minimum value, and the maximum value.

Table 11: Pearson's correlation matrix multiple regression

TSLA								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN_mv	1.00							
(2) SOL_mv	0.68**	1.00						
(3) SR_mv	0.47**	0.25**	1.00					
(4) SNRC_mv	0.48**	0.31**	0.88**	1.00				
(5) SSWN_mv	0.42**	0.39**	0.44**	0.51**	1.00			
(6) log_TV	-0.33**	-0.37**	-0.17**	-0.22**	-0.21**	1.00		
(7) log_MV	-0.20**	-0.18**	-0.29**	-0.31**	-0.17**	0.66**	1.00	
(8) MS	0.09*	-0.01	0.13**	0.13**	0.01	-0.07	-0.17**	1.00
AAPL								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN_mv	1.00							
(2) SOL_mv	0.72**	1.00						
(3) SR_mv	0.46**	0.71**	1.00					
(4) SNRC_mv	0.41**	0.71**	0.98**	1.00				
(5) SSWN_mv	0.38**	0.49**	0.32**	0.31**	1.00			
(6) log_TV	-0.43**	-0.23**	-0.28**	-0.24**	-0.20**	1.00		
(7) log_MV	-0.30**	-0.14**	-0.30**	-0.27**	-0.12*	0.65**	1.00	
(8) MS	0.28**	0.31**	0.34**	0.34**	0.05	-0.04	-0.04	1.00
AMZN								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN_mv	1.00							
(2) SOL_mv	0.77**	1.00						
(3) SR_mv	0.52**	0.42**	1.00					
(4) SNRC_mv	0.48**	0.48**	0.85**	1.00				
(5) SSWN_mv	0.35**	0.38**	0.39**	0.41**	1.00			
(6) log_TV	-0.31**	-0.36**	-0.24**	-0.31**	-0.07	1.00		
(7) log_MV	-0.09*	-0.14**	-0.19**	-0.28**	-0.15**	0.45**	1.00	
(8) MS	0.24**	0.16**	0.25**	0.19**	0.01	-0.03	0.03	1.00
MSFT								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN_mv	1.00							
(2) SOL_mv	0.57**	1.00						
(3) SR_mv	0.27**	0.17**	1.00					
(4) SNRC_mv	0.27**	0.26**	0.79**	1.00				
(5) SSWN_mv	0.07	0.29**	0.23**	0.31**	1.00			
(6) log_TV	-0.24**	-0.16**	-0.09*	-0.10*	0.12**	1.00		
(7) log_MV	0.01	0.19**	0.03	0.00	0.12**	0.37**	1.00	
(8) MS	0.13**	-0.02	0.00	-0.10*	-0.14**	-0.09	-0.02	1.00
GOOGL								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN_mv	1.00							
(2) SOL_mv	0.66**	1.00						
(3) SR_mv	0.55**	0.43**	1.00					
(4) SNRC_mv	0.44**	0.50**	0.82**	1.00				
(5) SSWN_mv	0.31**	0.53**	0.37**	0.43**	1.00			
(6) log_TV	-0.18**	-0.15**	-0.15**	-0.26**	-0.06	1.00		
(7) log_MV	0.11*	0.43**	0.15**	0.33**	0.29**	0.04	1.00	
(8) MS	0.05	0.02	0.00	-0.06	0.12**	0.02	-0.01	1.00

Table 11 reports the Pearson Correlation matrix for all the variables in each sample during the preliminary phase. ** shows significance correlation at the $p<0.01$ level.

Table 12: Multiple regression including all averaged sentiment variables and control variables

Variables	TSLA	AAPL	AMZN	MSFT	GOOGL
Intercept	-18.58**(-2.762)	-3.821(-1.014)	-5.212(-1.671)	2.735(0.819)	4.866(1.707)
SREN_mv	23.44***(7.936)	6.688***(5.679)	6.015***(4.535)	3.882***(3.957)	-0.024(-0.041)
SOL_mv	-0.8054*(-2.018)	0.6424**(3.310)	0.5639**(2.991)	0.3222**(2.641)	0.3000*(0.0143)
SR_mv	42.28+(1.871)	1.105(0.122)	10.84(1.223)	15.07*(2.573)	9.868(1.664)
SNRC_mv	-9.676(-1.564)	-2.205(-0.909)	-2.833(-1.115)	-3.546*(-2.122)	-3.181+(-1.911)
SSWN_mv	6.730(1.527)	4.8235**(2.786)	0.9505(0.513)	1.377(1.069)	3.027*(2.164)
log_TV	1.357**(3.121)	0.0065(0.029)	0.2149(1.011)	-0.3497+(-1.693)	-0.355+(-1.803)
log_MV	-0.3323(-0.976)	0.309+(1.748)	0.1319(0.645)	0.3568+(1.698)	-0.081(-0.559)
MS	-0.4398(-1.623)	-0.631***(-4.724)	-0.2992*(-1.995)	-0.1249(-1.005)	-0.167(-1.229)
Adjusted R^2	0.1899	0.2937	0.2079	0.1352	0.0531

Table 12: presents the multiple OLS regressions on R_{it} for each stock specific sample. The model includes all sentiment variables and control variables. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Appendix C

Table 13: Simple OLS regression results on daily returns as robustness check

Variables	TSLA	AAPL	AMZN	MSFT	GOOGL	Full sample (mean)
Intercept	0.6139***(4.223)	-0.1814*(-2.6195)	0.1192(1.586)	-0.4986***(-4.73)	-0.3389***(-3.505)	
SREN	0.0136***(10.819)	0.0097***(14.03)	0.0175***(12.26)	0.0209***(7.634)	0.0112***(6.076)	0.0146 (10.14)
Adjusted R^2	0.1890	0.2841	0.2322	0.1031	0.0672	0.175
Variables	TSLA	AAPL	AMZN	MSFT	GOOGL	Full sample (mean)
Intercept	1.429e-01(0.462)	-4.516e-01*(-2.474)	-0.6955***(-3.918)	-0.3756**(-3.04)	-0.2066(1.878)	
SOL	-1.032e-05(-0.118)	2.550e-04***(3.408)	0.00064***(5.103)	0.00089***(4.91)	0.0005**(3.237)	0.00044 (3.226)
Adjusted R^2	-0.0019	0.0210	0.0482	0.0444	0.0186	0.0245
Variables	TSLA	AAPL	AMZN	MSFT	GOOGL	Full sample (mean)
R	-1.2538***(-4.654)	-0.4899*(-2.517)	-0.6881***(-3.569)	-0.3935**(-2.765)	-0.2626*(-2.244)	
SR	0.0421***(6.059)	0.0097*(3.370)	0.0331***(4.561)	0.0383***(4.206)	0.0292***(3.520)	0.0298(4.258)
Adjusted R^2	0.0669	0.0205	0.0385	0.0324	0.0223	0.0347
Variables	TSLA	AAPL	AMZN	MSFT	GOOGL	Full sample (mean)
R	-0.6973*(-2.314)	-0.2921(-1.482)	-0.4898*(-2.355)	-0.3319*(-2.222)	-0.1216(-1.063)	
SNRC	0.0039**(3.105)	0.0016*(2.322)	0.0055**(3.125)	0.0081***(3.497)	0.0036*(2.073)	0.00681(2.731)
Adjusted R^2	0.0170	0.0079	0.0174	0.0221	0.0066	0.013
Variables	TSLA	AAPL	AMZN	MSFT	GOOGL	Full sample (mean)
R	-0.3256(-1.092)	-0.2871(-1.698)	-0.4560**(-2.647)	-0.1514(-1.288)	-0.2110*(-2.006)	
SSWN	0.0026(1.706)	0.0035**(2.658)	0.0087***(3.728)	0.0111**(2.973)	0.0116***(3.565)	0.00704(2.8778)
Adjusted R^2	0.0038	0.0121	0.0254	0.0155	0.0229	0.0152

Table 13 presents the single OLS regressions executed as robustness checks. The sentiment variables measured as daily averaged down to number of messages scores (SREN_mv) are replaced by the absolute aggregate daily sentiment scores (SREN). The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Table 14 : Pearson's correlation matrix robustness check replaced sentiment variables

TSLA								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN	1.00							
(2) SOL	-0.24**	1.00						
(3) SR	0.10*	0.63**	1.00					
(4) SNRC	-0.11*	0.83**	0.91**	1.00				
(5) SSWN	-0.11*	0.86**	0.77**	0.90**	1.00			
(6) log_TV	-0.44**	0.57**	0.50**	0.61**	0.56**	1.00		
(7) log_MV	-0.35**	0.89**	0.64**	0.81**	0.82**	0.66**	1.00	
(8) MS	0.02	-0.09	-0.06	-0.08	-0.08	-0.07	-0.17**	1.00
AAPL								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN	1.00							
(2) SOL	0.02	1.00						
(3) SR	0.30**	0.71**	1.00					
(4) SNRC	0.18**	0.79**	0.97**	1.00				
(5) SSWN	-0.09	0.90**	0.55**	0.64**	1.00			
(6) log_TV	-0.29**	0.51**	0.32**	0.40**	0.48**	1.00		
(7) log_MV	-0.22**	0.83**	0.54**	0.64**	0.80**	0.65**	1.00	
(8) MS	0.20**	0.09	0.25**	0.22**	-0.02	-0.04	-0.04	1.00
AMZN								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN	1.00							
(2) SOL	0.27**	1.00						
(3) SR	0.37**	0.74**	1.00					
(4) SNRC	0.19**	0.84**	0.90**	1.00				
(5) SSWN	0.18**	0.80**	0.70**	0.75**	1.00			
(6) log_TV	-0.33**	0.27**	0.24**	0.31**	0.37**	1.00		
(7) log_MV	-0.13**	0.77**	0.65**	0.76**	0.68**	0.45**	1.00	
(8) MS	0.19**	0.09	0.18**	0.12**	0.03	-0.03	0.03	1.00
MSFT								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN	1.00							
(2) SOL	0.82**	1.00						
(3) SR	0.73**	0.83**	1.00					
(4) SNRC	0.74**	0.87**	0.95**	1.00				
(5) SSWN	0.61**	0.81**	0.78**	0.82**	1.00			
(6) log_TV	0.05	0.22**	0.25**	0.27**	0.31**	1.00		
(7) log_MV	0.65**	0.85**	0.83**	0.86**	0.76**	0.37**	1.00	
(8) MS	0.08	-0.01	0.01	-0.03	-0.07	-0.09	-0.02	1.00
GOOGL								
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) SREN	1.00							
(2) SOL	0.57**	1.00						
(3) SR	0.63**	0.89**	1.00					
(4) SNRC	0.51**	0.93**	0.95**	1.00				
(5) SSWN	0.40**	0.88**	0.80**	0.85**	1.00			
(6) log_TV	-0.16**	0.13**	0.05	0.06	0.14**	1.00		
(7) log_MV	0.48**	0.85**	0.81**	0.85**	0.79**	0.04	1.00	
(8) MS	0.02	-0.03	-0.03	-0.04	0.01	0.02	-0.01	1.00

Table 14 reports the Pearson Correlation matrix for all the variables in each sample for the executed robustness check. ** shows significant correlation at the $p<0.01$ level.

Table 15: Descriptive statistics of the robustness check sample

TSLA							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R	54		-6.44	-2.11	-0.013	1.75	8.60
SREN_mv	54		-0.1753	0.0103	0.0334	0.0702	0.1777
SOL_mv	54		0.701	1.893	2.077	2.319	3.091
SR_mv	54		0.0154	0.0319	0.0409	0.0492	0.0598
SNRC_mv	54		0.0148	0.1288	0.1480	0.1713	0.1972
SSWN_mv	54		0.0507	0.0916	0.1076	0.1211	0.1607
log_MV	54		7.765	8.027	8.268	8.471	8.959
log_TV	54		16.60	16.97	17.16	17.32	17.71
Variable	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
AAPL							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R	50		-3.53	-0.787	-0.013	0.782	2.46
SREN_mv	50		-0.0516	0.0414	0.0739	0.1129	0.1694
SOL_mv	50		1.244	1.726	1.962	2.232	3.066
SR_mv	50		0.0451	0.0664	0.0715	0.0783	0.0919
SNRC_mv	50		0.149	0.19	0.2098	0.2309	0.2675
SSWN_mv	50		0.0145	0.0523	0.0713	0.091	0.1373
log_MV	50		6.681	6.913	7.107	7.228	8.229
log_TV	50		17.80	18.05	18.20	18.36	18.83
AMZN							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R	50		-3.07	-1.04	0.0136	0.9265	2.21
SREN_mv	50		-0.0809	0.0137	0.0588	0.0988	0.1754
SOL_mv	50		1.096	1.74	2.070	2.45	2.91
SR_mv	50		0.0098	0.0504	0.0598	0.0708	0.0826
SNRC_mv	50		0.0535	0.1713	0.1964	0.2230	0.2803
SSWN_mv	50		-0.0664	0.0705	0.0969	0.1288	0.1931
log_MV	50		6.347	6.591	6.829	6.938	8.070
log_TV	50		14.52	14.78	15.04	15.27	15.85
SPY							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R	56		-2.12	-0.2583	0.0814	0.5631	1.535
SREN_mv	56		-0.2520	-0.1556	-0.1025	-0.084	-0.0256
SOL_mv	56		1.017	1.467	1.666	1.881	2.199
SR_mv	56		0.0067	0.0163	0.0221	0.0265	0.0364
SNRC_mv	56		0.0396	0.0774	0.0893	0.1025	0.1229
SSWN_mv	56		0.0401	0.0662	0.0821	0.0956	0.1232
log_MV	56		17.56	17.77	18.00	18.18	18.72
log_TV	56		7.261	7.380	7.552	7.66	8.13
QQQ							
Variables	Obs	Std.Dev.	Min	25th QT	Mean	75th QT	Max
R	54		-2.588	-0.548	0.0851	0.768	2.207
SREN_mv	54		-0.2363	-0.1603	-0.1047	-0.0564	0.044
SOL_mv	54		0.8087	1.353	1.738	2.016	2.534
SR_mv	54		0.0076	0.0303	0.0372	0.0437	0.067
SNRC_mv	54		0.0685	0.1214	0.1338	0.1491	0.1952
SSWN_mv	54		-0.019	0.0389	0.0569	0.0756	0.1418
log_MV	54		6.532	6.721	6.853	6.93	7.39
log_TV	54		16.87	17.25	17.45	17.65	18.33

Table 15: represents the descriptive statistics of each variable. It reports the number of observations, standard deviations, the mean, the 25% quantile, 75% quantile, the minimum value, and the maximum value.

Table 16: Single OLS regression results robustness check samples

Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-0.8446*(-2.236)	-0.8684**(-3.411)	-0.4673(-1.996)	1.2302***(6.507)	0.8710***(4.212)	
SREN_mv	24.87***(4.978)	11.56***(4.22)	8.18**(2.982)	9.787***(6.59)	7.51***(4.6)	12,38(4,674)
Adjusted R^2	0.3097	0.2555	0.1387	0.44	0.2756	0,2834
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-7.0198***(-3.757)	-2.328**(-2.766)	-3.079***(-4.102)	-3.092***(-7.915)	-2.784***(-6.915)	
SOL_mv	3.373***(3.823)	1.18**(2.807)	1.495***(4.215)	1.913***(8.239)	1.651***(7.342)	1,922(5,284)
Adjusted R^2	0.2044	0.1231	0.2549	0.5533	0.4995	0,327
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-4.585**(-3.094)	-0.6356(-0.502)	-0.121(-0.150)	-0.2655(-0.857)	-0.9947*(-2.246)	
SR_mv	112.4**(3.190)	8.713(0.497)	2.251(0.172)	15.64(1.179)	29.03*(2.551)	33,61(1,518)
Adjusted R^2	0.1476	-0.0156	-0.0202	0.0072	0.094	0,033
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-4.434*(-2.327)	-0.957(-0.749)	-0.2142(-0.231)	0.2312(0.439)	-1.832**(-2.786)	
SNRC_mv	29.87*(2.37)	4.502(0.459)	1.16(0.25)	-1.678(-0.290)	14.329**(2.972)	10,71(1,152)
Adjusted R^2	0.0802	-0.00911	-0.0195	-0.0173	0.1287	0,0325
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-5.309**(-2.917)	-0.666(-1.412)	-0.5427(-1.279)	-0.6903(-1.767)	-0.989***(-3.933)	
SSWN_mv	49.22**(2.975)	9.16(1.492)	5.736(1.450)	9.351*(2.037)	18.88***(4.802)	18,47(3,507)
Adjusted R^2	0.129	0.0244	0.0219	0.0551	0.2939	0,1049

Table 16 reports the results of the single OLS regressions executed as robustness check in the preliminary phase. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Table 17: Pearson's correlation matrix for the robustness check different samples

TSLA							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN_mv	1.00						
(2) SOL_mv	0.88**	1.00					
(3) SR_mv	0.75**	0.58**	1.00				
(4) SNRC_mv	0.59**	0.41**	0.89**	1.00			
(5) SSWN_mv	0.53**	0.62**	0.44**	0.44**	1.00		
(6) log_TV	-0.16	0.05	-0.25	-0.33*	0.02	1.00	
(7) log_MV	-0.26	-0.05	-0.37**	-0.27	0.18	0.73**	1.00
AAPL							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN_mv	1.00						
(2) SOL_mv	0.87**	1.00					
(3) SR_mv	0.14	-0.07	1.00				
(4) SNRC_mv	0.17	-0.01	0.84**	1.00			
(5) SSWN_mv	0.63**	0.69**	-0.02	-0.03	1.00		
(6) log_TV	-0.17	0.02	-0.03	-0.10	0.10	1.00	
(7) log_MV	0.51**	0.63**	-0.16	-0.05	0.45**	0.46**	1.00
AMZN							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN_mv	1.00						
(2) SOL_mv	0.77**	1.00					
(3) SR_mv	0.22	0.03	1.00				
(4) SNRC_mv	0.25	0.05	0.90**	1.00			
(5) SSWN_mv	0.46**	0.56**	0.25	0.16	1.00		
(6) log_TV	-0.26	0.00	-0.29*	-0.38**	0.03	1.00	
(7) log_MV	0.19	0.43**	-0.29*	-0.34*	0.21	0.65**	1.00
SPY							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN_mv	1.00						
(2) SOL_mv	0.84**	1.00					
(3) SR_mv	0.31*	0.27*	1.00				
(4) SNRC_mv	-0.04	-0.06	0.69**	1.00			
(5) SSWN_mv	0.21	0.28*	0.25	0.20	1.00		
(6) log_TV	-0.60**	-0.53**	0.00	0.15	-0.17	1.00	
(7) log_MV	-0.55**	-0.43**	0.06	0.15	-0.15	0.81**	1.00
QQQ							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN_mv	1.00						
(2) SOL_mv	0.80**	1.00					
(3) SR_mv	0.50**	0.53**	1.00				
(4) SNRC_mv	0.49**	0.58**	0.83**	1.00			
(5) SSWN_mv	0.38**	0.58**	0.27*	0.45**	1.00		
(6) log_TV	-0.53**	-0.42**	-0.23	-0.28*	-0.13	1.00	
(7) log_MV	-0.58**	-0.31*	-0.15	-0.19	-0.09	0.80**	1.00

Table 17 presents Pearson's Correlation matrix for the robustness check sample. ** shows significant correlation at the p<0.01 level.

Table 18: Simple OLS regression results for the robustness check sample with replaced sentiment variables

Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-0.5667(-1.581)	-0.2911(-1.233)	-0.3581(-1.588)	0.8562***(6.072)	0.7652***(4.264)	
SREN	0.0048***(4.910)	0.0025+(1.785)	0.0059*(2.573)	0.0032***(6.451)	0.0062***(4.912)	0,0045(4,126)
Adjusted R^2	0.3037	0.0427	0.1029	0.4293	0.3038	0,2365
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
Intercept	-2.172(-1.998)	-1.188e-01(-0.369)	-0.5059(-1.514)	-1.335**(-2.825)	-2.058***(-4.470)	
SOL	0.00025*(2.132)	3.964e-05(0.397)	0.00025+(1.840)	0.00044**(3.055)	0.0013***(4.807)	0,00046(2,4462)
Adjusted R^2	0.0627	-0.0175	0.0464	0.1336	0.2944	0,1039
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
R	-2.836*(-2.565)	0.1238(0.251)	-0.2948(-0.568)	0.1634(0.602)	-0.5346(-1.23)	
SR	0.0175**(2.724)	-0.0015(-0.298)	0.0054(0.636)	-0.0018(-0.324)	0.017(1.50)	0,00732(0,8476)
Adjusted R^2	0.108	-0.0189	-0.0123	-0.0168	0.0231	0,0166
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
R	-1.675(-1.43)	0.0750(0.163)	-0.3515(-0.616)	0.7868*(2.390)	-0.4149(-0.743)	
SNRC	0.0028(1.51)	-0.0003(-0.208)	0.0019(0.676)	-0.004*(-2.236)	0.0039(0.924)	0,00086(0,1332)
Adjusted R^2	0.0236	-0.0199	-0.111	0.0689	-0.0027	0,0117
Variables	TSLA	AAPL	AMZN	SPY	QQQ	Full sample (mean)
R	-1.423(-1.40)	-1.188e-01(-0.369)	-0.3649(-1.2)	0.0156(0.044)	-0.8427**(-3.268)	
SSWN	0.0031(1.51)	3.964e-05(0.397)	0.0038(1.546)	0.0004(0.19)	0.0171***(4.068)	0,0049(1,542)
Adjusted R^2	0.0235	-0.0174	0.0275	-0.0182	0.2268	0,0484

Table 18 reports the results of the single OLS regressions executed as robustness check in the preliminary phase. This robustness check contains different samples and the replaced sentiment variables. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Table 19: Pearson's correlation matrix – robustness check altered sentiment variables on the robustness sample

TSLA							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN	1.00						
(2) SOL	0.49**	1.00					
(3) SR	0.58**	0.81**	1.00				
(4) SNRC	0.32*	0.78**	0.92**	1.00			
(5) SSWN	0.26	0.91**	0.76**	0.83**	1.00		
(6) log_TV	-0.07	0.57**	0.40**	0.39**	0.51**	1.00	
(7) log_MV	-0.11	0.75**	0.57**	0.70**	0.80**	0.73**	1.00
AAPL							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN	1.00						
(2) SOL	0.94**	1.00					
(3) SR	0.83**	0.87**	1.00				
(4) SNRC	0.86**	0.90**	0.98**	1.00			
(5) SSWN	0.89**	0.94**	0.84**	0.85**	1.00		
(6) log_TV	0.16	0.36*	0.43**	0.40**	0.38**	1.00	
(7) log_MV	0.84**	0.92**	0.91**	0.92**	0.87**	0.46**	1.00
AMZN							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN	1.00						
(2) SOL	0.78**	1.00					
(3) SR	0.59**	0.66**	1.00				
(4) SNRC	0.60**	0.68**	0.96**	1.00			
(5) SSWN	0.73**	0.90**	0.65**	0.64**	1.00		
(6) log_TV	0.08	0.49**	0.39**	0.41**	0.46**	1.00	
(7) log_MV	0.61**	0.90**	0.71**	0.75**	0.79**	0.65**	1.00
SPY							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN	1.00						
(2) SOL	-0.03	1.00					
(3) SR	-0.22	0.54**	1.00				
(4) SNRC	-0.59**	0.44**	0.79**	1.00			
(5) SSWN	-0.28*	0.47**	0.53**	0.56**	1.00		
(6) log_TV	-0.75**	0.30*	0.40**	0.65**	0.36**	1.00	
(7) log_MV	-0.77**	0.58**	0.56**	0.78**	0.52**	0.81**	1.00
QQQ							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) SREN	1.00						
(2) SOL	0.15	1.00					
(3) SR	-0.05	0.59**	1.00				
(4) SNRC	-0.26	0.64**	0.85**	1.00			
(5) SSWN	0.12	0.62**	0.33*	0.46**	1.00		
(6) log_TV	-0.70**	0.14	0.22	0.37**	0.11	1.00	
(7) log_MV	-0.79**	0.40**	0.42**	0.62**	0.21	0.80**	1.00

Table 19 reports Pearson's correlation matrix for the robustness check sample in combination with the altered measurement of sentiment variables. ** shows significant correlation at the p<0.01 level.

Table 20: Robustness check results – relatedness test company specific sentiment

AAPL	Stock specific sentiment variable related to:			
Variables	TSLA	AMZN	MSFT	GOOGL
Intercept	0.1527(1.853)	0.1194(1.573)	0.019(0.151)	-0.0198(-0.187)
SREN	0.0011(1.512)	0.0078***(5.444)	0.0031(0.945)	0.0037(1.840)
Adjusted R^2	0.0026	0.0548	-0.0002	0.0048
TSLA	Stock specific sentiment variable related to:			
Variables	AMZN	AAPL	MSFT	GOOGL
Intercept	0.1131(0.734)	0.1628(1.008)	-0.1384(-0.562)	0.1895(0.908)
SREN	-0.0012(-0.397)	-0.0016(-0.997)	0.0008(1.292)	-0.0022(-0.551)
Adjusted R^2	-0.0017	0	0.00013	-0.0014
AMZN	Stock specific sentiment variable related to:			
Variables	TSLA	AAPL	MSFT	GOOGL
Intercept	0.1328(1.463)	-0.0237(-0.27)	-0.0791(-0.573)	0.081(0.694)
SREN	0.0007(1.005)	0.0042***(4.771)	0.006+(1.683)	0.0006(0.279)
Adjusted R^2	0	0.0423	0.0037	-0.0019
MSFT	Stock specific sentiment variable related to:			
Variable	TSLA	AAPL	AMZN	GOOGL
Intercept	0.1618*(2.221)	0.078(1.077)	0.139*(2.026)	0.0185(0.197)
SREN	0.0008(1.253)	0.0018*(2.521)	0.0055***(4.263)	0.0032+(1.788)
Adjusted R^2	0.0011	0.011	0.0336	0.0044
GOOGL	Stock specific sentiment variable related to:			
Variables	TSLA	AAPL	MSFT	AMZN
Intercept	0.1362+(1.770)	-0.021(-0.270)	-0.1174(-0.994)	0.0676(0.926)
SREN	0.0019**(2.983)	0.0027***(3.603)	0.0059+(1.945)	0.0053***(3.805)
Adjusted R^2	0.0156	0.0237	0.0056	0.0266

Table 20 reports the results of the robustness check to test relatedness of company specific sentiment. For example, all other company specific SREN variables are used as predictor in the AAPL sample. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Table 21: Robustness check results – relatedness test company specific sentiment – SOL_mv

AAPL	Stock specific sentiment variable related to:			
Variables	TSLA	AMZN	MSFT	GOOGL
Intercept	-0.4927(-1.148)	-1.793***(-5.604)	-0.7208*(-2.26)	-0.6726**(-2.929)
SOL_mv	0.2238(1.433)	0.6805***(6.127)	0.3003**(2.693)	0.2933***(3.629)
Adjusted R^2	0.0021	0.0689	0.0125	0.024
TSLA	**Stock specific sentiment variable related to:**			
Variables	AAPL	AMZN	MSFT	GOOGL
Intercept	1.952*(2.508)	-0.6188(-0.946)	-0.6318(1.022)	-0.2619(-0.574)
SOL_mv	-0.5418*(-2.408)	0.2618(1.153)	0.268(1.241)	0.1391(0.868)
Adjusted R^2	0.0096	0	0.0011	0
AMZN	**Stock specific sentiment variable related to:**			
Variables	TSLA	AAPL	MSFT	GOOGL
Intercept	-0.7029(-1.49)	-2.119***(-4.993)	-0.4638(-1.316)	-0.5136*(-2.021)
SOL_mv	0.2982+(1.738)	0.6556***(5.341)	0.2044+(1.659)	0.231*(2.576)
Adjusted R^2	0.004	0.053	0.004	0.011
MSFT	**Stock specific sentiment variable related to:**			
Variables	TSLA	AAPL	AMZN	GOOGL
Intercept	-0.2556(-0.67)	-1.0302**(-2.946)	-0.8466**(-2.896)	-0.266(-1.299)
SOL_mv	0.1437(1.035)	0.343***(3.394)	0.35***(3.449)	0.1488*(2.065)
Adjusted R^2	0	0.021	0.022	0.006
GOOGL	**Stock specific sentiment variable related to:**			
Variables	TSLA	AAPL	MSFT	AMZN
Intercept	-0.9147*(-2.264)	-1.624***(-4.405)	-0.6497*(-2.198)	-1.298***(-4.206)
SOL_mv	0.3618*(2.460)	0.4971***(4.667)	0.257*(2.488	0.486***(4.537)
Adjusted R^2	0.01	0.0404	0.01	0.038

Table 21 reports the results of the robustness check to test relatedness of company specific sentiment. For example, all other company specific SOL_mv variables are used as predictor in the AAPL sample. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Appendix D

Table 22: Descriptive statistic augmented CAPM model

TSLA							
Variables	Obs	Std.Dev.	Min	25^{th} QT	Mean	75^{th} QT	Max
RC_{it}	493	0.0337	-0.1496	-0.0156	0.0008	0.0174	0.1627
Rf	493	0.0003	0.00236	0.00275	0.00298	0.00328	0.00336
(Rm-Rf)	493	0.0095	-0.0452	-0.0061	-0.0026	0.0027	0.0459
SREN	493	109.41	-625.89	-61.61	-36.77	23.44	270.44
SOL_mv	493	0.4972	1.070	2.407	2.706	3.069	4.087
Log_TV	493	0.4513	16.55	17.18	17.50	17.75	18.94
Log_MV	493	0.5689	5.389	6.607	6.901	7.191	8.720
MS	493	0.5349	0.4277	0.9076	1.251	1.557	3.84
AAPL							
Variables	Obs	Std.Dev.	Min	25^{th} QT	Mean	75^{th} QT	Max
RC_{it}	489	0.0174	-0.1049	-0.0065	0.0011	0.0102	0.0681
Rf	489	0.0003	0.00236	0.00275	0.00298	0.00328	0.00336
(Rm-Rf)	489	0.0095	-0.0452	-0.0063	-0.0026	0.0027	0.0459
SREN	489	95.24	-1298.82	3.153	30.387	73.46	287.62
SOL_mv	489	0.6812	1.085	2.973	3.395	3.831	5.412
Log_TV	489	0.3704	17.73	18.29	18.57	18.80	19.77
Log_MV	489	0.3966	4.787	6.194	6.417	6.600	8.464
MS	489	0.5350	0.4277	0.9076	1.251	1.557	3.84
AMZN							
Variables	Obs	Std.Dev.	Min	25^{th} QT	Mean	75^{th} QT	Max
RC_{it}	489	0.0192	-0.0814	-0.0067	0.00097	0.011	0.0902
Rf	489	0.0003	0.00236	0.00275	0.00298	0.00328	0.00336
(Rm-Rf)	489	0.0095	-0.0452	-0.0063	-0.0026	0.0027	0.0459
SREN	489	52.98	-237.03	-24.35	-0.7484	26.33	292.01
SOL_mv	489	0.6724	0.7561	2.376	2.799	3.199	4.989
Log_TV	489	0.4271	14.45	14.95	15.29	15.57	16.52
Log_MV	489	0.4342	4.094	5.823	6.025	6.263	7.587
MS	489	0.5358	0.4277	0.9076	1.251	1.557	3.84
MSFT							
Variables	Obs	Std.Dev.	Min	25^{th} QT	Mean	75^{th} QT	Max
RC_{it}	493	0.0154	-0.0558	-0.0059	0.0013	0.0099	0.0729
Rf	493	0.0003	0.00236	0.00275	0.00298	0.00328	0.00336
(Rm-Rf)	493	0.0095	-0.0452	-0.0063	-0.0026	0.0027	0.0459
SREN	493	23.82	-25.35	17.42	30.30	37.91	172.07
SOL_mv	493	0.7041	-0.1462	2.325	2.772	3.222	6.551
Log_TV	493	0.3477	4.477	5.021	5.227	5.38	6.65
Log_MV	493	0.3424	16.01	16.83	17.08	17.29	18.53
MS	493	0.5339	0.4277	0.9076	1.257	1.575	3.84
GOOGL							
Variables	Obs	Std.Dev.	Min	25^{th} QT	Mean	75^{th} QT	Max
RC_{it}	494	0.0165	-0.0779	-0.0067	0.00056	0.0098	0.0918
Rf	494	0.0003	0.00236	0.00275	0.00298	0.00328	0.00336
(Rm-Rf)	494	0.0095	-0.0452	-0.0063	-0.0026	0.0027	0.0459
SREN	494	38.43	-226.44	10.62	35.57	54.36	210.60
SOL_mv	494	0.9644	-2.329	2.2	2.675	3.288	4.691
log_TV	494	0.3827	3.258	4.585	5.046	5.472	7.203
log_MV	494	0.5895	13.37	14.05	14.31	14.51	15.71
MS	494	0.5338	0.4277	0.9076	1.257	1.575	3.84

Table 22: represents the descriptive statistics of each variable inserted in the augmented CAPM model.It reports the number of observations, standard deviations, the mean, the 25% quantile, 75% quantile, the minimum value, and the maximum value.

Table 23 : Pearson's correlation matrix augmented CAPM model

TSLA							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) RC_{it}	1.00						
(2) β(Rm-Rf)	0.36**	1.00					
(3) SREN	0.45**	0.11*	1.00				
(4) SOL_mv	0.18**	0.12**	0.49**	1.00			
(5) log_TV	-0.01	-0.03	-0.45**	-0.38**	1.00		
(6) log_MV	-0.04	-0.02	-0.35**	-0.18**	0.65**	1.00	
(7) MS	-0.03	-0.04	0.01	-0.02	-0.08	-0.18**	1.00
AAPL							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) RC_{it}	1.00						
(2) β(Rm-Rf)	0.74**	1.00					
(3) SREN	0.54**	0.32**	1.00				
(4) SOL_mv	0.41**	0.33**	0.56**	1.00			
(5) log_TV	-0.16**	-0.21**	-0.30**	-0.22**	1.00		
(6) log_MV	-0.09	-0.11*	-0.22**	-0.14**	0.64**	1.00	
(7) MS	-0.07	-0.03	0.20**	0.31**	-0.05	-0.04	1.00
AMZN							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) RC_{it}	1.00						
(2) β(Rm-Rf)	0.75**	1.00					
(3) SREN	0.49**	0.29**	1.00				
(4) SOL_mv	0.42**	0.35**	0.70**	1.00			
(5) log_TV	-0.12**	-0.17**	-0.33**	-0.36**	1.00		
(6) log_MV	-0.01	-0.01	-0.13**	-0.14**	0.44**	1.00	
(7) MS	0.03	-0.04	0.18**	0.15**	-0.03	0.03	1.00
MSFT							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) RC_{it}	1.00						
(2) β(Rm-Rf)	0.85**	1.00					
(3) SREN	0.31**	0.18**	1.00				
(4) SOL_mv	0.30**	0.23**	0.54**	1.00			
(5) log_TV	-0.13**	-0.22**	0.05	-0.14**	1.00		
(6) log_MV	0.08	-0.02	0.65**	0.20**	0.36**	1.00	
(7) MS	-0.01	-0.03	0.07	-0.03	-0.09	-0.03	1.00
GOOGL							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) RC_{it}	1.00						
(2) β(Rm-Rf)	0.77**	1.00					
(3) SREN	0.27**	0.12**	1.00				
(4) SOL_mv	0.22**	0.16**	0.63**	1.00			
(5) log_TV	-0.12**	-0.16**	-0.16**	-0.15**	1.00		
(6) log_MV	0.03	0.03	0.48**	0.43**	0.02	1.00	
(7) MS	-0.03	-0.04	0.01	0.01	0.02	-0.02	1.00

Table 23 reports the Pearson Correlation matrix for all the variables inserted in the augmented CAPM model.** shows significance at the 0.01 level.

Table 24: Variance Inflation Factor (VIF) values for the augmented CAPM model

Variables	ViF TSLA	ViF AAPL	ViF AMZN	ViF MSFT	ViF GOOGL
β(Rm-Rf)	1.021	1.199	1.170	1.119	1.052
SREN	1.504	1.565	2.015	2.627	1.854
SOL_mv	1.409	1.626	2.114	1.546	1.764
log_MV	1.853	1.702	1.255	2.197	1.366
log_TV	2.054	1.805	1.428	1.282	1.071
MS	1.040	1.133	1.052	1.037	1.003
Mean ViF	1.480	1.505	1.506	1.635	1.352

Table 24 reports the VIF-factors for all explanatory variables in the augmented CAPM model.

Table 25: Analysis of Variance (ANOVA) test results

	AAPL	TSLA	AMZN	MSFT	GOOGL
	RSS	RSS	RSS	RSS	RSS
Model 1 (Simple CAPM)	0.0665	0.4856	0.0788	0.0331	0.0553
Model 2 (Augmented)	0.0476	0.3662	0.0629	0.0293	0.05
Sum of Sq < Pr(>Chi)	0.0188< 0***	0.1194 < 0***	0.0159 <0***	0.0038 <0***	0.0053 < 0***

Table 25 reports the results of the ANOVA test comparing the simple CAPM model (Model1) with the augmented CAPM model (Model 2). The Chi-test is significant at ***p < (0.001) for each sample.

Table 26 : Sample splitting accuracy test

	AAPL		TSLA		AMZN		MSFT	
Metric	Simple	Augmented	Simple	Augmented	Simple	Augmented	Simple	Augmented
ME	-0.002	-0.0022	-0.0046	0.0096	0.0026	0.0017	0.0013	0.0014
RMSE	0.0085	0.0081	0.0282	0.0284	0.0088	0.0089	0.0061	0.0054
MAE	0.0065	0.0063	0.0182	0.0183	0.0067	0.0070	0.0046	0.0042
MPE	194.30	760.83	-2013.75	-207.89	89.48	25.12	33.83	-26.21
MAPE	644.94	1375.79	2914.57	404.45	241.70	148.68	212.94	189.23

	GOOGL	
Metric	Simple	Augmented
ME	0.0005	0.0026
RMSE	0.0059	0.0177
MAE	0.0046	0.0134
MPE	116.03	103.52
MAPE	465.65	527.05

Table 26 reports the sample-splitting accuracy metrics of both the simple and augmented CAPM models.

Table 27 : Out of sample accuracy test AAPL

	AAPL	
Metric	Simple	Augmented
ME	0.0035	0.0066
RMSE	0.0125	0.0137
MAE	0.0087	0.0101
MPE	-92.92	199.20
MAPE	713.54	377.60

Table 27 reports the accuracy metrics of the out of sample test.

Table 28: Results additional test – replacing SREN with SREN_mv

Variables	TSLA	AAPL	AMZN	MSFT	GOOGL
α	-0.1279*(-2.018)	-0.078**(-2.701)	-0.0556*(-2.537)	-0.0647***(-3.487)	-0.0112(-0.606)
β(Rm-Rf)	1.086***(7.814)	1.1623***(20.59)	1.368***(22.11)	1.337***(33.97)	1.308***(25.57)
SREN_mv	0.2274***(8.943)	0.0473***(5.673)	0.0447***(5.034)	0.0194***(3.656)	0.0050(1.440)
SOL_mv	-0.0091*(-2.485)	0.0021*(2.007)	0.0009(0.746)	0.00009(1.489)	0.0012(1.611)
log_MV	-0.0041(-1.342)	0.0007(0.409)	-0.0011(0.451)	0.0031**(2.65)	-0.0008(-0.893)
log_TV	0.0111**(2.734)	0.0040*(2.174)	0.0042**(2.757)	0.0028*(2.415)	0.0011(0.847)
MS	-0.0039(-1.529)	-0.0045***(-4.615)	-0.0002(-0.173)	0.0003(0.458)	-0.0003(-0.313)
Adjusted R-squared	0.2689	0.6165	0.6088	0.7391	0.5934

Table 28 reports the results of the augmented CAPM model. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Table 29: Results additional test - Augmented CAPM model – SREN only

Variables	TSLA	AAPL	AMZN	MSFT	GOOGL
α	-0.2957***(-5.186)	-0.0573*(-2.232)	-0.0684***(-3.466)	-0.0238+(-1.949)	-0.0112(-0.628)
β(Rm-Rf)	1.119***(8.531)	1.154***(22.71)	1.355***(23.571)	1.333***(34.30)	1.30***(26.527)
SREN	0.0002***(12.50)	0.00007***(13.49)	0.0002***(10.84)	0.0002***(6.341)	0.0009***(7.192)
log_MV	-0.0004(-0.133)	0.001(0.687)	-0.0007(-0.534)	-0.0024(-1.623)	-0.0029**(-3.262)
log_TV	0.0176***(4.634)	0.003+(1.898)	0.0051***(3.553)	0.0028*(2.483)	0.0018(1.503)
MS	-0.0006(-0.228)	-0.004***(-4.571)	-0.0003(-0.264)	0.0001(0.187)	-0.0003(-0.36)
Adjusted R-squared	0.3379	0.6716	0.6457	0.7442	0.6226

Table 29 reports the results of the augmented CAPM model with only SREN included as sentiment variable. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Table 30: Results additional test- Augmented CAPM model – SOL_mv only

Variables	TSLA	AAPL	AMZN	MSFT	GOOGL
α	-0.149*(-2.188)	-0.0334(-1.162)	-0.0628**(-2.803)	-0.056**(-2.993)	-0.0087(-0.474)
β(Rm-Rf)	1.221***(8.198)	1.219***(21.28)	1.399***(22.18)	1.353***(34.15)	1.306***(25.51)
SOL_mv	0.0116***(3.767)	0.0061***(7.192)	0.0056***(5.902)	0.0022***(4.115)	0.0019***(3.454)
log_MV	-0.0051(-1.534)	0.00002(0.013)	-0.0007(-0.516)	0.0027*(2.361)	-0.0011(-1.271)
log_TV	0.0091*(2.067)	0.0012(0.648)	0.0036*(2.33)	0.0023*(1.971)	0.0009(0.743)
MS	-0.0012(-0.451)	-0.0038***(-3.839)	0.00085(0.806)	0.0007(1.074)	-0.0002(-0.258)
Adjusted R-squared	0.1504	0.5917	0.5891	0.7324	0.5925

Table 29 reports the results of the augmented CAPM model with only SOL_mv included as sentiment variable. The coefficients are reported for each variable, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.

Appendix E

Table 31: Granger Causality results

Sample	Relationship stock returns Granger causes sentiment	Pr > F	Relationship sentiment Granger causes stock returns	Pr > F
TSLA				
	Stock Returns = SREN_mv	0.491	SREN_mv = Stock Returns	0.0889
	Stock Returns = SREN	0.725	SREN = Stock Returns	0.0246*
	Stock Returns = SOL_mv	0.733	SOL_mv = Stock Returns	0.0335*
AAPL				
	Stock Returns = SREN_mv	0.0015**	SREN_mv = Stock Returns	0.0002***
	Stock Returns = SREN	0.0437*	SREN = Stock Returns	0.0002***
	Stock Returns = SOL_mv	0.0094**	SOL_mv = Stock Returns	0.0006***
AMZN				
	Stock Returns = SREN_mv	0.4635	SREN_mv = Stock Returns	0***
	Stock Returns = SREN	0.0796	SREN = Stock Returns	0***
	Stock Returns = SOL_mv	0.0099**	SOL_mv = Stock Returns	0.0042**
MSFT				
	Stock Returns = SREN_mv	0.2428	SREN_mv = Stock Returns	0.1508
	Stock Returns = SREN	0.0797	SREN = Stock Returns	0.0289*
	Stock Returns = SOL_mv	0.1416	SOL_mv = Stock Returns	0.1214
GOOGL				
	Stock Returns = SREN_mv	0.3333	SREN_mv = Stock Returns	0.138
	Stock Returns = SREN	0.6866	SREN = Stock Returns	0.0894
	Stock Returns = SOL_mv	0.4955	SOL_mv = Stock Returns	0.2251

Table 31 reports the results of the two directional Granger Causality analysis. The coefficients are reported for each Granger test, followed by the significance level (*** p<0.001, ** p<0.01, * p<0.05, + p<0.1) with corresponding t-statistic between the brackets.