Social Media Sentiment and Stock Return

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Master of Science in Finance

In the Department of Finance and Management Science

Edwards School of Business

University of Saskatchewan

Saskatoon, Saskatchewan, Canada

By

AMIR HOSSEIN SHAMSHIRI

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for Master of Science in Finance degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Finance and Management Science
Edwards School of Business
University of Saskatchewan
25 Campus Drive
Saskatoon, Saskatchewan S7N 5A7 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

ABSTRACT

Technology advancements increase people's engagements in social media and make a great framework for analyzing investors' behavior. Social media sentiment is the overall mood of investors shared on social media platforms and can affect readers' investment decisions and the stock return. During the literature review, we found inconsistent empirical results for the relationship between social media sentiment and stock return. In addition, this relationship's moderators must be theoretically and empirically analyzed. Using a panel regression over investors' sentiment extracted from the two biggest financial and non-financial social media platforms (StockTwits and Twitter), we have checked if social media sentiment can predict S&P500 stock return and if the sentiment effect varies among different firm characteristics and market conditions. The findings suggest that positive (negative) sentiment has a significantly positive (negative) effect on stock return, while negative sentiment has a stronger effect. In addition, the sentiment effect is stronger among the firms that are smaller, high volatile, and has higher attention on social media platforms, specifically during bearish markets. Our findings indicate that social media sentiment is a risk factor separated from other previously studied risk factors.

Keywords: Social Media Sentiment, Stock Return, Firm Characteristics, Market Condition

ACKNOWLEDGEMENTS

To start with, I would like to express my greatest appreciation to my supervisors, Dr. Shan Wang and Dr. Fan Yang. I am truly thankful for their valuable time and guidance on my process. This work would not be finished without their patience, knowledge, and supports.

I would also like to thank my committee member Dr. Craig Wilson and my external examiner Dr. Hamilton Elkins for their helpful comments. In addition, I would like to thank Dr. Marie Racine, the Graduate Chair in Finance, for her support and help during the past two years.

In the end, I want to express my thankfulness to my family, Fatemeh Moniri, Mansour Shamshiri, and Shiva Shamshiri, and also my friends for their unconditional love and support, without whom I could not be where I am today.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

Under the behavioral finance framework, investment decisions are based on not only the fundamental facts but also the sentiment of investors about a firm. The investors' sentiment is how much people are optimistic or pessimistic about an individual stock or the whole market that can not be justified by the facts of the firm or the market (Baker & Wurgler, 2006). Investors' sentiment may influence their investment decisions and affect stock prices (DeLong et al., 1990; Ho & Hung, 2009). Previous studies measure investors' sentiment using survey results and mass media content such as newspapers. However, technology advancements increase people's engagement with social media platforms and web forums, in which they can freely share their sentiment toward any firm and stock. Social media content brings more detailed information about the investors' sentiment compared with the previously mentioned sources (Das & Chen, 2007). Thus, investors' sentiment can be captured from the shared content on social media (Sprenger et al., 2014), and it can be used for explaining changes in stock returns.

Social media content shows the publisher's sentiment and may affect the readers' sentiment at the same time (Sprenger et al., 2014; Tan & Tas, 2021). In general, people process information by two routes: central and peripheral routes (Petty & Cacioppo, 1984). Under the central route, they interpret the quantitative facts – the fundamental information such as changes in a firm's cash flow or its EPS that might be published in the firms' financial reports or mass media. However, people consider the sentimental and qualitative contents using the peripheral route, like when investors use social media. By this route, investors interpret others' sentiment by reading their shared content on social media, and social media sentiment can stimulate such readers' behavioral biases. Thus, social media sentiment affects the investment decisions of the readers as well.

Social media sentiment may affect stock returns differently among companies. If the valuation of a firm is difficult due to information asymmetry, investors will search more on social media to get information. Accordingly, the firm's value is likely to depend more on social media content (Seok et al., 2019). In addition, investors' vulnerability toward sentiment varies in different market conditions, i.e. sentiment affects investors differently in the bullish market compared to the bearish market (Chung et al., 2012). In this work, we have empirically analyzed if firm characteristics and market conditions are moderating the predictability of social media sentiment.

Empirical findings of the relationship between social media sentiment and stock return are inconsistent with the theories behind the causality effect of investors' sentiment. Some researchers have shown that social media sentiment can perfectly predict stock returns (Risius et al., 2015; Bartov et al., 2017; Sul et al., 2017). However, others like Kim and Kim (2014) and Wang et al. (2022) presented no predictive power for the social media sentiment which is inconsistent with the theories behind the sentiment. In addition, among those studies that have divided social media sentiment into different dimensions such as positive/negative, some works mentioned predictability power for just one of the dimensions, either positive or negative (Risius et al., 2015; Wang et al., 2022), while each dimension is supposed to have a different effect on the stock return compared to the other dimensions (Hassanein et al., 2021).

Such inconsistencies could be the result of inaccurate measurements for the social media sentiment and methodology limitations for comparing the positive and negative social media sentiment with each other. Previous studies have been done mostly on Twitter, which is a non-financial platform and may bring irrelevant information about investors' sentiment (Sprenger et al., 2014). In addition, these works have used traditional text analyzing methods such as using word dictionaries for counting positive and negative words, though modern techniques such as using artificial intelligence and machine learning procedures for distinguishing the tones of social media shared contents would come up with more accurate results (Das & Chen, 2007). In this study, we have used Bloomberg's expert-trained machine learning process which analyzes the sentiment of the shared messages on the two biggest financial and non-financial social media platforms (StockTwits and Twitter, respectively), and brings us an opportunity to perfectly distinguish different dimensions of social media sentiment and compare the positive and negative social media sentiment effects with each other.

The objective of this thesis is to use social media sentiment to explain changes in S&P500 stock return. We have contributed to the literature by first, clarifying the role of social media sentiment in predicting stock return which has been inconsistent with the theories in the literature. Compared with the previous studies, this work not only used machine-learning-based approaches whose accuracy is higher than other approaches (Das & Chen, 2007; Loughran and McDonald, 2011) but also extracted sentiment from both StockTwits and Twitter, the biggest financial and non-financial platforms, which brings more relevant information about investors' sentiment (Sprenger et al., 2014). The second contribution is theoretically and empirically comparing the effects of positive and negative social media sentiment. Third, to our best knowledge, this study is the first that investigates whether social media sentiment varies among different firm characteristics and market conditions.

The findings of this work is significant to both the policymakers and investors (whether institutional or individual). Policymakers may use the findings to increase market efficiency and encourage more arbitrage activities among firms with stronger social media sentiment effects. By doing this, they would decrease the possibility of huge losses for the investors, likewise the GameStop hike and down story which started from a shared content on Twitter in early 2021. From the investors' side, social media sentiment can be used for predicting stock return. The investors can use the findings of the comparison between positive and negative social media sentiment and the moderators of this relationship, to construct investment plans based on this systematic risk factor.

The remained parts of this work are structured as section II provides a literature review about the studies on investors' sentiment and its relationship with stock/market return, in sections III and IV, we have our hypothesis development, empirical model, and the methods used in this work, respectively. Section V provides the result interpretations and the implemented robustness tests, and we have provided a comprehensive discussion in section VI. Finally, we will have a conclusion in section VII.

CHAPTER 2: Literature Review

2.1 Theories behind investors' sentiment

Investors act completely rational under the Efficient Market Hypothesis (EMH), and all information is adjusted in asset prices fast enough to not bring any opportunity for beating the market (Fama, 1965). However, several abnormal events in the financial market have proven that stock mispricing is unavoidable, and made the researchers think about factors beyond traditional finance, such as those related to investors' behavior. Delong et al. (1990) introduced a noise trader model in which the irrational behavior of traders is the main reason behind the market mispricing. Under their framework, the behavioral biases of noise traders make them have irrational investment decisions such as avoiding diversification, risks, and other fundamental principles. To beat such irrational behaviors, rational arbitrageurs try to find the mispriced firms and fill the gaps with their investments. However, limits on arbitrage activities would increase the market vulnerability to the noise traders, and the number of mispriced firms due to their irrational trades.

Following Delong et al. (1990)'s model, Lee et al. (2002) investigated the mutual fund activities under different levels of investors' sentiment. They showed that investors' sentiment is an appropriate indicator of irrational noise trader activities and can be used for calculating the intrinsic value of the mutual funds. Baker and Wurgler (2006) also defined the reasons behind stock mispricing as either sentimental demand shocks or limits on arbitrageurs. Having extracted investors' sentiment based on their propensity to speculation, they showed that salient characteristics of firms (such as "high dividend" or "high earning") would affect the investment decisions of the investors, so the stock return. Their empirical findings show that investors' sentiment can be used to predict financial bubbles and individual stock returns.

2.2 Different sources for extracting investors' sentiment

We can divide the sources for extracting investors' sentiment into three different groups. The first and the most typical one is the surveys filled by the investors. Using the findings from a survey done by the University of Michigan, Charoenrook (2005) showed that investors' sentiment has an explanatory power on the market return. Following their work, Lemmon and Portniaguina (2006) used Conference Board surveys and showed that investors' sentiment can predict stock return. Using sentiment extracted from Investor Intelligence surveys, Lee et al. (2002) have indicated a positive relationship between investors' sentiment and the market return in the long run. Charoenrook (2005) and Lemmon and Portniaguina (2006) used the same surveys as Lee et al. (2002), though they found the relationship between sentiment and market return with opposite signs. In addition to the inconsistent results, sentiment extracted from surveys could not be divided into different dimensions such as positive and negative (Sprenger et al., 2014). That is why the researchers tended to use other sources for extracting investors' sentiment.

The second group of sources is using price and market-based variables to measure investors' sentiment. Edelen et al. (2010) stated that retail investors' tendency toward risky assets can be used for measuring the sentiment. Previously, Baker and Wurgler (2006) constructed a sentiment index based on six different market variables such as closed-end fund discounts, stock turnover, the equity shares in new issues, the IPO returns in their first days and the dividend premium. The problem with this group of sources is that they are related to the whole market's sentiment, and it is not applicable for measuring investors' sentiment uniquely for each firm.

The third group of sources that can be used for measuring investors' sentiment is the media. Prior studies approve that media contents have a significant effect on stock returns by distributing new sentiment to the market (See Qian & Rasheed, 2007; Fang & Peress, 2009). Media content can be divided into two main categories: mass media and social media. Mass media includes newspapers, journals, and any other ways that investors get access to experts' analyses. However, the common people cannot publish their own beliefs on such media. Using the Wall Street Journal's content, Tetlock (2008) built an investor sentiment index and shows that it can be used for predicting stock market return. Garcia (2013) also shows a strong relationship between the mass media sentiment and Dow Jones Industrial Average Index return using New York Times content between 1905 till 2005. However, Antweiler et al. (2004) found sentiment extracted from Yahoo Finance cannot predict stock returns. The same result is found in Nardo et al. (2016) who investigated if the Wall Street Journal's positive and negative content has a predictive power of future market return, and the result was inconsistent with the theories behind sentiment which expects a relationship between sentiment and stock return.

On the other hand, some studies believe that sentiment can better explain investors' behavior if is extracted from social media in which not only experts but also common investors can share their sentiment about a specific stock or the whole market (Chen et al., 2010). Simply, social media platforms have richer language than mass media which makes social media got attracted more investors (Wang et al., 2022). Twitter, Facebook, StockTwits, SeekingAlpha, and any other social media platform in which people can freely share their moods, can be categorized in this group. Using 5.5 million tweets about S&P100 firms shared on Twitter, Risius et al. (2015) show that social media sentiment can predict stock return. Consistently, Bartov et al. (2018) measure the sentiment of investors based on their tweets and indicate a significant predictive power of the investors' aggregate mood. Also, having analyzed 2.5 million Twitter posts around S&P500 firms, Sul et al. (2017) show that trading strategies based on social media sentiment can bring up to 15%

annual return. Hassanein et al. (2021), also, found a significant positive relationship between Twitter sentiment and stock return.

Social media platforms other than Twitter are also used in the literature for measuring social media sentiment. Renault (2017) believes that one of the great features of the StockTwits platform is that writers can assign bullishness or bearishness tags to the shared content. Distinguishing investors' sentiment based on this feature, they show that StockTwits sentiment can predict the S&P500 contemporaneous index return. Consistently, Lachana and Schroder (2022) compared the predictive power of mass media sentiment and social media sentiment and indicate a significantly higher predictive power of Seeking Alpha sentiment than Wall Street Journal contents. The reason behind the higher predictive power of social media sentiment compared to the mass media sentiment is that social media includes a larger number of ideas, leading to having much larger information[1] that can be used for firm valuations (Lanchana & Schroder, 2022). In this work, shared content on both Twitter and StockTwits has been used for measuring social media sentiment.

2.3 Importance of social media over mass media

There are several benefits of social media over mass media, all have made social media have more effects on the readers or investors than the mass media. First, social media platforms are free to use. However, mass media users should pay membership fees to get access to all the ideas. Investors use 2.8 channels, on average, to get the required information for their investment decisions, and the internet forums and social media got more attractive due to their low-cost characteristics (Wang et al., 2022).

[1] It has been called as "Wisdom of Crowd" in the literature.

Second, social media platforms have larger amounts of contributors, and people believe that such a high number of publishers can result in more information than the mass media, in which only a limited number of experts can post. Simply, the social media platforms have more wisdom of crowds, which makes such platforms got attracted more than the mass media (Sprenger et al., 2014; Lanchana & Schroder, 2022). Third, readers can interpret the notes faster when they are using social media compared to the mass media because of their brief contents. Social media platforms, such as Twitter and StockTwits, restrict their users to publish not more than 140 characters per message, which makes the messages be summarized and interpreted much faster than other sources (Oh & Sheng, 2011).

The fourth reason behind the potential higher tendency of investors toward social media is that social media content is pushed to the users, due to the recommendation algorithms of such platforms. Mass media users have to look for their favorite content themselves, but recommendation algorithms of social media platforms analyze their behavior, find their favorite areas, and push the users to see the content with much less effort (Li et al., 2018). Last but not least, mass media are updated less frequently than social media. Wall Street Journal, as a famous mass media, publishes each day around midnight Eastern Time (ET), when the market is closed (Wall Street Journal, 2022). However, social media platforms do not have any specific date and time and get the investors updated with any types of information more frequently. In this work, we have used social media which has a higher attraction to the investors.

2.4 Dimensions of Social Media Sentiment

Literature has proposed different dimensions for social media sentiment. Baker and Wurgler (2006) define sentiment as how much people are optimistic or pessimistic about an individual stock or the whole market. In line with this definition, two main dimensions are mostly assumed for the sentiment in the literature: positive and negative. Bhayani and Huang (2009) defined these two as "personal positive and negative feelings". Having assumed that investors are human beings with lots of moods, Risius et al. (2015) divided human moods based on their "valence" into positive and negative, too. Valence helps to distinguish between moods behind a content by dividing content into what is perceived as pleasant (positive) and unpleasant (negative) (Risius et al., 2015).

Antweiler et al. (2004) believe that words with positive tones increase the aggregate demand toward a specific stock, and using 1.5 million shared messages on social media, show that the number of positive words has statistically significant predicting power of Dow Jones stock return. Li (2008) believes that only negative words are affecting investors' emotions and investment decisions. Counting the negative words of the shared contents, they show that the lower the negative words, the higher the return of the stocks. Consistent with this result, Loughran and McDonald (2011) changed the list of negative words and count the same content's negative words again to check if the findings of the previous studies are robust. Their findings show that the number of negative words still has significant predictive power on stock performance. Tetlock et al. (2008) also found the number of negative words as the reason behind the relationship between social media sentiment and downward movements in the Dow Jones Industrial Average index. Checking this relationship at the stock level, Chen et al. (2014) show the total number of negative words among the shared contents in the Seeking Alpha platform negatively affects the future stock return. However, none of the mentioned studies have used both positive and negative sentiment in one model to compare them with each other, while the prospect theory (Kehneman & Tversky, 1979) has stated that the negative information would be assessed more than the positive ones, and this comparison could help to empirically approve this theory in the sentiment framework.

2.5 Approaches for measuring social media sentiment

The frequency of emotional words indicates the tone of the report and can be used for distinguishing the sentiment of investors (Das & Chen, 2007; Loughran & McDonald, 2011). Among the literature, the widely used approaches for measuring social media sentiment are using dictionary-based methods. Dictionaries are defined as the list of words that are categorized based on their tones and valences into different groups of emotions (Loughran & McDonald, 2011). Several studies have come up with using Loughran and McDonald's (2011) lexicon to count sentimental words and classify words into positive and negative groups. Other lexicons such as Harvard-IV can also be used for categorizing words into positive and negative groups, but Loughran and McDonald's dictionary has a wider range of financial tonal words. Simply, 73.8% of the negative word counted using Harvard-IV are not considered negative in a financial context (Loughran & McDonald, 2011). Thus, better findings are expected from Loughran and McDonald's dictionary. Using this approach, Lachana and Schroder (2022) indicate a strong relationship between the sentiment of Seeking Alpha's content and the stock return. However, using dictionary-based approaches would be subjective to the list of tonal words and may avoid the researcher to extract an accurate investor sentiment from the shared content (Das & Chen, 2007).

Another approach for measuring investors' sentiment from their shared contents is using machine learning techniques such as Natural Language Processing (NLP). In this process, experts will manually train an algorithm and teach it how to distinguish the tone of a note. Then this program can automatically categorize the messages into different tones such as positive/negative or any other classification. Compared to the former approach, this process is not restricted to a list anymore, and it gets far from being a subjective technique and has fewer biases (Das & Chen, 2007; Loughran and McDonald, 2011). In this work, we have benefited from a supervised machine learning technique that has been regularly trained by Bloomberg's experts and can perfectly distinguish the tones of the share content on Twitter and StockTwits.

2.6 Social media sentiment and stock return

The direction of the relationship between social media sentiment and stock return is skeptical. While several empirical studies show that sentiment is predicting stock return (Sun et al., 2016), other studies indicate that stock return is causing sentiment (Das et al, 2007; Kim & Kim, 2014). For shedding light on this direction, Ganesh & Iyer (2021) used Vector Auto-Regressive (VAR) approach and showed that social media sentiment is causing changes in stock return while there is no relationship oppositely. Consistently, Wang et al. (2022) investigated the relationship between social media sentiment and stock return in the Chinese stock market and stated that the investors' social media sentiment can predict the current day's stock return, while daily stock return cannot predict the current day's social media sentiment. In this work, we also have used the Ganger causality technique to make sure about the direction of this relationship.

Our work is also related to that research about how long the sentiment effect will remain. The short-term effect of social media sentiment on stock return has been approved in previous studies as well. Tetlock (2008) believes that the effect of sentiment starts to wane about half an hour after the publication but lasts for several days. Previously and consistently, the empirical study of Gidofalvi (2001) showed that we can expect the sentiment effect starts approximately 20 minutes after the shared content is released on social media. Findings of Tan & Tas (2021) also implicate that social media sentiment has higher predictive power of contemporaneous stock return, rather than the next day's return. The common characteristic among all the mentioned works is that the social media sentiment effect is restricted to the short time horizons, and we can expect more effects in the short term. Thus, in this work, we also have assumed that the social media sentiment effect lasts shortly and would show itself in the current day's stock return.

Empirical findings of the relationship between social media sentiment and stock return are inconsistent with each other, even among those research that has assumed short-term effects of social media sentiment effect. Using over 100,000 shared messages on social media, Sprenger et al. (2014) and Renault (2017) both showed that there is a positive relationship between investors' sentiment and stock return. Wang et al. (2022) indicated that even if the abnormal return is used instead of the stock return, the predictability power of the social media sentiment will remain significant. Having used a Chinese social media platform (EastMoney), they showed that social media sentiment has a positive relationship with the stock's abnormal return. On the other hand, Schmeling (2009) investigated the relationship between investors' sentiment and stock return in 18 industrialized countries and found a negative relationship between these two which is inconsistent with what was previously mentioned. A negative relationship between social media sentiment and stock return also appeared even after an out-sample analysis in Huang et al. (2022)'s work. The empirical findings are inconsistent with each other and with the theories behind the sentiment and require to be clarified at this stage.

2.7 Moderators of social media sentiment effect

Relationship between the social media sentiment and stock return depends on several factors. Sprenger et al. (2014) state that active users that have less activity on social media (silent majorities) have posted more relevant messages and will have a higher effect on other investors' investment decisions. Their study implicates that sentiment extracted from the BitcoinTalk platform has a more significant relationship with Bitcoin return when the experiment is done among the salient majority. In addition to the users' activity, the empirical study of Tetlock (2011) finds that firms with higher attention on social media have higher sentiment effects. Chung et al. (2012) extract the investors' sentiment from market variables (but not social media) and state that the sentiment effect is significant only during market expansion states, but not in recessions.

Several studies investigated the sentiment effect differences among various firm characteristics, though the investors' sentiment measurements used in their work are not derived from social media. Baker and Wurgler (2006) used market variables and IPO information to build an investor sentiment index and stated that the sentiment effect is stronger for the firms that are harder to measure their intrinsic value. As the firm characteristics of such firms, they selected size, age, volatility, unprofitability, book-to-market ratio, dividend payout ratio, and leverage, then empirically showed that the sentiment effect on stock return is stronger among the firms that are small, young, volatile, unprofitable, having a higher book to market ratio, non-dividend paying and distressed. Consistently, Yang et al. (2017) implicate that firms that are small, volatile, and with higher book-to-market ratios and lower prices are harder to value, and the sentiment effect is stronger for them. However, the sentiment used in all these works was not extracted from social media, but from the price and market-based variables.

Our work contributes to the literature from various views. First, we have empirically and theoretically investigated which social media sentiment dimension (positive/negative) has a stronger effect on explaining the stock return. Second, we have checked if the social media sentiment effect can be moderated by the firm characteristics, social media activity, and market conditions. As the robustness test, we have checked the causality relationship between social media sentiment and stock return and investigated if social media sentiment is a systematic risk separated from other previously studied risk factors or not.

CHAPTER 3: Research Hypothesis

When it comes to predicting the stock/market return, there are two schools of thought: those researchers who believe in the traditional finance theories such as the Efficient Market Hypothesis (EMH) and those who analyze the market based on behavioral finance theories. Based on EMH, new information should be adjusted in the price quickly (Fama, 1965). Gradual Information Flow theory states that if information diffuses slowly, there will be a time difference between information release and its adjustment in stock price. Meanwhile, individual investors may use the media information to better assess the information they need to calculate the intrinsic value of the stock (Sul et al., 2017).

On the other hand, behavioral finance states that investors have behavioral biases, and their investment decisions are affected by their sentiment (DeLong et al., 1990; Ho & Hung, 2009). Investors express their sentiment in their shared content on social media when they are talking about a specific stock (Sprenger et al., 2014). By analyzing social media content, we can figure out investors' sentiment that is affecting the readers' sentiment toward a specific stock (Lucey and Dowling, 2005). Thus, social media sentiment can explain the changes in stock return, and our first group of the hypothesis will be developed as follows:

HYPOTHESIS 1: There is a significant relationship between sentiment and changes in stock price.

Social media content can have positive and negative sentiment, each may have different effects on stock return: the frequency of positive words in a shared content not only shows the positive sentiment of the writer toward that stock but also will send positive sentiment to the readers as well (Sprenger et al., 2014). The affected investors by the positive sentiment bring new demands toward stock and increase the stock price. Thus, the first subset of our first hypothesis will be:

HYPOTHESIS 1a: There is a significant positive relationship between positive sentiment and changes in stock price.

The negative sentiment of the shared content makes the investors have negative prospects for the stock. Thus, the relationship between negative social media sentiment and stock return is expected to have an opposite direction compared with the positive sentiment effect, and the second subset of our first group of hypotheses will be:

HYPOTHESIS 1b: There is a significant negative relationship between negative sentiment and changes in stock price.

The effect of positive and negative sentiment on the readers is not the same. Prospect theory (Kehneman & Tversky, 1979) states that investors face loss-aversion and they would over-weight small probabilities of the expected loss whilst underweighting the expected return. So that when investors read negative shared messages on social media, their investment decisions would get more affected compared to when they read about positive shared content. In addition, negative messages are distributed more easily and faster than positive messages on social media (Kimmel, 2010). Thus, it is expected that investors looking for new information on social media will get affected by negative content more than positive ones, and the negative social media sentiment would affect the stock price more than the positive social media sentiment. Thus:

HYPOTHESIS 1c: Relationship between negative sentiment and changes in stock price is economically and statistically stronger than the relationship between positive sentiment and changes in stock price.

Arbitrage activities and information are not distributed symmetric among all the firms, and the firms with fewer arbitrage activities and more information asymmetry are much harder to value, compared to those that have more arbitrageurs and symmetric information distribution (Baker & Wurgler, 2006). As a result, investors may search more about these firms on social media to get any relevant information helping them to find the intrinsic value of the firms. Hence, such investors will be more vulnerable to getting affected by social media sentiment, shared by other people about these firms. Based on this, our second group of hypotheses will be:

HYPOTHESIS 2: Relationship between sentiment and changes in stock price is higher for firms that are difficult to price due to information asymmetry or difficulty to engage in arbitrage.

Smaller firms are harder to value compared to the bigger firms, due to the more available information and more arbitrage activities in bigger firms. In addition, bigger firms have more budget for fulfilling their marketing and broadcasting needs, and there will be higher media channels covering their news and publishing their information (Wang et al., 2022). Thus, they will be more vulnerable to the sentiment effect of these firms and the hypothesis can be developed as:

HYPOTHESIS 2a: Relationship between sentiment and changes in stock price is higher for small firms, compared to big firms.

It is also harder to find the intrinsic value for highly volatile firms, compared to those firms that have less volatility in their prices since the arbitrageurs mostly avoid the higher volatile firms compared to the lower volatile ones (Yang et al., 2017). For such firms, investors look more into social media to get information helping them in finding the intrinsic value of the firms.

HYPOTHESIS 2b: Relationship between sentiment and changes in stock price is higher for high volatile firms, compared to the firms with lower volatility in their stock prices.

In addition, the distribution of the shared content in social media is not equal among the firms – some have higher attention while others have lower. When investors are spending time on social media, they will get attracted more if the number of publications about a firm is higher than other firms – when they have become trending firms (Sprenger et al., 2014). Hence, the investors will read content about such trending firms more than other ones and they will get affected by the sentiment of such content more than the others. Here, the social media algorithms also push the readers to see more content about the firms that became trending in the previous days (Li et al., 2018). Thus, it can be expected that the social media sentiment effect would be higher for the trending firms since the readers would see these firms' content more than the others, and get affected more by their sentiment.

HYPOTHESIS 2c: The relationship between sentiment and changes in stock price is higher for the firms that get higher attention on social media than those with less shared content.

It is also believed that investors react differently to the information in different market conditions. During bearish markets, investors react to the information more than when the market is uptrend and all assets are facing growth. Simply, investors in bearish markets act pessimistically, overestimate the possibility of the information, and their investment decisions get more affected by the sentiment (Chung et al., 2012). Thus, they would be more vulnerable to the sentiment behind the shared content on social media during bearish markets.

HYPOTHESIS 2d: Relationship between sentiment and changes in stock price is higher in bearish markets, compared to bullish markets.

CHAPTER 4: Research Methodology

4.1 Measurement

Twitter and StockTwits are the most frequently used social media platforms for investors to share their sentiment. By 2022, there are 217 million daily active users on Twitter which had steady growth since 2010 (Twitter, 2022). Also, Twitter's high volume and real-time characteristics help this platform to dispute information among investors and users quickly (Chen & Chong, 2010). On the other hand, StockTwits is a financial social media platform where active users will not stock and markets. By 2017, the number of StockTwits's members reached 300,000 individual active users and the growth rate significantly increased till 2021, when it has 5 million members with over 250,000 messages per day (Stocktwits, 2021). Although the active users of StockTwits are much less than on Twitter, since financial social media platforms have more relevant information about stocks than other platforms (Sprenger et al., 2014), both the most worldwide-used financial and non-financial social media platforms have been used in our work.

The sentiment variables in this thesis are measured by the data that comes from the expert-trained Bloomberg's machine-learning process. Based on Bloomberg's definition, "First, a human expert manually assigns a positive, negative, or neutral score to each news story or tweet. The labeling is based on the question: If an investor having a long position in the security mentioned were to read this news or tweet, is he/she bullish, bearish, or neutral on his/her holdings? Then, the annotated data is fed into machine-learning models, such as a support vector machine. Once the model is trained, when new information comes, the model automatically assigns a probability of being positive, negative or neutral to each news story or tweet." Then, the number of distinguished positive and negative content will be used to build our positive and negative sentiment variables.

We can construct our positive and negative sentiment measurements, following Tetlock et al. (2008)'s approach:

$$PosSent_{i,t} = \ln\left(\frac{1+Positive_{i,t}}{1+Total\_Pub_{i,t}}\right) \qquad (4.1)$$

$$NegSent_{i,t} = \ln\left(\frac{1+Negative_{i,t}}{1+Total\_Pub_{i,t}}\right) \qquad (4.2)$$

In which $Positive_{i,t}$ ($Negative_{i,t}$) is the number of shared contents about stock *i* that are distinguished as positive (negative) at day *t* by the Bloomberg's expert-trained machine learning process. Also $Total\_Pub_{i,t}$ is the total number of published messages about stock *i* on day *t*. It should be mentioned that $Total\_Pub_{i,t}$ is not just the summation of $Positive_{i,t}$ and $Negative_{i,t}$, since we have divided total shared content into three groups (we have neutral messages as well as positive and negative). In addition to positive and negative sentiment variables ($PosSent$ and $NegSent$), we have used another sentiment variable called $Company - level\ Sentiment$ as the total sentiment score of a firm on each day. It is measured as the average of positive and negative sentiment, weighted based on the confidence of the publishers (Bloomberg, 2022). $Company - level\ Sentiment$ can get any value between -1 (as the most negative sentiment) to +1 (as the most positive sentiment).

Sentiment variables are not the only variable used in this work. For checking the relationship between social media sentiment and stock return, we have used price variables such as stock $Return$ and stock $Abnormal\ Return$ as the dependent variables of our regression models. $Return$ is measured as the natural logarithm of the changes in the adjusted price of a firm compared to its previous day's adjusted price. $Abnormal\ Return$ is also measured as the difference between the stock's realized return and its estimated return from Fama-French three-factor model (1993) on each day.

As the control variable, we have followed the works in which the relationship between social media sentiment and stock return has been investigated (See Sprenger et al., 2014; Li et al., 2018; Tan & Tas, 2020). Following Tan and Tas (2020), the firm size is measured by the natural logarithm of the stock's market cap on the previous day. $Last\ week's\ Return$ is measured as the stock return during the last five trading days. Stock's $Abnormal\ Turnover$ is calculated as the difference between the stock's log turnover on each day and its average log turnover during the last five trading days (Tetlock, 2011). Following Amihud (2002), $Amihud\ Illiquidity$ is measured as the dollar volume-weighted average of the stock return during the last week. As the last control variable, $Park\ volatility$ (Parkinson, 1980) has been measured as follows:

$$Park\ Volatility = \frac{\left(Ln(H_t - \ln(L_t))\right)^2}{4\ln{(2)}} \qquad (4.3)$$

In which H_t and L_t are the high and low prices of a specific stock during the last five trading days.

In addition to the dependent variables, control variables, and sentiment factors, we have used the moderators of the relationship between social media sentiment and stock returns. The moderators are used as dummy variables and named as firm size dummy ($Size\_Dummy$), stock volatility ($Volatility\_Dummy$) and total publication of the firm ($Total\_Pub\_Dummy$). On each day, firms will be sorted based on each of these three moderators from the highest to the lowest, and will get a value of 1 if they are in the top 50 percent, and a value of 0 if they are in the bottom 50 percent. We also have used the market condition ($Trend\_Dummy$) as our last moderator. Following Lunde and Timmermann (2004), we have divided the market into bullish and bearish states using 10 percent and 5 percent thresholds, respectively. Simply, if the SP500 index has experienced increases by 10 percent, it will be assigned to a bullish phase, while decreases by 5 percent are assigned to the bearish market[2].

4.2 Data Collection

Sentimental variables of this work are extracted from the Bloomberg database. Bloomberg's expert-trained machine learning can distinguish between positive and negative messages shared on Twitter and StockTwits. In addition, price variables and firm characteristics are gathered from CRSP and Compustat databases, respectively, and have been back-tested with the Bloomberg database to assure their reliability. Our sample includes the S&P500 firms from the years 2015 till 2021. There are 505 firms in the S&P500 each year. However, for relaxing survivorship bias and having unbiased results, we have used 641 firms that have been included in the S&P500 during the period 2015 – 2021. We have used Bloomberg's database which has used Twitter and StockTwits, at the same time, to extract investors' sentiment data, and for the firm characteristics and stock price data, we have used Compustat and CRSP, respectively.

4.3 Sample Descriptive Analysis

As shown in Table A.1 (Appendix A), the mean daily total publication of each stock is 99.657 content. On average, 7.24 of them are recognized as having positive sentiment and 6.55 of content is categorized as negative, by the trained Bloomberg's machine learning procedure. Company-level sentiment, which is an aggregate sentiment score with a range between -1 and +1, has a significant mean of 2.5 percent. It shows that the number of positive content shared by the investors is higher than the number of negative content, which is consistent with the previous

[2] As a robustness test, the threshold of 10-10 percent is also investigated.

studies (Sprenger et al., 2014; Hassanein et al., 2021). Daily stock return in our sample has a mean of 0.0003 percent, which is significantly different from zero.

Table A.2 indicates the correlation coefficients among the variables used in our study. According to Table A.2, daily stock return is positively correlated with the positive sentiment (at 1 percent), and negatively correlated with negative sentiment (at 1 percent). However, the correlation coefficient between daily return and negative sentiment is 3.4 percent higher than the correlation coefficient between stock return and positive sentiment. The company-level sentiment is positively correlated (at 1 percent) with stock return. Among the variables used in our empirical model, Table A.2 shows that there is a statistical correlation between the independent variables. However, all of them are small enough to be economically insignificant.

4.4 Regression Model

The relationship between social media sentiment and stock return is so skeptical. The sentiment effect starts to decrease right after the publishment but lasts for several days (Tetlock, 2008). Gidofalvi (2001) and Lie et al. (2018) show that the highest effect of social media sentiment occurs around 20 minutes after the release and lasts for a day, which is consistent with Tan & Tas (2020). Also, Wang et al. (2022) indicate no predictive power of social media sentiment when it comes to predicting the next day's return, but significant causality for the same day's return. In this work, we investigate if there is a significant relationship between social media sentiment and stock return on the same day. For doing so, we have used below fixed-effect regression which is widely used in sentiment analysis literature (See Tetlock, 2011; Sprenger et al., 2014; Tan & Tas, 2020):

$$Ret_{i,t} = \alpha_{i,t} + \beta_1 PosSent_{i,t} + \beta_2 NegSent_{i,t} + \beta_n ControlVariables_{i,t} + u_{i,t} + \varepsilon_{i,t} \quad (4.4)$$

In which $Ret_{i,t}$, $PosSent_{i,t}$, $NegSent_{i,t}$ and $ControlVariables_{i,t}$ are stock return, positive sentiment, negative sentiment, and control variables of stock *i* at day *t*, respectively. Following the literature investigating the relationship between investors' sentiment and stock return (See Sprenger et al., 2014; Dalen et al., 2018; Tan & Tas, 2020), the control variables that have been used in this work are firm size, return of the last week, abnormal turnover, last week's Parkinson volatility (Parkinson, 1980), and last week's Amihud illiquidity (Amihud, 2002). As the fixed effect term ($u_{i,t}$), we have used Global Industry Classification Standard (GICS) classification for dividing the firms into the industries, and the years are considered for controlling the time effect in our regression.

Coefficients of our sentiment measurements will be used to test our first group of hypotheses. A positive and significant β_1 is expected by our hypothesis 1a, while a negative and significant β_2 is expected by hypothesis 1b. Our hypothesis 1c predicts that the absolute value of β_2 is significantly different and higher than the absolute value of β_1. Following Shrout and Yip-Bannicq (2017), we have used a Z-test for comparing the coefficients β_1 and β_2, and it is expected to have a larger β_2 compared to β_1 based on our hypothesis 1c.

The relationship between investors' sentiment and stock return depends on size, volatility, the firms' attention in social media and market condition (Baker & Wurgler, 2006; Chung et al., 2012; Yang et al., 2017). Based on each of these four moderating effects and at each day, we have divided firms into two groups: firms with 50 percent high and 50 percent low of the moderating variable. Then using dummy variable and below pooled regression model, we have separately investigated if any of these moderating effects are significant in the relationship between stock return and social media sentiment:

$$Ret_{i,t} = \alpha_{i,t} + \beta_1 Sentiment_{i,t} + \beta_2 Dummy_{i,t} + \gamma_1 Dummy_{i,t} \times Sentiment_{i,t} + \beta_n ControlVariables_{i,t} + \varepsilon_{i,t} \tag{4.5}$$

Where $Ret_{i,t}$, $Sentiment_{i,t}$ and $ControlVariables_{i,t}$ are stock return, sentiment, and control variables of stock *i* at day *t*, respectively. Here, we have used Bloomberg's company-level sentiment score as an aggregate measurement of social media sentiment, which is defined as "average of story-level sentiment [that can] deliver one value ranging from -1 to 1, with -1 being the most negative sentiment and 1 being the most positive sentiment" (Bloomberg, 2016). In addition, $Dummy_{i,t}$ is each of our four moderating variables used as a dummy variable in equation (4.5). These variables are 1) Size, which will get 1 if the firm is in high 50 percent based on its market cap, otherwise get zero. 2) Volatility, which will get 1 if the firm is in high 50 percent based on its Park Volatility (Parkinson, 1980), otherwise get zero, 3) Market condition which will get 1 if it is a bullish market, otherwise get zero and 4) Total Publication which will get 1 if the firm is in high 50 percent based on its number of shared contents about the firm, otherwise get zero. Equation (4.5) is regressed for each of our four moderating variables, separately. Coefficients of these four regressions will be used to test the second group of our hypotheses. According to hypotheses (2a till 2d), we shall observe γ_1 is significant and positive for the interaction of sentiment effect with volatility and total publication, and is significant and negative for size and market conditions' moderating effect.

4.5 Robustness Tests

Three robustness tests have been used in this work to decrease the concerns about the results. First, some of the independent variables used in this work are non-stationary in nature. As it can be inferred from the Figure A1 (Appendix A), the independent variables such as positive and negative sentiment, size and Amihud illiquidity are not stationary and we would have spurious regression in the case of not using the first difference of the independent variables in the model (Granger & Newbold, 1974). Equation (4.6) and Equation (4.7) will be used to regress the main regressions of this work at the first difference levels to make sure that our findings are robust.

$$\Delta Ret_{i,t} = \alpha_{i,t} + \beta_1 \Delta PosSent_{i,t} + \beta_2 \Delta NegSent_{i,t} + \beta_n \Delta ControlVariables_{i,t} + \Delta\varepsilon_{i,t} \tag{4.6}$$

$$\Delta Ret_{i,t} = \alpha_{i,t} + \beta_1 \Delta Sentiment_{i,t} + \beta_2 Dummy_{i,t} + \gamma_1 Dummy_{i,t} \times \Delta Sentiment_{i,t} + \beta_n \Delta ControlVariables_{i,t} + \Delta\varepsilon_{i,t} \tag{4.7}$$

In which $\Delta Ret_{i,t}$, $\Delta PosSent_{i,t}$, $\Delta NegSent_{i,t}$ and $\Delta ControlVariables_{i,t}$ are changes in stock return, positive sentiment, negative sentiment, and control variables of stock *i* at day *t*, respectively. The control variables are firm size, return of the last week, abnormal turnover, last week's Parkinson volatility (Parkinson, 1980), and last week's Amihud illiquidity (Amihud, 2002). In addition, $Dummy_{i,t}$ is each of our four moderating variables used as a dummy variable in equation (4.5). These variables are 1) Size, 2) Volatility, 3) Market condition and 4) Total Publication. However, we cannot consider industry (or time) fixed effect term in the model anymore, since the fixed factors would be omitted in first difference models (Wooldridge, 2008).

Second concern is about how much the sentiment effect is separated from other systematic risk factors. For checking this, we have used abnormal stock return which is measured as the difference between the expected return using Fama & French three-factor model and the realized return (Fama & French, 1993). By doing so, the relationship between social media sentiment and stock return will be independent of other risk factors such as market return, the book to market ratio, and the size effect. The expected return of each stock has been predicted using the previous year's estimates by equation (4.8):

$$R_t - r_{f_t} = \alpha_t + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_t$$

(4.8)

In which R_t and r_f are returns of the individual stock and risk-free asset on day *t*, respectively. We have used MKT_t, SMB_t, and HML_t as the market risk factor, size effect and book to market effect. We have run the regression for each stock on each day using the previous year's data, and the estimates of β_1, β_2, and β_3 have been used to predict the daily expected return. The abnormal return is measured as equation (4.9):

$$AbnRet_{i,t} = R_{i,t} - \overline{R_{\iota,t}}$$

(4.9)

In which $\overline{R_{\iota,t}}$ is the expected return of stock *i* on day *t*, predicted using the estimates of equation (4.8). Also, $R_{i,t}$ is the realized return of stock *i* on the same day. In the next step, our main regressions for testing the relationship between stock return and social media sentiment (equations 4.4 and 4.5) will be repeated, but this time using the abnormal return instead of stock return.

Third and last robustness test has been used with the purpose of checking the causality effect of social media sentiment and stock return; simply, the direction of the relationship between social media sentiment and stock return. Previously, it has been argued that the social media contents might be just users' reactions to the stock performances, and changes in the social media sentiment is caused by the changes in stock return (Tan & Tas, 2020; Ganesh & Iyer, 2021). In this case, we may have significant regression results for social media sentiment and stock return, but no predictability power. Using Granger causality test (Granger, 1969), we have checked whether social media sentiment causes the stock return or the stock return is causing the social media sentiment.

CHAPTER 5: Results

5.1 Social Media Sentiment and Stock Return

In our first regression model (Equation 4.4), we have regressed the daily return of S&P500 stocks on our positive and negative sentiment measurements, controlling for size, last week's return, volatility, abnormal turnover, and illiquidity. To avoid the potential multicollinearity problem in our models, we have done the Variance Inflation Factor (VIF) test. According to Tan and Tas (2021), since none of the VIFs of our independent variables were higher than 2, there would be no concern about the multicollinearity, which shows that correlations between the variables have no disturbing effect on the results of our regression. Table 5.1 shows the result of the fixed-effect panel regression. In the first column, positive and negative sentiment measurements have been used along with the control variables and the industry effect as the fixed effect of the panel regression. The coefficient on the positive sentiment, 0.414 percent, is significant at 1%, showing that if the daily positive sentiment increases by one standard deviation, the average daily stock return will increase by 0.414 percent. Based on the descriptive statistic (Table A.1), the average daily stock return is 0.03 percent, and a 0.414 percent increase in daily stock return would be considered as economically and statistically significant, and approves our hypothesis 1a, which tells the positive relationship between positive social media sentiment and stock price. The second column is the same as the first one but considers the year fixed effect instead of the industry. Column 2's coefficient on positive sentiment, 0.416 percent is also significant at 1 percent. Simply, one value increase in the positive social media sentiment would bring a 0.416 percent increase in stock return which shows that our hypothesis 1a is approved either using the industry fixed effect or the year fixed effect. This result is inconsistent with those studies that mentioned that positive sentiment has no relationship with stock return (Risius et al., 2015), however, positive sentiment and stock return are expected to have a positive relationship with each other based on the theories behind the sentiment.

Regarding the negative social media sentiment, the coefficient is -0.467 percent (significant at 1 percent) in the first column. It shows that one standard deviation increase in the negative social media sentiment brings a 0.467 decrease in the average daily stock return, even after controlling for size, the last week's return, abnormal turnover, Park volatility, and Amihud Illiquidity. In the second column, the fixed effect is changed from the industry effect to the year effect, compared with the first column. The result shows that the negative social media sentiment still has a significant negative coefficient (-0.473 percent significant at 1 percent) and approves our hypothesis 1b. This finding is well consistent with Loughran and McDonald (2011), Tetlock (2008), and Risius et al. (2015) but inconsistent with Wang et al. (2022) which indicate that the social media sentiment effect is restricted to the positive sentiment, but not the negative one. However, the chosen sample in their work is China stock market, in which short sales were not allowed due to the regulations and may restrict the downside movements followed by the negative social media content.

Hypothesis 1c states that the negative social media sentiment might have a stronger effect than the positive social media sentiment. Comparing the numerical value of these two measurements' coefficients in columns 1 and 2, we can conclude that the negative sentiment effect is stronger than the positive sentiment effect by 0.053 and 0.057 percent, respectively. Following Shrout and Yip-Bannicq (2017), the results from a Z-test for comparing the coefficients of positive and negative social media sentiment show a test statistic of 161.97, significant at 1 percent, which approves our hypothesis 1c.

Table 5.1 – Panel regression of the stock return on the social media sentiment

Table 5.1 presents the Panel regression result of the relationship between social media sentiment and stock return. T-stats are robust with Newey-West's (1987) autocorrelation robust standard errors for any potential multicollinearity. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively. Standard errors are presented in brackets below the estimated coefficients.

	(1)	(2)	(3)	(4)
	Pos. & Neg.	Pos. & Neg.	Comp-level Sentiment	Comp-level Sentiment
Pos Sent	4.14e-3***	4.16e-3***		
	(100.06)	(101.78)		
neg sent	-4.67e-3***	-4.73e-3***		
	(-119.41)	(-120.69)		
company-level sentiment			1.60e-2***	1.62e-2***
			(81.18)	(81.63)
Size	-1.8e-4***	-3.35e-4***	7.01e-5**	1.09e-5
	(-5.84)	(-10.86)	(2.46)	(0.38)
Last Week's Return	-1.45e-2***	-1.51e-2***	-1.04e-2***	-1.09e-2***
	(-22.23)	(-23.24)	(-15.83)	(-16.54)
Abnormal Turnover	4.31e-4***	3.44e-4***	7.12e-3***	6.72e-4***
	(5.43)	(4.33)	(8.82)	(8.33)
Park Volatility	5.37e-3***	-3.16e-3	3.19e-3***	-1.34e-3
	(5.64)	(-1.15)	(3.32)	(-1.33)
Amihud Illiquidity	4.1e-3***	2.1e-3*	4.61e-3***	3.03e-3**
	(2.90)	(1.95)	(3.22)	(2.13)
Constant	-6.71e-4**	1.83e-3***	-1.16e-3***	9.86e-6***
	(-1.93)	(5.15)	(-3.33)	(5.17)
F-value	2,469.46***	2,502.93***	1,141.73***	1,155.55***
Adjusted r-squared	2.75%	2.79%	1.12%	1.13%
Industry Fixed effect	YES	NO	YES	NO
Year Fixed effect	NO	YES	NO	YES
|FD_neg Sent| minus |FD_pos Sent|	5.3e-4***	5.7e-4***		

In the third and fourth columns of Table 5.1, the company-level sentiment has been used as the social media sentiment measurement. It is an aggregate sentiment factor of the firms on each trading day and has a range between -1 (as the most pessimistic) and +1 (as the most optimistic). Based on the descriptive statistic (Table A.1), the average of this measurement is positive 2.5 percent (significant at 1 percent), which shows that the positive content is more than the negative ones, on average. Since the coefficients of positive sentiment are positive in the first and second columns of Table 5.1, we should expect a significant and positive coefficient for the company-level sentiment in the third and fourth columns. The coefficient of the company-level sentiment in Table 5.1's column 3 is 1.6 percent (significant at 1 percent), which indicates that one standard deviation increase in the company-level sentiment brings a 1.6 percent increase in the average daily stock return. In the fourth column of Table 5.1, the changes in years are considered as the fixed effect of our panel regression. The coefficient of the company-level sentiment, 1.62 percent is significant at 1 percent and is much higher than the average daily stock return (0.03 percent) to be considered an economically significant coefficient. The results from columns 3 and 4 of Table 5.1 indicate that the company-level sentiment has a significant positive relationship with the stock return, either controlling for the year or the industry fixed effect. The positive relationship between company-level sentiment and daily stock return even after controlling other factors is consistent with Tan and Tas (2020).

5.2 Moderating Effects of the Firm Characteristics

In this study, we have tested four moderating factors to investigate if they affect the relationship between social media sentiment and stock return. For doing so, we have used each of these factors (separately) and regressed equation (4.5). Table 5.2 shows the results from these regressions. In column 1, we have selected firm size as the moderating effect and divided firms into two groups with high 50% market cap and low 50% ones. In the first column, the coefficient of the interaction (SENTIMENT * DUMMY_EFFECT (HIGH)), -0.56 percent is significant at 1% and shows that the sentiment effect on stock return is 0.56 percent higher among smaller firms, compared to the bigger ones. Since the average daily stock return change is 0.03 percent, 0.56 percent difference between the sentiment effect of small and big firms is also economically significant. This result strongly approves our hypothesis 2a which says that smaller firms have a higher sentiment effect due to their information asymmetries, and it is consistent with Baker and Wurgler (2006) and Yang et al. (2017). However, all mentioned works used investors' sentiment derived from price and market-based variables, but not social media with the potentially better ability for presenting investors' sentiment (Sprenger et al., 2014). In addition, the coefficient of sentiment in the first column, 1.85 percent (significant at 1%) implicates the sentiment effect among small firms. It means that one standard deviation change in the sentiment of small firms will result in a 1.85 percent change in the stock return, on average.

Table 5.2 - Panel regression of the return on the moderators of sentiment effect

Table 5.2 presents the Panel regression result of the moderating factors of the relationship between social media sentiment and stock return. T-stats are robust with Newey-West's (1987) autocorrelation robust standard errors for any potential multicollinearity. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively. Standard errors are presented in brackets below the estimated coefficients.

	(1)	(2)	(3)	(4)
moderating factors	Effect: Size	Effect: Volatility	Effect: Market Condition	Effect: Total Publication
Sentiment	.0185***	.0099511 ***	.0188717 ***	.0077488 ***
	(71.27)	(33.54)	(61.35)	(24.53)
Sentiment * Dummy_effect (high)	-.0056369***	.0107649 ***	-.0049158 ***	.0133762 ***
	(-14.61)	(27.17)	(-12.30)	(0.004)
	.0001954 ***	-.0004575 ***	.0020884 ***	.0001973 ***
Dummy_Effect (high)	(3.17)	(-7.03)	(4.56)	(8.16)
Size	-	.0000362 ***	.0001265 ***	-7.64e-06
		(1.33)	(33.99)	(-0.25)
Last Week's Return	-.0123664 ***	-.0109474 ***	-.0121021 ***	-.0105846***
	(-19.11)	(-16.68)	(-18.34)	(-16.10)
Abnormal Turnover	.0006229 ***	.0006878 ***	.0007691 ***	.0005581***
	(8.02)	(8.56)	(9.54)	(6.87)
Park Volatility	.0006719	-	.006588***	-.0007985
	(0.79)		(7.10)	(-0.85)
Amihud Illiquidity	.0027525**	.0037266***	.0024839 ***	.0032963**
	(0.045)	(2.63)	(1.75)	(2.32)
Constant	-.0002536 **	-.0001639**	-.003388 ***	.0000473 ***
	(-2.10)	(-0.58)	(-9.85)	(-2.55)
Adjusted r-squared	1.25%	1.24%	1.29%	1.31%

Using volatility as our other moderating factor, column 2 of table 5.2 implicate that the coefficient of the sentiment is 0.99 percent (significant at 1%), indicating the relationship between the sentiment and stock return among high volatile firms. This column also includes the coefficient of the interaction effect, which is 1.08 percent, significant at 1 percent. In other words, the relationship between social media sentiment and stock return is 1.08 percent stronger in high volatile firms, compared to less volatile firms. By this, our hypothesis 2b is accepted which states that the sentiment effect is stronger among high volatile firms, due to their hard-to-arbitrage characteristics.

The third column of Table 5.2 indicates the moderating effect of market conditions. Following Lunde and Timmermann (2004), we have divided the market into bullish and bearish states using 10% and 5% thresholds, respectively. The coefficient of interaction between sentiment and market condition, -4.92 percent, is significant at 1% and is an indicator of a stronger sentiment effect during bearish markets, compared with the bullish market, and approves our hypothesis 2d. As a robustness test, we have changed the 10%-5% threshold to 10-10%, though we still had significant results. This result is inconsistent with Chung et al. (2012) who used NBER to divide the market into expansions and recessions phases and showed that the sentiment effect is significant only in the expansion phase. However, our results support the theories behind the relationship between market conditions and the sentiment effect, which was not empirically approved by Chung et al. (2012).

As the last moderating effect, we have used total publications as an indicator of “attention” toward specific firms on social media. Table 5.2’s fourth column implicates that the coefficient of interaction between social media effect and total publication is 1.34 percent (significant at 1%), states that the sentiment effect is 1.34 percent stronger among the firms that have higher attention on social media. Previous studies by Tan and Tas (2020) also show a positive sentiment effect for firms with higher amounts of shared content on social media which is consistent with our study’s result.

5.3 Robustness Test Results

Figure A1 (Appendix A) shows that the positive sentiment, negative sentiment, total publication, and some of our control variables are not stationary and it would have brought spurious results if we use them at their level (Granger & Newbold, 1974). For avoiding this, we have repeated equations (4.4) and (4.5), but this time using their first difference form. Table A.3 indicates the results from regression Equation (4.6) in which the first difference of the current stock return is regressed on the first difference of social media sentiment measurements and the control variables. In the first model (column 1), positive and negative sentiment measurements have been used without any control variable. The coefficient on changes in positive sentiment, 0.37 percent, is significant at 1%. It means that if the daily changes of our positive sentiment measurement increase by one standard deviation, we would have a 0.370 percent increase in the average daily stock return changes. Given that the average change in daily stock return is 0.0095 percent, a 0.370 percent change is both statistical and economically significant, which shows that the previously mentioned relationship between positive social media sentiment and stock return is robust. On the other hand, the coefficient on changes in negative sentiment, -0.426 percent is statistically (at 1 percent) and economically significant as well, which tells that negative sentiment has a negative relationship with the stock return, and shows that the previously mentioned approval of hypothesis 1b is robust even after using the first difference version of our models.

Comparing the numerical value of these two measurements' coefficients, we can conclude that the explanatory power of negative sentiment is higher than positive sentiment. Also, following Shrout and Yip-Bannicq (2017), the results from a Z-test for comparing the coefficients of positive and negative social media sentiment show a test statistic of 123.3, significant at 1 percent, which approves the robustness of hypothesis 1c approve.

In the second column of Table A.3, we have added control variables to the previous model to check if our sentiment measurements are still significant or not. As Table A.3's column 2 shows coefficients of positive and negative sentiment, 0.208 and -0.254 percent, respectively, remained significant at 1%. Since they are much higher than the average daily return of our sample (0.0095 percent), positive and negative sentiment are both economically and statistically significant. In addition, the result of the Z test on the difference between positive and negative coefficients indicates a Z-statistic of 98.5, significant at 1%, consistent with our hypothesis 1c even after controlling the other factors.

In column 3 of Table A.3, we have just used the company-level sentiment score which is an aggregate sentiment factor of the firms on each day. The coefficient of company-level sentiment in column three, 1.20%, is statistically (at 1 percent) and economically significant and shows that 1 value change in company-level sentiment will bring a 1.20% change in the daily stock return changes. Compared with column 4, which is the same as column four but with adding the control variables, the coefficient of company-level sentiment is still economically and statistically (at 1%) significant, showing that our results are robust.

We also repeated the moderating regression (Equation 4.5) but this time using the first difference of the variables. Previously, we found out that the social media sentiment effect is stronger among smaller, high volatile, and trending firms on social media. In addition, social media sentiment affects the stock return stronger during bearish market conditions. However, we have tried changing the variables into their first differences to check if the mentioned results are robust. Table A.4's first column represents the moderating effect of the firm size on the relationship between social media sentiment and stock return. The coefficient of the interaction term ($\Delta Company - level\ Sentiment * Size\_Dummy$) is -0.211 percent, significant at 1 percent. It shows that the sentiment effect is 0.211 percent stronger among the smaller firms, compared with the bigger ones, which indicates the robustness of our previously mentioned results even after using the first difference levels of the variables.

In the second column, the volatility moderating effect has been used to check if the relationship between social media sentiment and stock return varies between firms with different volatilities. The interaction term ($\Delta Company - level\ Sentiment * Volatility\_Dummy$) has a coefficient of 0.622 percent, significant at 1 percent, indicating that social media sentiment effect is 0.622 percent stronger among high volatile firms. This result is perfectly consistent with what we found when we used the variables at their levels (instead of the first differences). The third column of Table A.4 shows how market conditions moderate the social media sentiment effect. The significant and negative sign of the interaction coefficient (-0.275 percent significant at 1 percent) shows that the social media sentiment effect is stronger during bearish markets and approves the robustness of our previous results.

Lastly, total publications about a firm on social media have been investigated as the moderator of the relationship between social media sentiment and stock return. The coefficient of the interaction ($\Delta Company - level\ Sentiment\ *\ Publication\_Dummy$) is 0.322 percent, significant at 1 percent. It can be inferred that social media sentiment affects the trending firms on social media stronger than those with fewer total publications, which shows that our results are robust even after using the first difference level of the variables.

As the second robustness test, we have checked if the previously mentioned results are robust when we use the abnormal return instead of the realized return to check if the social sentiment effect is independent of the other systematic risk factors or not. Table A.5 (Appendix A) implicates the results. The coefficient of changes in positive sentiment, 0.0879 percent significant at 1%, shows that one standard deviation increase in positive social media sentiment would come with a 0.0879 percent increase in daily abnormal return, on average. Since the average daily abnormal return is -0.00795 percent (significant at 1%), 0.0879 percent changes in abnormal return are both statistically and economically significant. Regarding the coefficient of negative sentiment in the first column of Table A.5, -0.099 percent (significant at 1%) indicates a significant negative relationship between the stock abnormal return and negative social media sentiment. Significant coefficients of the positive and negative social media sentiment indicate that these measurements can explain changes in the abnormal return, which has not been explained by the other risk factors such as size, book-to-market ratio, and the market risk factor. In other words, social media sentiment can be considered a separate systematic risk factor. Comparison between the coefficients of the positive and negative sentiment using a Z test (Shrout & Yip-Bannicq, 2017) shows that negative sentiment has a statistically stronger explanatory power for stock abnormal return since the absolute value of the negative sentiment is 0.0187 percent (significant at 1 percent) more than the absolute value of the positive sentiment, which is perfectly consistent with our hypothesis 1c.

The second column of Table A.5 is the same regression as the first column, but controls for size, last week's return, Amihud illiquidity measurement, abnormal turnover, and Park volatility measurement. Consistent with the previous result, the coefficient of positive sentiment is 0.0881 percent (significant at 1%) showing a positive relationship between stock abnormal return and positive social media sentiment, in the same line with our hypothesis 1a. On the other hand, the coefficient of negative sentiment in the second column, -0.12 percent (significant at 1%) indicates an economically and statistically significant relationship between negative social media sentiment and stock abnormal return, supporting our hypothesis 1b even after controlling for other variables.

The difference between the coefficients of positive and negative sentiment, 0.0208 percent (significant at 1%) shows that negative sentiment has a stronger power in explaining changes in stock abnormal return, compared to the positive sentiment, and approves our hypothesis 1c.

Table A.5's third column indicates the coefficient of the company-level sentiment, which is the daily total sentiment score with a range between -1 (as the most negative) and +1 (as the most positive). Since the average of this measurement is 2.5% (significant at 1% level), it shows that the number of positive publications is higher than negative ones, and due to the positive relationship between positive sentiment and stock return, the coefficient of this measurement is expected to be significant and positive. As column three of Table A.5 shows, the coefficient of the company-level sentiment is 0.565 percent (significant at 1%), which indicates that one standard deviation increase in the changes of this measurement brings a 0.565 percent increase in the daily abnormal stock return changes. Adding the control variables to the model, column 4 of Table A.5 shows that the positive relationship between the company-level sentiment and stock abnormal return is robust even after controlling for size, abnormal turnover, last week's return, Park volatility, and Amihud illiquidity measurement. To conclude, all three hypotheses which had been proved in the previous parts are strongly rousted even after using abnormal return instead of realized return, showing that the social media sentiment effect is independent of the other risk factors such as market risk, and size effect, and book to the market risk factor.

As the third and last robustness test, we have checked the causality effect of social media sentiment on stock return. Using the Granger causality approach (Granger, 1969) to figure out if social media sentiment is causing the stock return changes, or if the opposite direction is happening. Table A.6 presents the result of the Granger causality test. Among 641 firms in the sample, 427 firms (66.61 percent of the sample) show the causality effect of social media sentiment on stock return at a 90 percent confidence level, while only 135 (21.06 percent) firms indicate that the stock return causes social media sentiment at the same confidence level. Among these 135 firms, the causality effect of social media sentiment on stock return was also approved in 129 firms. Thus, there are only 6 firms (0.93 percent) showing that social media sentiment is the effect of stock return and cannot be used to predict stock return. Increasing the preciseness of our Granger causality test by considering the lower significance levels, the stock return of 376 firms (58.66 percent of the sample) is caused by their social media sentiment, while there are only 85 firms (13.26 percent) whose social media sentiment is caused by the stock return. Among these 85 firms, 81 firms show the causality effect for both directions. Considering the results at a 1 percent significance level, the social media sentiment of 268 firms (41.81 percent) causes their stock return, while the opposite causality is approved for only 31 firms (4.84 percent), in which 28 firms have a significant causality effect for both directions. The findings show that the number of firms whose social media sentiment is causing the stock return is much higher than those with the opposite causality directions, and it is consistent with Tan and Tas (2020) and Ganesh and Iyer (2021) that mentioned a significant causality effect of social media sentiment on the stock return.

CHAPTER 6: Discussion

6.1 Result Discussion

In this work, we have checked 1) the relationship between the investors' sentiment and stock return under the social media framework, 2) whether positive or negative social media sentiment has a stronger effect on the stock return, and 3) how this relationship differs among various firm characteristics and different market conditions. Among the literature, most works have used market variables, surveys, and mass media to measure the investors' sentiment. However, with the increasing engagement of investors on social media platforms (Sprenger et al., 2014), a new underpinning for measuring investors' sentiment has been introduced. Having a broader database, brief and straightforward messages, and richer language (Das & Chen, 2007), besides its free-to-use characteristic, have made social media platforms have better information about the investors' sentiment compared to other traditional sources. On social media platforms, investors not only will write about their sentiment toward specific firms but also, would read about people's sentiment and get affected by them. In this work, we have empirically shown that we can use social media sentiment extracted from Twitter and Stocktwits platforms for predicting stock returns.

Our findings show that the positive sentiment has a positive relationship with the stock return, even after controlling for firm characteristics and other previously studied risk factors. This is inconsistent with those works in which the positive sentiment is not important in explaining the stock return changes, such as Risius et al. (2015) which used Twitter for measuring the investors' sentiment. However, empirical findings of Sprenger et al. (2014) show that financial social media platforms such as StockTwits have more relevant information for the investors compared to non-financial ones such as Twitter and Facebook. It can be counted as one of the reasons behind finding no explanatory power of positive social media sentiment in Risius et al. (2015)'s work.

Talking about the negative social media sentiment, our findings show that negative sentiment has a negative relationship with the stock return, which is consistent with Loughran and MacDonald (2011) and Risius et al. (2015). These two works have used dictionary-based approaches for counting positive and negative words. However, such approaches are subjective and restricted to words existing in the dictionary (Das & Chen, 2007). In this work, we have used the Bloomberg-trained machine learning approach which has been getting trained since 2015 and can distinguish the valence of the shared contents with high preciseness.

Theoretical and empirical comparison between the positive and negative sentiment effect is one of the contributions of this work. In this work, we showed that the negative sentiment has a 0.057 percent stronger effect on the stock return, which is both statistically and economically significant. It is because investors' loss aversion, leads them having overestimated the expected loss and got affected more by the bad news and negative sentiment. Such a stronger effect of the negative sentiment is expected to be weaker among the countries in which short sales are not allowed and can be checked in further research.

6.2 Research Theoretical Implications

Behavioral finance theories state that investors' sentiment affects their investment decisions and so, the stock prices. Our model also showed that investors' sentiment can significantly explain the changes in stock returns, and clarify the inconsistent results in the literature. We also had a comprehensive comparison between positive and negative sentiment effects. The comparison showed that negative social media sentiment has a stronger relationship with the stock return, compared with the positive one. This result perfectly approves the prospect theory (Kehneman & Tversky, 1979) which states that investors overestimate the expected losses when they hear about bad news, while they underestimate the expected returns from good news, leading to getting affected more by the negative content.

Based on our literature review, we found a need for empirically testing the theories behind the moderation of the sentiment effect, under the social media framework. Investors' sentiment affects their investment decisions and may cause mispricing in firms. Simply, the price will have a gap with its intrinsic value. In such cases, arbitrageurs find the mispriced firms and the gap will tend to be filled by their investments (Baker & Wurgler, 2006). However, there might be some situations in which arbitrageurs cannot offset the mispriced firms well, and the sentiment effects would be stronger among those firms and situations. Our findings perfectly approve this fact and show that smaller and high volatile firms (as a common characteristic among the firms with limits on arbitrage) have a stronger social media sentiment effect. In other words, the moderation of the sentiment effect is empirically and theoretically approved in this work.

This work also tries to solve the conflicts between traditional finance and behavioral finance frameworks and contributes to the literature by stating that the sentiment effect is a systematic risk separated from the previously studied systematic risk factors. Our result showed that the social media sentiment effect can explain a part of stock return changes which has not been explained by the traditional finance risk factors such as size, book-to-market ratio, and the market return. It can be concluded that asset pricing models work much better when they include both behavioral and traditional risk factors beside each other.

6.3 Managerial Implications

6.3.1. Research Implication to the Policy Makers

Policymakers and market regulators can use our findings to increase market efficiency, which means there is much less predictability of the stock prices by the investors. However, social media sentiment can be used to predict the stock return, and any effort for decreasing this predictability would increase the market efficiency.

We showed that the predictability of social media sentiment is stronger among the firms that are facing limits in arbitrage activities. Simply, the smaller firms are facing more information asymmetry, so the arbitrage activities would be fewer for such firms, and the sentiment effect would be stronger. However, if the regulators encourage such firms to be more transparent in their information and have more effective ways to publish the firm's news, their information asymmetry would decrease, and the relationship between social media sentiment and stock return would be weakened. We also showed that the social media sentiment effect is stronger during bearish market conditions, compared to bullish ones, and among trending firms compared with those firms with a smaller number of publications on social media. It shows that such arbitrage encouragement policies would result better if they were done during market downtrends and among these firms.

6.3.2. Research Implication to the Investors

Our findings show that social media sentiment is a risk factor separated from the other previously studied risk factors such as the Fama French three factors (1996) and having it in the asset pricing models increases the chance of predicting the stock return. However, the social media sentiment effect is restricted to the short-term horizons. Daily traders who look for speculations and gain from the daily changes in stock prices, might have better results by entering the social media sentiment into their models. Since our findings indicate that the negative social media sentiment effect is statistically and economically stronger than the positive one, traders and investors gain more by assigning more weights to the firms with high negative sentiment in their strategies.

In addition, our findings show that the social media sentiment effect is stronger among the firms with smaller size, higher volatility, higher attention on social media, and during bearish market conditions. Thus, the expected return from the strategies based on social media sentiment would be maximized if gets done on such firms. Simply, investors and traders can have a long-short investment strategy based on the social media sentiment factors of the smaller, high-volatile, and trending firms on social media, especially during bearish markets.

6.4 Research Limitation and Future Research

One of the challenges of this work was sentiment database limitation. We have used data resulting from Bloomberg's expert-trained machine learning process which has been started work since 2015, leading to seven years data frame. The firm characteristic variables are updated quarterly makes us have 28 data points for each firm. However, a wider data range for the sentiment database could have helped us to use more firm characteristic data in our models. As another data limitation, we should mention the frequency of our sentiment variables. Based on the literature mentioned in the work, the sentiment effect is likely to be short-term and sentiment variables with high frequencies (such as intraday data) result in better predictions (Tetlock, 2008; Tan & Tas, 2020). However, the highest frequency of the sentiment variables measured by Bloomberg's machine-learning process was daily. We would have gotten better explanatory results if such variables were measured intraday. As the last found limitation, the findings of this work cannot be compared to that of research in which investors' sentiment are measured from the market variables. We have mentioned several benefits of social media for measuring sentiment compared with the market variables, and it was better if we could empirically test them. However, works such as Baker and Wurgler (2006), have come up with investors' sentiment as a cumulative measurement of the market sentiment, but not as a cross-section among the firms. But our work has a panel of investors' sentiment among the SP500 firms and cannot be compared with the market sentiment.

For further research, some ideas can be provided here. We have stated and empirically showed that the negative sentiment has a statistically and economically stronger effect on the stock return, compared with the positive return. The reason behind it is that investors got affected more by negative content, due to their loss aversion and overestimating the expected losses. Such investors would make downward movements by closing their already-opened positions, or by doing short sales on the stocks. However, short sales are not allowed in all countries, and it may weaken the negative social media sentiment effect. It can be tested if the negative sentiment effect is still stronger among the countries in which the short sale activities are forbidden or not. In addition, we know that arbitrage activities are fewer among the countries with emerging economies, compared with developed countries. By constructing a trained machine learning process and extracting the investors' sentiment from the famous financial and non-financial social media platforms of those countries (such as EastMoney in China), it is worth testing if there is any difference between the sentiment effect based on the development of countries' economic situations or not.

CHAPTER 7: Conclusion

In this study, we have examined the relationship between social media sentiment and S&P500 stock return between the years 2015 and 2021. During the literature review presented in this work, we found inconsistent empirical results for the relationship between social media sentiment and stock return especially when it comes to different dimensions of social media sentiment such as positive and negative. To fill the gap, we have used Bloomberg's machine learning process which is not only a newer and more efficient technique than the previously used methods but also, extracted the investors' sentiment from the two most popular financial and non-financial social media platforms. Our findings indicate that one standard deviation increase in positive (negative) social media sentiment would change the daily stock return by 0.416 (-0.473) percent, and sentiment can be used to predict the current day's stock return. In addition, this study shows that negative sentiment has a statistically and economically stronger effect on the stock return, compared with the positive sentiment.

The sentiment effect is not the same for all companies. Based on the theories, firms with more limits for their arbitrageurs would have more noise trader activities which are based on nothing than their sentiment. We have checked these theories under the social media framework, which is a better source for measuring the investor's sentiment, compared with the previously used sources such as surveys, price and market-based variables, and mass media (Sprenger et al., 2014). We have used size, volatility, market condition, and total publication, and investigated if any of them moderate the relationship between sentiment and stock return. Our findings empirically approve the theories behind the moderators and show that the sentiment effect significantly varies among different firm characteristics, social media activities, and market conditions. Simply, small and high volatile firms and the firms with a higher portion of publications on social media would have higher sentiment effects on stock return changes. Also, the sentiment effect is stronger during bearish markets.

REFERENCES

Antweiler, W., & M. Z. Frank. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance,* 59(3), 1259–94. doi:10.1111/j.1540-6261.2004.00662.x

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56. Doi: https://doi.org/10.1016/S1386-4181(01)00024-6

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 61 (4),1645–80. doi:10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler. J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21 (2), 129–52. doi:10.1257/jep.21.2.129

Bartov, E., Faurel, L., & Mohanram, P. (2018). Can Twitter Help Predict Firm Level Earnings and Stock Returns?. *The Accounting Review.*, 93 (3), 25–57. doi:10.2308/accr-51865

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2 (1), 1–8. doi:10.1016/j.jocs.2010.12.007

Charoenrook, A. (2005). Does sentiment matter. Vanderbilt University (Unpublished working paper).

Chen, H., Chong, T. T. L., & Duan, X. (2010). A principalcomponent approach to measuring investor sentiment. *Quantitative Finance*, 10(4), 339–347. Doi: https://doi.org/10.1080/14697680903193389

Chen, H., De, P., Hu, Y. J, & Hwang, B. H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media. *Review of Financial Studies*, 27 (5), 1367–403. doi:10.1093/rfs/hhu001

Chung, S.L., Hung, C.H., & Yeh, C.Y. (2012). When does investor sentiment predict stock returns?. *Journal of Empirical Finance*, 19 (2), 217–240. doi:10.1016/j.jempfin.2012.01.002.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, 53(9), 1375–1388. Doi: 10.1287/mnsc.1070.0704

DeLong, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), 703–738. doi:10.1086/261703

Edelen, R., Marcus, A., Tehranian, H. (2010). Relative sentiment and stock returns. *Financial Analysts Journal*, 66 (2010), pp. 20-32

Fama, E. (1965) . The behavior of stock market prices. *The Journal of Business*, 38(1), 34–105. Doi: https://www.jstor.org/stable/2350752

Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5.

Fang, L.,& Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023–2052. Doi: https://doi.org/10.1111/j.1540-6261.2009.01493.x

Fawcett, T., & Provost, F. (1999). Activity monitoring: noticing interesting changes in behavior. Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA. Doi: https://doi.org/10.1145/312129.312195

Ganesh, A., & Iyer, S. (2021). Impact of Firm-Initiated Tweets on Stock Return and Trading Volume, *Journal of Behavioral Finance*, 3(1), 1-12. DOI: 10.1080/15427560.2021.1949717

Garcia, D. (2013). Sentiment during Recessions. The journal of finance, 68(3), 1267-1300. https://doi.org/10.1111/jofi.12027

Gidofalvi, G. (2001). Using news articles to predict stock price movements, Master thesis of Department of Computer Science and Engineering, University of California, SanDiego.

Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market, in: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 1–8. Doi: https://ojs.aaai.org/index.php/ICWSM/article/view/14023

Bhayani, R., & Huang. L. (2009). Twitter Sentiment Classification Using Distant Supervision. *CS224N Project Report*, 1 (12). Stanford.

Gidofalvi, G. (2001). Using news articles to predict stock price movements. San Diego University Thesis.

Granger, C. W. K. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424-438. https://doi.org/10.2307/1912791.

Granger, C. W. J., & Newbold, N. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2(2), 111-120.

Hassanein, A., Mostafa, M. M., Benameur, K. B., & Al-Khasawneh, J. A. (2021). How do big markets react to investors' sentiment on firm tweets?. *Journal of Sustainable Finance & Investment*. DOI: 10.1080/20430795.2021.1949198

Ho, C., & Hung, C. H. (2009). Investor Sentiment as Conditioning Information in Asset Pricing. *Journal of Banking & Finance*, 33(5), 892–903. doi:10.1016/j.jbankfin.2008.10.004

Huang, D., Jiang, F., Tu,J. & Zhou, G. (2022). Investor sentiment aligned: a powerful predictor of stock returns. Review of Financial Studies, vol. 28, no. 3, pp. 791–837. https://doi.org/10.1016/j.irfa.2022.102028

Kahneman, D., & and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, 47(2), 263 – 292. https://doi.org/10.2307/1914185

Kim, S. H., & Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*, 107 (11), 708–729. Doi: https://doi.org/10.1016/j.jebo.2014.04.015

Kimmel, A. J. (2010). Connecting with consumers: Marketing for new marketplace realities. Oxford University Press. Doi: https://doi.org/10.1080/0267257X.2014.946246

Lachana, I., & Schroeder, D. (2022). Investor Sentiment, Social Media and Stock Returns: Wisdom of Crowds or Power of Words. http://dx.doi.org/10.2139/ssrn.3842039

Lee, W.Y., Jiang, C.X., Indro, D.C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. J. Banking Finance 26, 2277–2299. https://doi.org/10.1016/S0378-4266(01)00202-3

Lemmon, M., Portniaguina, E. (2006). Consumer confidence and asset prices: some empirical evidence. Rev. Finan. Stud. 19, 1499–1529. DOI:10.2139/ssrn.335240

Li, T., Dalen, J. van., & van Rees, P. J. (2018). More than Just Noise? Examining the Information Content of Stock Microblogs on Financial Markets. *Journal of Information Technology*, 33 (1), 50–69. doi:10.1057/s41265-016-0034-2

Loughran, T. , & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance,* 66, 67–97. Doi: https://doi.org/10.1111/j.1540-6261.2010.01625.x

Lucey, B.M., & Dowling, M. (2005). The role of feelings in investor decision-making. *Journal of Economic Surveys*, 19(2), 211–237. Doi: https://doi.org/10.1111/j.0950-0804.2005.00245.x

Lunde, A., & Timmermann, A. (2004). Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets. *Journal of Business & Economic Statistics, American Statistical Association*, 22(1), 253-273.

Nardo, M., Petracco, M., & Naltsidis, M. (2016). Walking down wall street with a tablet: a survey of stock market predictions using the web. *Journal of Economic Survey*, 30(2), 356–369. Doi: https://doi.org/10.1111/joes.12102

Newey, W. K., & West, K. D. (1987). A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703–8. doi:10.2307/1913610

Oh,O., &Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. Proceedings of the International Conference on Information Systems, Shanghai.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53(2), 61–65. Doi: https://www.jstor.org/stable/2352357

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*(1), 69–81. https://doi.org/10.1037/0022-3514.46.1.69

Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, *26*(1), 25–33. Doi: https://doi.org/10.1007/s10489-006-0001-7

Renault, T. (2017). Intraday Online Investor Sentiment and Return Patterns in the US Stock Market. *Journal of Banking & Finance*, 84(2), 25–40. doi:10.1016/j.jbankfin.2017.07.002

Risius, M., Akolk, F., & Beck, R. (2015). Differential emotions and the stock market—The case of company-specific trading. *European Conference on Information Systems*, Completed Research Papers, Munster, Germany, Paper147.

Schmeling, M. (2009). Investor sentiment and stock returns: some international evidence. *Journal of Empirical Finance*, vol. 16, no. 3, pp. 394–408. DOI:10.1016/j.jempfin.2009.01.002

Seok, S. I., Cho, H., & Ryu, D. (2019). Firm-specific investor sentiment and daily stock returns. *The North American Journal of Economics and Finance*, 50(C), 1-18. Doi: 10.1016/j.najef.2018.10.005

Shrout, P. E., Yip-Bannicq, M. (2017). Inferences about competing measures based on patterns of binary significance tests are questionable. *Psychological Methods*, 22(84), 1-12.

Sprenger, T. O., Sandner, P. G., Tumasjan, A., & Welpe, I. M. (2014). News or Noise? Using Twitter to Identify and Understand Company-specific News Flow. *Journal of Business Finance & Accounting*, 41 (7-8), 791–830. Doi: https://doi.org/10.1111/jbfa.12086

Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48, 454–488. Doi: https://doi.org/10.1111/deci.12229

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: a high-frequency perspective. *J. Bank. Finance*, 73(2), 147–164. Doi: http://www. sciencedirect.com/science/article/pii/S0378426616301595 .

Tan, S., & Tas, O. (2021). Social Media Sentiment in International Stock Returns and Trading Activity. *Journal of Behavioral Finance*, 22(2), 221–34. Doi: https://doi.org/10.1080/15427560.2020.1772261

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, 63(3), 1437–67. doi:10.1111/j.1540-6261.2008.01362.x

Tetlock, P. C. (2011). All the News That's Fit to Reprint: Do Investors React to Stale Information?. *Review of Financial Studies*, 24 (5), 1481–512. doi:10.1093/rfs/hhq141

Wang, X., Xiang, Z., Xu, W., & Yuan, P. (2022). The Causal Relationship between Social Media Sentiment and Stock Return: Experimental Evidence from an Online Message Forum. *Economics Letters*, http://dx.doi.org/10.2139/ssrn.4035197

Welpe. (2014). Tweets and Trades: The Information Content of Stock Microblogs. *European Financial Management*, 20 (5), 926–57. doi:10.1111/j.1468-036X. 2013.12007.x

Wooldridge, J. M. (2008). Introductory Econometrics: A Modern Approach (Book).

Xiong, X., Meng, Y., Li, X., & Shen, D. (2020). Can overnight return really serve as a proxy for firm-specific investor sentiment? Cross-country evidence. *Journal of International Financial Markets, Institutions and Money, 64(1), 12-71* https://doi.org/10.1016/j.intfin.2019.101173

Yang, H., Ahn, H. J., Kim, M. H., & Ryu, D. (2017). Information asymmetry and investor trading behavior around bond rating change announcements. *Emerging Markets Review*, 32(1), 38–51. Doi: https://doi.org/10.1016/j.ememar.2017.05.004

APPENDIX

Table A.1 – Statistical Summary

Table A.1 presents the statistical summary of the variables used for checking the relationship between social media sentiment and stock return. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively.

Variable Names	Variable Description	Observations	Mean	Std, Dev.	Minimum	Maximum
$\boldsymbol{Ret}$	Daily Stock Return	659,504	0.00036***	0.0236034	-1.79	1.33
$\boldsymbol{Total\_Pub}$	Number of Daily Publications	744,094	99.657***	659.589	0.00	142267.00
$\boldsymbol{Positive}$	Number of Positive Content	744,094	7.248***	59.913	0.00	22332.00
$\boldsymbol{Negative}$	Number of Negative Content	744,094	6.552***	63.365	0.00	19063.00
$\boldsymbol{PosSent}$	Positive Sentiment Ratio	744,094	-1.735***	0.925	-7.77	0.13
$\boldsymbol{NegSent}$	Negative Sentiment Ratio	744,094	-1.980***	0.987	-8.12	0.00
$\boldsymbol{Company - level\ Sentiment}$	Aggregate Sentiment Score	735,010	0.025***	0.152	-1.00	1.00
$\boldsymbol{Size}$	Firm's Size	683,323	9.938***	1.188	1.93	14.90
$\boldsymbol{Last\ week's\ Return}$	Return during the last week	657,884	0.0013***	0.0461	-1.80	2.09
$\boldsymbol{Abnormal\ Turnover}$	Turnover higher than the weekly average	681,727	0.0018***	0.3929	-15.39	9.78
$\boldsymbol{Park\ Volatility}$	Parkinson (1980) volatility	706,105	0.1261***	0.0360	0.01	0.77
$\boldsymbol{Amihud\ Illiquidity}$	Amihud (2002) illiquidity measurements	706,089	0.0009***	0.0337	0.00	9.56
$\boldsymbol{Size\_Dummy}$	Dummy variable for the size effect	755,460	0.4516***	0.4976	0.00	1.00
$\boldsymbol{Volatility\_Dummy}$	Dummy variable for the volatility effect	755,460	0.4666***	0.4989	0.00	1.00
$\boldsymbol{Publication\_Dummy}$	Dummy variable for total publication effect	755,460	0.4780***	0.4995	0.00	1.00
$\boldsymbol{Trend\_Dummy}$	Dummy variable for size effect	755,460	0.5997***	0.4899	0.00	1.00
$\boldsymbol{AbnRet}$	Daily Abnormal Return	659,504	0.0003568***	0.0236034	-1.79	1.33

Table A.2 – The Pearson Correlation Coefficient Matrix

Table A.2 presents the correlations for all the variables used in checking the relationship between social media sentiment and stock return, for the entire sample. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Variables	
1.000	0.008***	0.103***	0.001	-0.013***	0.012***	-0.005**	0.002***	0.061***	-0.095***	Ret	(1)
	1.000	-0.023***	0.228***	0.022***	0.046***	0.010***	-0.005***	-0.221***	-0.169***	$Total\_Pub$	(2)
		1.000	-0.003	0.057***	0.029***	-0.042***	-0.005***	0.361***	-0.280***	$Company-level\ Sentiment$	(3)
			1.000	0.045***	0.002***	-0.354***	-0.163***	-0.304***	-0.327***	$Size$	(4)
				1.000	0.013***	-0.094***	-0.006	0.028***	-0.057***	$Last\ week's\ Return$	(5)
					1.000	-0.096***	0.004***	0.004***	-0.037***	$Abnormal\ Turnover$	(6)
						1.000	0.132***	-0.013***	0.053***	$Park\ Volatility$	(7)
							1.000	0.039***	0.045***	$Amihud\ Illiquidity$	(8)
								1.000	0.526***	$PosSent$	(9)
									1.000	$NegSent$	(10)

Figure A.1 – Stationary Test Results

Figure A.1 shows the results of Auto Correlation Function (ACF) on the stock return, stock's abnormal return, positive sentiment ratio, negative sentiment ratio, total publication, and the company-level sentiment. The grid line indicates the 1 percent significant level.

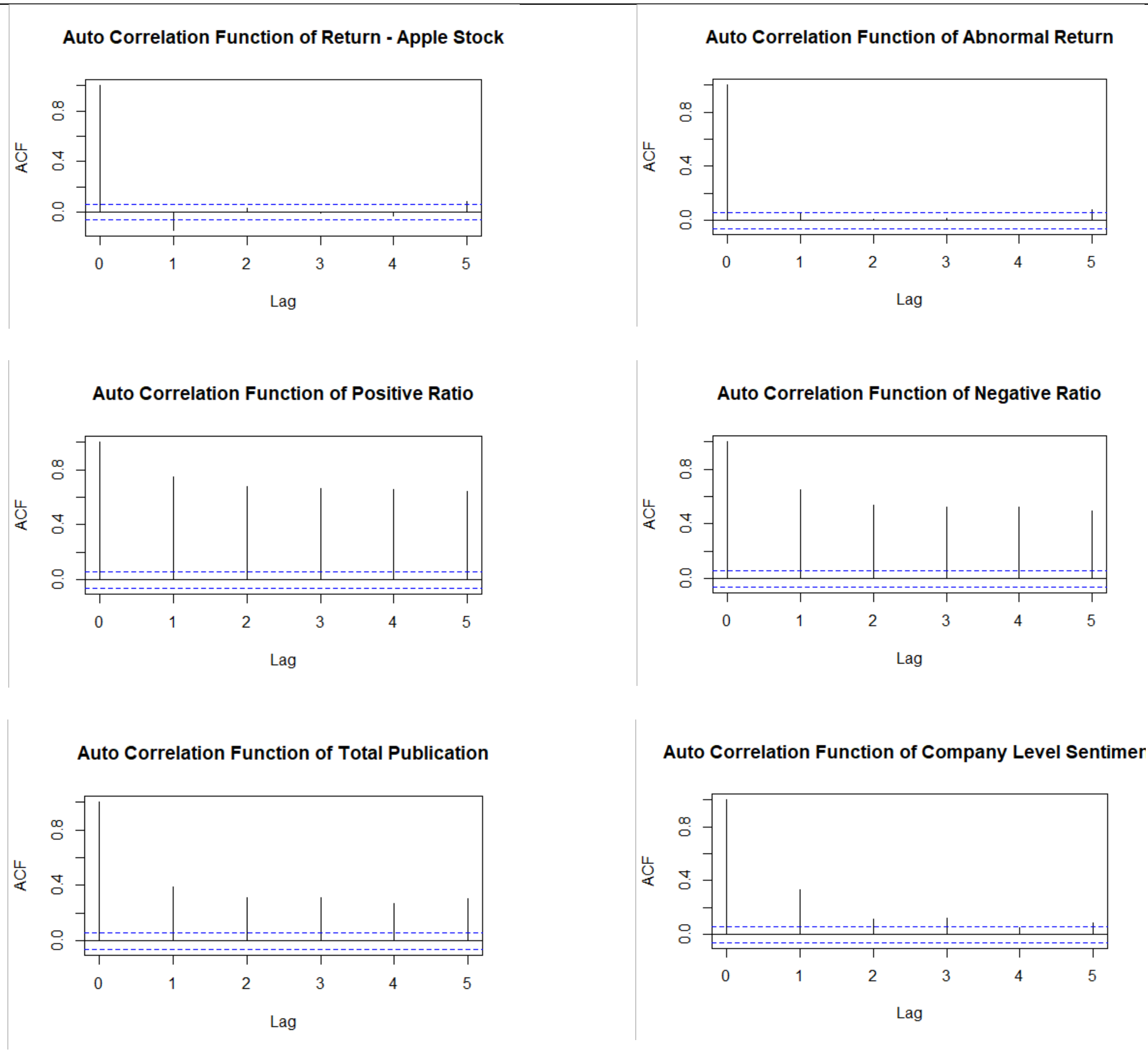

Table A.3 – Panel regression of the Stock return on Social Media Sentiment (First Difference)

Table A.3 presents the Panel regression result of the relationship between social media sentiment and stock return at their first difference level. T-stats are robust with Newey-West's (1987) autocorrelation robust standard errors for any potential multicollinearity. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively. Standard errors are presented in brackets below the estimated coefficients.

	(1) Just Pos. & Neg.	(2) Pos. & Neg. with Control Variables	(3) Just Comp-level Sentiment	(4) Comp-level Sentiment with Control Variables	(5) Just Control Variables
$\Delta \boldsymbol{PosSent}$	3.70e-3***	2.08e-3***			
	(74.32)	(44.09)			
$\Delta \boldsymbol{NegSent}$	-4.26e-3***	-2.54e-3***			
	(-92.27)	(-48.48)			
$\Delta \boldsymbol{Company}$ $- \boldsymbol{level\ Sentiment}$			1.20e-2***	6.59e-3***	
			(55.55)	(32.95)	
$\Delta \boldsymbol{Size}$		-8.57e-1***		-8.83e-1***	-8.64e-1***
		(-24.30)		(-24.29)	(-24.43)
$\Delta \boldsymbol{Last\ week's\ Return}$		-8.81e-2***		-8.79e-2***	-8.80e-2***
		(-5.02)		(-4.97)	(-4.98)
$\Delta \boldsymbol{Abnormal\ Turnover}$		6.33e-4***		7.22e-4***	7.19e-4***
		(4.31)		(4.85)	(4.86)
$\Delta \boldsymbol{Park\ Volatility}$		1.19e-2		1.07e-2	1.03e-2
		(1.01)		(0.91)	(0.87)
$\Delta \boldsymbol{Amihud\ Illiquidity}$		3.76e-2*		3.76e-2*	3.74e-2*
		(1.68)		(1.68)	(1.67)
Constant	9.42e-5***	1.44e-4***	9.78e-5***	1.50e-4***	1.51e-4***
	(2.24)	(4.47)	(2.29)	(4.60)	(4.66)
Adj. R-SQUARED	1.58%	48.47%	0.47%	48.10%	47.95%
F-value	5233.37***	6,583.10***	3,085.71***	2,853.59***	2,452.81 ***
$\lvert \Delta NegSent \rvert$ minus	4.98e-4***	4.11e-4***	-	-	-
$\lvert \Delta PosSent \rvert$	(123.3)	(98.5)			

Table A.4 - Panel regression Stock Return on the moderators of Sentiment (First Difference)

Table A.4 presents the Panel regression result of the moderating factors of the relationship between social media sentiment and stock return at the first difference level. T-stats are robust with Newey-West's (1987) autocorrelation robust standard errors for any potential multicollinearity. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively. Standard errors are presented in brackets below the estimated coefficients.

	(1) Effect: Size	(2) Effect: Volatility	(3) Effect: Market Condition	(4) Effect: Total Publication
$\Delta \boldsymbol{Company - level\ Sentiment}$	7.22e-3***	2.71e-3***	7.87e-3***	4.25e-3***
	(17.18)	(18.51)	(19.07)	(15.60)
$\Delta \boldsymbol{Company - level\ Sentiment}$ * $Dummy\_effect$ (high)	-2.11e-3***	6.22e-3***	-2.75e-3***	3.22e-3***
	(-4.22)	(11.31)	(-8.19)	(8.16)
$Dummy\_Effect$ (high)	-1.09e-3***	4.05e-4***	2.04e-3***	1.01e-3***
	(-10.11)	(6.12)	(27.32)	(8.16)
$\Delta \boldsymbol{Size}$	-	-8.33e-1***	-8.81e-1***	-8.79e-1***
		(-30.08)	(-33.99)	(-34.09)
$\Delta \boldsymbol{Last\ week's\ Return}$	-7.61e-2***	-7.55e-2***	-7.80e-2***	-7.95e-2***
	(-5.88)	(-5.02)	(-6.02)	(-6.17)
$\Delta \boldsymbol{Abnormal\ Turnover}$	7.61e-4***	7.77e-4***	7.16e-4***	6.64e-4***
	(4.91)	(4.99)	(3.63)	(3.37)
$\Delta \boldsymbol{Park\ Volatility}$	1.01e-2	-	1.60e-2	9.63e-3
	(1.06)		(1.54)	(0.92)
$\Delta \boldsymbol{Amihud\ Illiquidity}$	3.93e-2**	3.61e-2**	3.84e-2**	3.86e-2**
	(2.88)	(2.11)	(2.43)	(2.45)
Constant	7.41e-4***	-6.06e-5**	-1.09e-3***	-3.66e-4***
	(18.23)	(-2.14)	(-25.35)	(-10.47)
Adj. R-SQUARED	47.11%	44.31%	48.34%	48.26%
F-value	4,901.23***	4,879.11***	5,045.91***	4,859.61***

Table A.5 - Panel regression of Stock Abnormal Return on Social Media Sentiment

Table A.5 presents the Panel regression result of the relationship between social media sentiment and stock return at their first difference level. T-stats are robust with Newey-West's (1987) autocorrelation robust standard errors for any potential multicollinearity. ***, **, and * stand for the statistical significance level at 1%, 5%, and 10%, respectively. Standard errors are presented in brackets below the estimated coefficients.

	(1) Just Pos. & Neg.	(2) Pos. & Neg. with Control Variables	(3) Just Comp-level Sentiment	(4) Comp-level Sentiment with Control Variables	(5) Just Control Variables
$\Delta PosSent$	8.79e-04***	8.81e-04***			
	(20.70)	(19.63)			
$\Delta NegSent$	-9.96e-04***	-1.20e-03***			
	(-23.85)	(-26.87)			
$\Delta Company - level\ Sentiment$			5.65e-03***	6.10e-03***	
			(26.19)	(27.41)	
$\Delta Size$		-2.47e-04***		-1.30e-04***	-1.33e-04***
		(-7.18)		(-4.19)	(-4.29)
$\Delta Last\ week's\ Return$		-2.03e-02***		-1.98e-02***	-1.85e-02***
		(-24.45)		(-23.89)	(-22.32)
$\Delta Abnormal\ Turnover$		-3.76e-04***		-2.89e-04**	-2.40e-04***
		(-3.83)		(-2.95)	(-2.45)
$\Delta Park\ Volatility$		8.83e-03***		1.02e-02***	9.88e-03***
		(6.15)		(7.23)	(7.02)
$\Delta Amihud\ Illiquidity$		3.22e-2**		3.07e-2**	3.01e-2**
		(2.01)		(1.99)	(1.97)
Constant	-1.35e-06**	-2.38e-06**	-1.20e-06**	-2.19e-06**	-2.47e-06**
	(-1.99)	(-2.22)	(-1.98)	(-2.16)	(-2.33)
F-value	342.086***	251.72***	685.848***	287.471***	172.006***
Adj. R-SQUARED	1.88%	23.91%	2.02%	21.88%	19.11%
$\|\Delta NegSent\|$ minus $\|\Delta PosSent\|$	1.87 e-4***	2.08 e-4***			
	(25.96)	(27.88)			

Table A.6 - Granger Causality test between Stock Return and Social Media Sentiment

Table A.6 presents the Granger causality test for the relationship between social media sentiment and stock return. The results are divided into three significance levels, and the number of firms with significant causality statistics is counted based on these three significance levels.

Significance Levels	At 10 percent	At 5 percent	At 1 percent
Number of firms whose sentiment causes the stock return	427	376	268
Number of firms whose stock return causes the sentiment	135	85	31
Number of common firms between the two groups	129	81	28