

https://dspace.org/

A tale of tweets: the influence of Twitter on the technology sector

Eltringham, Bradley

2022

https://hdl.handle.net/10539/37280

Downloaded from a DSpace repository. https://dspace.org/

School of Economics and Finance
Faculty of Commerce, Law and Management

A tale of tweets: The influence of Twitter on the technology sector.

Master of Commerce in Finance

Bradley Eltringham **[1604611]**



Supervisor: Prof. Yudhvir Seetharam
School of Economics and Finance

DECLARATION

I, Bradley Eltringahm, declare that this research report is my own unaided work. It is submitted in fulfilment of the requirements for the degree of Master of Commerce (M.Com) in Finance at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

ACKNOWLEDGEMENTS

I would like to acknowledge all the special people who made it possible for me to reach this point. In particular, I would like to express my sincerest gratitude to the following people:

- To my parents, for their selfless support, and enabling me to pursue a path I am passionate about. I would also like to thank them for constantly providing the encouragement I needed to push the boundaries of what I thought I was capable of.

- To my brother, for lending an ear to listen and for his invaluable advice. Thank you for always steering me in the right direction and for your unwavering support. You are my primary source of motivation and the reason I decided to embark on this journey.

- To my entire family for their understanding and boundless love, and for supporting me in all of my efforts.

- To Prof. Yudhvir Seetharam, for his constant guidance, patience and swift communication throughout the development of this study, despite his many commitments. May many other students be given the privilege to experience his unapologetic passion for research.

- To peers who have turned into friends, for the memories we shared together over the duration of this master's program.

- Last but not least, I would like to thank all the lectures on the program for their unwavering commitment to teaching and imparting their knowledge for the benefit of those who are willing to learn and eager to better themselves

ABSTRACT

Investor sentiment has emerged to be a very topical subject within the context of behavioural finance. In more recent times textual analysis has emerged to be one such way of attempting to quantify investor sentiment. This study utilises textual sentiment and attempts to examine the predictive power of social media networks with regard to the technology sector during the course of unprecedented market wide volatility as a consequence of a global pandemic. More specifically, this study analyses the association between *tweet* features (bullishness/sentiment , message volume and overall investor agreement) and market features (daily trading volume, volatility and raw returns, closing price, intraday high, intraday low) with regard to the 10 current largest technology sector firms by market capitalisation. The study spans over the period of February 2016 to December 2021, which encompasses the COVID-19 pandemic. The first finding of this study indicates that most of the *tweet* features are observed to be contemporaneously associated to the stock returns of the ten biggest technology firms by market capitalisation. Second, the results show the that for the most part there is no monotonic relationship between *tweet* features and technology sector stock market returns with exception of the relationship between the magnitude of message agreement and stock returns. Third, there is no evidence that the past values of *tweet* features contain any useful information that could be used to predict future stock returns. Finally, by comparing the pre-COVID data and intra-COVID data, it is noted that for the most part there is no monotonic relationship between *tweet* features and technology sector stock market returns. There is however evidence that the use of increased usage of Twitter as an investment tool following an exogenous shock to the market.

Table of Contents

ACKNOWLEDGEMENTS **3**

ABSTRACT **4**

TABLE OF CONTENTS **5**

1. INTRODUCTION **9**

1.1. Background and overview 9
1.2. Motivation of the study 11
1.3. Knowledge gap 11
1.4. Research objective 11
1.5. Problem statement 12
1.6. Hypotheses 12
1.6.1. Primary Hypothesis *12*
1.6.2. Secondary Hypothesis *12*
1.6.3. Tertiary hypothesis *13*
1.7. Potential benefits of the study 13
1.8. Organisation 13

2. LITERATURE REVIEW **15**

2.1.1. The history and development of the Efficient Market Hypothesis *15*
2.1.2. What does the EMH entail? *16*
2.1.3. Are financial markets efficient? *16*
2.2. The role of behavioural finance 17
2.3. The amalgamation of two concepts- the Adaptive Market Hypothesis 20
2.4. An overview of textual sentiment 21
2.5. The role of social media 22
2.5.1. Twitter *24*
2.6. The change in investor behaviour preceding and during the COVID-19 pandemic 25
2.7. The impact of COVID-19 on social media 27
2.8. Summary of literature review 28

3. DESCRIPTION OF DATA AND RESEARCH METHODOLOGY **32**

3.1. Data 32
3.2. Variables 33
3.2.1. Tweet Features *33*
3.2.2. Market Features *35*
3.3. Econometric model estimation 36
3.3.1. The relationship between tweet features and stock market features *36*
3.3.2. Pooled quantile regressions *37*
3.4. Granger Non-Causality Analysis 39
3.5. Robustness checks 40
3.6. Limitations 41

4. EMPIRICAL ANALYSES **42**

4.1. Overall sample 42
4.1.1. Descriptive statistics *42*
4.1.2. The correlation between tweet features and market features *45*
4.1.3. The contemporaneous association between market and tweet features *46*
4.1.4. The association between the size of stock returns and that of tweet features *50*
4.1.5. The relationship between the past value of tweet features and stock returns *59*
4.2. Comparison of Pre-COVID and Intra-COVID data 62
4.2.1. Comparison of descriptive statistics *62*

4.2.2. Comparison of the contemporaneous association between market and tweet features before and during the COVID-19 pandemic ... *66*
4.2.3. Comparison of the association between the size of stock returns and that of tweet features before and during the COVID-19 pandemic ... *69*
4.2.4. Comparison of the relationship between the past value of tweet features and stock returns before and during the COVID-19 pandemic. ... *83*

5. CONCLUSION ... **87**

6. REFERENCES ... **91**

7. APPENDIX A ... **98**

7.1. FULL SAMPLE ROBUSTNESS CHECKS ... 98
7.1.1.1. Quantile regressions using alternative conditional quantiles ... *98*

8. APPENDIX B ... **102**

8.1. PRE-COVID AND INTRA-COVID ROBUSTNESS CHECKS ... 102
8.1.1. Pre-COVID and intra-COVID quantile regressions using alternative conditional quantiles ... *102*

FIGURE 1: NUMBER OF SOCIAL MEDIA USERS WORLDWIDE ACCORDING TO STATISTA 23
FIGURE 2: AVERAGE MONTHLY TIME SPENT WITH LEADING SOCIAL PLATFORMS IN THE UNITED STATES FROM 2018 TO 2020 (IN MINUTES) ACCORDING TO STATISTA 28
FIGURE 3: DAILY DISTRIBUTION OF AGGREGATE TWEETS 43
FIGURE 4: VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO BULLISHNESS AT A 5% LEVEL OF SIGNIFICANCE 52
FIGURE 5: VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO MESSAGE VOLUME AT A 5% LEVEL OF SIGNIFICANCE 54
FIGURE 6:VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO THE AGREEMENT INDEX AT A 5% LEVEL OF SIGNIFICANCE 56
FIGURE 7:VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO MARKET RETURN AT A 5% LEVEL OF SIGNIFICANCE 58
FIGURE 8: DISTRIBUTION OF AGGREGATE RETURNS (PRE-COVID SAMPLE) 64
FIGURE9: DISTRIBUTION OF AGGREGATE RETURNS (INTRA-COVID SAMPLE) 65
FIGURE 10:VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO PRE-COVID BULLISHNESS AT A 5% LEVEL OF SIGNIFICANCE 72
FIGURE 11: VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO COVID BULLISHNESS AT A 5% LEVEL OF SIGNIFICANCE 73
FIGURE 12:VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO PRE-COVID MESSAGE VOLUME AT A 5% LEVEL OF SIGNIFICANCE 75
FIGURE 13:VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO COVID MESSAGE VOLUME AT A 5% LEVEL OF SIGNIFICANCE 76
FIGURE 14:VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO THE PRE-COVID AGREEMENT INDEX AT A 5% LEVEL OF SIGNIFICANCE 78
FIGURE 15: VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO THE COVID AGREEMENT INDEX AT A 5% LEVEL OF SIGNIFICANCE 79
FIGURE 16: VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO PRE-COVID MARKET RETURN AT A 5% LEVEL OF SIGNIFICANCE 81
FIGURE 17: VISUALISATION OF QR AND OLS ESTIMATES WITH REGARD TO COVID MARKET RETURN AT A 5% LEVEL OF SIGNIFICANCE 82
FIGURE 18: VISUALISATIONS OF FULL SAMPLE QUANTILE ESTIMATORS (T= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] 100
FIGURE 19:VISUALISATIONS OF PRE-COVID SAMPLE QUANTILE ESTIMATORS (T= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9]) 105
FIGURE 20:VISUALISATIONS OF PRE-COVID SAMPLE QUANTILE ESTIMATORS (T= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9]) 106

TABLE 1: SUMMARY STATISTICS OF THE VARIABLES (FULL SAMPLE) 42
TABLE 2: PAIRWISE CORRELATION MATRIX (FULL SAMPLE) 45
TABLE 3: HAUSMAN TEST (FULL SAMPLE) 47
TABLE 4: CONTEMPORANEOUS OLS REGRESSIONS (FULL SAMPLE) 47
TABLE 5: POOLED QUANTILE REGRESSION ESTIMATES (FULL SAMPLE) 51
TABLE 6: SUMMARY OF PANEL UNIT ROOT TESTS (FULL SAMPLE) 60
TABLE 7: SUMMARY OF GRANGER NON-CAUSALITY TESTS (FULL SAMPLE) 61
TABLE 8: SUMMARY STATISTICS OF THE VARIABLES (PRE-COVID SAMPLE) 62
TABLE 9: SUMMARY STATISTICS OF THE VARIABLES (INTRA-COVID SAMPLE) 63
TABLE 10: CONTEMPORANEOUS OLS REGRESSIONS (PRE-COVID SAMPLE) 66
TABLE 11: CONTEMPORANEOUS OLS REGRESSIONS (COVID SAMPLE) 67
TABLE 12: POOLED QUANTILE REGRESSION ESTIMATES (PRE-COVID SAMPLE) 69
TABLE 13:POOLED QUANTILE REGRESSION ESTIMATES (COVID SAMPLE) 70
TABLE 14: SUMMARY OF GRANGER NON-CAUSALITY TESTS (PRE-COVID SAMPLE) 83
TABLE 15: SUMMARY OF GRANGER NON-CAUSALITY TESTS (COVID SAMPLE) 84
TABLE 16: POOLED QUANTILE REGRESSION ESTIMATES (FULL SAMPLE) WITH T= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] 99

TABLE 17: POOLED QUANTILE REGRESSION ESTIMATES (FULL SAMPLE) WITH T=∈[0.05,0.2,0.4,0.6,0.8,0.95] .. 101
TABLE 18: POOLED QUANTILE REGRESSION ESTIMATES (PRE-COVID SAMPLE) WITH T= ∈[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] 103
TABLE 19: POOLED QUANTILE REGRESSION ESTIMATES (COVID SAMPLE) WITH T= ∈[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] 104
TABLE 20: POOLED QUANTILE REGRESSION ESTIMATES (PRE-COVID SAMPLE) WITH T=∈[0.05,0.2,0.4,0.6,0.8,0.95]..... 107
TABLE 21: POOLED QUANTILE REGRESSION ESTIMATES (COVID SAMPLE) WITH T=∈[0.05,0.2,0.4,0.6,0.8,0.95] 108

1. Introduction

1.1. Background and overview

In the past century a topic of concern that has beleaguered participants of the financial world has been the ability of financiers to predict stock price movements. Numerous academics such as Cowles (1933), Kendall (1953) and finally, Fama (1965b, 1970) find evidence to support the idea that stock prices follow a random walk and can therefore not be predicted. According to the Efficient Market Hypothesis (EMH) postulated by Fama (1970), if all information is reflected in the stock price, neither technical analysis[1] nor, fundamental analysis[2] can generate consistent alpha (risk adjusted excess returns). Although widely accepted, the EMH received a lot of criticism due its inability to explain deviations in prices from their fair values.

In an attempt to explain what the EMH could not, academics and practitioners in the fields of finance and psychology, the most notable being that of Kahneman and Tversky (1974, 1979, 1989), Thaler (1985,1999) and De Bondt and Thaler (1985), laid the foundation for the ideology known today as behavioural finance. This new school of thought noted a crucial element inherent in financial markets that the EMH tends to overlook - this being the fact that market participants are human and as such their decisions are based on emotions and biases; and are occasionally irrational. A component of behavioural finance that has attracted significant amounts of attention in recent years is the influence that sentiment has on the stock market. Kearney and Liu (2014) differentiate between two forms of sentiment that each play fundamental roles in driving the stock market: investor sentiment and textual sentiment. Investor sentiment relates to the expectations of market participants with regard to both the expected return and risk when making financial decisions (De Long et al., 1990). In contrast, textual sentiment differs from investor sentiment in that it is concerned with the disposition of text-based posts with regard to financial markets.

[1] Technical analysis can be briefly defined as a trading discipline that is underpinned by the analysis of statistical trends which are based on historical trading activity.

[2] Fundamental analysis can be briefly defined as the measurement of a securities intrinsic value by analysing both economic and financial factors.

It is innocuous to assume that new information plays a pivotal role in influencing the dynamics of the stock market. In fact, information is the primary factor in shaping market perception and therefore overall market investment. Prior to the rise of the internet, dissemination of new financial information to market participants took a relatively long time. With the introduction of microblogging services such as Twitter, that at their very core, are driven by new information (Cropper 2011), people gained access to increasing amounts of information at rates never before observed in the finance world, thus the prolonged dissemination of information, once inherent in financial markets, has begun to disappear. Social media affords people the ability to share ideas, information and most important for the context of this study, their opinions with other people or communities across the globe simultaneously. The works of Bollen, Mao and Zeng (2011), Mao, Counts and Bollen (2011), Gilbert and Karahalios (2010) and Sprenger and Welpe (2010, 2014) suggest that sentiment of the general public and market participants on social media platforms is highly influential at a micro-economic level. Wysocki (1998) amongst other findings notes that changes in overnight posting volume on internet stock message boards is related with next day trading volume and abnormal stock returns. Blankespoor, Miller and White (2014) establish that there is an increase in dissemination of firm news via Twitter and this increase is associated with a reduction in information asymmetry and an increase in trading volumes.

With this in mind, it is worth noting that over the last 20 years, the global economy has been subject to significant changes comparable to those that occurred during the period of the Industrial Revolution . This is primarily due to the profound technological revolution that began at the turn of the millennium and still continues. Innovations in both the digital computing and automation realm have triggered tectonic shifts in consumer and business behaviour across economies around the world. The firms responsible for this innovation form part of the technology sector and are typically focused on developing and producing technology for the rest of the economy. The technology sector is an area of specific focus in the context of this study due to the attractive returns and rapid growth present within the last two years. This growth is thought to be aided in part by the discovery of the COVID-19. The pandemic is thought to have acted as a catalyst, accelerating digital transformation, which in turn propelled many organisations into the future. Work environments changed drastically as remote work, a once foreign concept, became a reality and as such market demands evolved. In addition to this, the COVID-19 pandemic affords a unique

opportunity to observe how the global financial markets responded to an exogenous shock of this magnitude.

Although academic literature on the subject of textual analysis is broad, the number of studies that focus on specific sectors of the financial markets is relatively scarce. This study aims to address this by focusing on the effect of specific aspects of social media on the stock market features of the largest publicly listed companies within the technology sector. In addition to this, the influence of social media on the technology sector will be analysed utilising both pre-pandemic and intra-pandemic data.

1.2. Motivation of the study

This study aims to fill the gap in academic literature that focusses on the predictive power of social media networks with regards to the technology sector during the course of a worldwide pandemic which brought about unprecedented market wide volatility.

1.3. Knowledge gap

Investor sentiment has emerged to be a very topical subject within the context of behavioural finance. Whilst difficult to observe there have been many attempts to not only identify suitable proxies but also to quantify the effect of sentiment on asset markets. In more recent times textual analysis has emerged to be one such way of attempting to quantify investor sentiment. Several studies have focused on the ability of textual sentiment analysis to predict stock returns in different markets however, both, the area of research focusing on sector specific analysis and a pre- and intra- pandemic analysis of stock market returns and those focussing on the technology sector in this space is relatively scarce.

1.4. Research objective

The primary objective of this study is to examine the ability of textual analysis of Twitter posts in predicting returns related to the technology sector on a global scale . Furthermore, this study aims to provide commentary on the use of social media as an investment tool during a period of unrivalled volatility across markets which was brought about by the coronavirus disease (COVID-19).

1.5. Problem statement

In the same vein as studies attempting to corroborate the use of fundamental and technical analysis in predicting stock market movements, textual analysis has received a great amount of attention. The results from such studies vary which makes the usefulness of these analyses questionable. This study attempts to address this problem by providing reliable, well-defined results as it examines the relationship between various tweet features and market features not only at a global level but at a market specific level. The widespread adoption of social media platforms by firms as a means of communication calls for additional academic commentary. This study aims to add to existing literature by providing an analysis of specific tweet features and their relationship with stock market features before and during the COVID-19 pandemic.

1.6. Hypotheses

1.6.1. Primary Hypothesis

The primary hypothesis of this study is to determine if there is a relationship between the past value of tweet features and the future value(s) of stock returns within the technology sector:

$H_{0,\,A}$: *Past values of tweet features contain statistically significant information that could be used to predict future returns*

$H_{1,\,A}$: *Past values of tweet features contain no statistically significant information that could be used to predict future returns.*

1.6.2. Secondary Hypothesis

The secondary hypothesis of this study is to examine if the there is a monotonic association between the size of market returns and *tweet* features within the technology sector:

$H_{0,B}$: *The size of firm-level tweet features is monotonically related to the size of market features (stock returns, volatility and trading volume) within the technology sector.*

$H_{1,B}$: : *The size of firm-level tweet features is not monotonically related to the size of market features (stock returns, volatility or trading volume within the technology sector).*

1.6.3. Tertiary hypothesis

The tertiary hypothesis of this study is to identify the pervasiveness of Twitter as an investment tool across during times of an exogenous shock to the market such as the introduction of the COVID- 19 pandemic:

$H_{0,C}$: *Twitter is an investment tool that is utilised to a greater extent during times in which an exogeneous shock is introduced to the market.*

$H_{1,C}$: *Twitter is not an investment tool that is utilised to a greater extent during times in which an exogeneous shock is introduced to the market.*

1.7. Potential benefits of the study

The body of literature in this area of research is at best inconclusive and is relatively scarce. Therefore, the focus of this study may shed additional light on the topic surrounding the role of investor sentiment indicators in the technology sector, more specifically, quantifying the ability of social media as a sentiment indicator of the technology sector around the world, thus providing a basis for future academic research on the subject. This research study seeks to provide sound empirical evidence to enable a better understanding of the dynamics of the influence of social networks as an indicator within the stock market. In doing so, this study may provide newfound knowledge to industry professional and academics alike.

1.8. Organisation

Chapter 1 offered an introduction and background to the importance of social media platforms such as Twitter and their value in providing metrics that could be used to explain movements in financial markets. Chapter 2 will discuss the key concepts as well as provide a review of the

existing literature on the theoretical framework underpinning this study as well as empirical studies showing the role played by investor sentiment, particularly from online texts, on the stock market. This chapter also includes a survey of methodological issues surrounding textual analysis. This is followed by Chapter 3 will provide an overview of the research methodology that encompasses the theoretical framework, research methodology and data. Chapter 4 presents the empirical results from the econometric and statistical specifications as well as various robustness checks to establish if the results are sensitive to methodological considerations. The final chapter, Chapter 5, summarises the findings and presents the pertinent conclusions of the study, outlines the limitations of the current research and make recommendations for further research.

2. Literature review

2.1.1. The history and development of the Efficient Market Hypothesis

It is important to understand the basis on which the Efficient Market Hypothesis (EMH) is formed and to do so, one must be cognisant of the history that underlies such a concept. The earliest forms of evidence in support of efficient markets dates back as early as the 18th century. This early evidence presented the idea of a random walk. In statistics a random walk is defined as the path taken by a point or quantity that moves in steps where the direction of each step is determined randomly. The idea of random walk was based on Robert Browns observation that grains of pollen suspended in water had a rapid oscillatory (random) motion when viewed under a microscope (Brown, 1828). In 1863, a French stockbroker by the name Jules Regnault, observed that the longer you hold a security, the more you can win or lose on its price variations, that is, the price deviation is directly proportional to the square root of time (Regnault, 1863). Although his work went unnoticed for several years Louis Bachelier, another Frenchman whose Ph.D. dissertation titled "The Theory of Speculation" (Bachelier 1900) introduced the first mathematical model of Brownian motion and its use for valuing stock prices. This was the first time in history advanced mathematics was used in the context of finance. In 1933 Alfred Cowles, an American economist, examined the performance of investment professionals and reached the conclusion that stock market forecasters cannot predict asset prices accurately (Cowles, 1933). The theory that stock prices move randomly was officially proposed by Maurice Kendall in his 1953 paper, “The Analytics of Economic Time Series, Part 1: Price” (Kendall, 1953). In 1965 Paul Anthony Samuelson finds proof that anticipated prices fluctuate randomly. Similarly, in 1965 Fama defines an “efficient” market for the first time, linking the concept to the theory of random walk. Finally, in 1970 Fama presented a paper titled “Efficient Capital Markets: A Review of Theory and Empirical Work” whereby the EMH was formalised and presented in the final form that is familiar today.

2.1.2. What does the EMH entail?

The theory of efficient markets is concerned with the ability of prices to "fully reflect" available information (Fama, 1970). Fama (1970) corroborates this definition of efficient markets into three levels of "efficiency" being that of weak form efficiency, semi-strong form efficiency and finally strong form efficiency. Weak form efficiency refers to the ability of a stock price to reflect all past information, thus future stock prices cannot be predicted given current or past pricing information. With this being said, the weak form of market efficiency alludes to the idea that technical analysis is fruitless. The semi-strong form of the EMH claims that prices encompass all publicly available information, thus attempting to beat the market using technical or fundamental analysis will not yield any reward. Finally, the strong form of the EMH posits that all information, both public and private is accounted for in the current stock price, thus no additional information can be advantageous for investors.

2.1.3. Are financial markets efficient?

After its creation in the 1960s, the EMH was viewed as one of the greatest theoretical and empirical findings in world of finance. In the decade following this innovation, hordes of academics and finance professionals alike conceptualized reasons as to why the EMH holds (Shleifer, 2000). Towards the end of the 1970s however, people began to challenge the foundation on which the EMH was based. Theoretical critiques included the fact that it is naïve to assume that investors always act rationally, in fact many investors react to immaterial information in forming their portfolios. Black (1986) stated that investors are likely to trade on noise rather than information. Another pitfall of the EMH was its inability to reasonably explain disruptions within financial markets. Throughout the 1980s and early 1990s academic literature set out with the focus of explaining these disruptions by arguing that investors, being human, have emotions which in turn have an influence on the decisions they make. These decisions are very often inefficient and irrational and could be the reason as to why disasters or disruptions in the stock market exist (Kapoor & Prosad, 2017). This new form of reasoning paved the path for what has become known today as behavioural finance.

2.2. The role of behavioural finance

2.2.1. Imperative findings

In the early 1980s the concept of behavioural finance started to gain popularity following the work of Kahneman and Tversky (1974, 1979) and that of Thaler (1985); however, this was not the first time such a concept had been explored. Seldon (1912) was the first to uncover and explore the psychological facet of finance in a book titled *Psychology of the Stock Market.* The book was based upon the belief that the movements of prices on the exchanges correlated to the to the mental attitude of those that participate in the market (Sewell, 2001). Unlike the traditional finance theory, behavioural finance relaxes the assumption of rationality and opts for an explanation that entails that real investor are influenced by their psychological biases when making financial decisions (Kapoor & Prosad 2017). Often considered the forefathers of behavioural finance are the prominent psychologists, Kahneman and Tversky who, in 1974 postulated that heuristics and biases affect decisions under uncertainty. Heuristics can be thought of as patterns regarding how people behave (Shefrin, 2001). Another definition of a heuristic is any approach to a problem that utilises a practical technique that is not guaranteed to be rational nor optimal, they are usually experience-based techniques that can viewed as "rules-of-thumb" or "short cuts" (Konstantinidis, *et al.*, 2012). Five years later, Kahneman and Tversky (1974) provide empirical evidence that illustrates that the principles of expected utility theory do not necessarily hold when making decisions under risk. As such, by combining concepts from both psychology and economics, they introduced an alternative decision-making model, known as Prospect Theory. Prospect Theory suggests that individuals value gains and losses differently and that the dissatisfaction from losing a sum of money is greater than the satisfaction derived from gaining the same amount. In addition to these findings, Prospect Theory argues that are risk averse for gains and risk seeking for losses. Kahneman and Frederick (2002), suggest that cognitive heuristics work via a process called attribute substitution which happens without conscious awareness. The most common cognitive heuristics that provide evidence of the irrational behaviour of investors are:

Herding: This behaviour is seen to occur when an investor follows the example of other investors or a group of investors as opposed to relying on their own private information, that may suggest doing something quite different, when making financial decisions (Banerjee 1992).

Overconfidence: De Bondt and Thaler (1995) state that evidence of overconfidence is "*perhaps the most robust finding in the psychology of judgement*". In the context of finance overconfidence refers to the tendency of investors to overestimate their level of skill, abilities, or perceived information and as a result, leads to the making of incorrect decisions (Konstantinidis, *et al.*, 2012).

Representativeness: Kahneman and Tversky (1974) argue that *"when people are asked to judge the probability that object or event A belongs to class or process B, probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B".* In simpler terms, people often attempt to fit a new or unknown subset of information into one that already exists or that they are familiar with, and in doing so, they are likely to discover common features in unrelated situations.

Anchoring: This cognitive heuristic entails that, individuals base their beliefs on an initial starting point or target number, which can produce erroneous results if this starting point or "anchor" does not accurately resemble the right information or is does not fit the true value precisely.

Mental accounting: Involves the process in which an individual generates mental accounts and register events they have experienced based on their unique traits. Thaler (1999) defines mental accounting as "the set of cognitive operations used by individuals and households to organize, evaluate, and keep track of financial activities."

Loss aversion: In their 1989 paper, Kahneman and Tversky propose that the more a person experiences loss, the more likely they are to become prone to loss aversion (Kahneman and Tversky, 1989). In short, loss aversion simply refers to the tendency of investors to focus on attempting to avoid losses more so than making gains.

Regret aversion: Landman (1993) defines regret as "*a more or less painful cognitive and emotional state of feeling sorry for misfortunes, limitations, losses, transgressions, shortcomings or mistakes"*. Regret aversion entails that an investors' fear of regret, that is the pain of making the wrong decision, is what drives them to be risk averse or motivates them to take higher risks.

The above-mentioned cognitive heuristics emanate from both biases in beliefs and biases in preferences, which have been the focus of academics' attempts to account for the irrational decisions made by participants in the financial markets for several decades. In recent times another facet of behavioural finance has caught the attention of academics and industry professionals alike: investor sentiment.

Investor sentiment can be broadly defined as the general prevailing attitude or beliefs of market participants with regard to future cash flows and investment risks that are necessarily justified by the facts at hand (Baker and Wurgler, 2007). Building on the work of Delong, Shleifer, Waldman and Summers (1990), Baker and Wurgler (2007) note that investors who dictate price and expected return in the market fall within the scope of two broad categories: the sentiment-free rational arbitrageurs and irrational traders whose decisions are likely influenced by some form of sentiment. With that said, mispricing appears to stem from two factors; the first being a change in sentiment on the side of the irrational investors and limits to arbitrage from the rational investors, which work hand in hand in driving prices away from their fundamental values. Investor sentiment is often difficult to gauge or quantify and as such several studies have attempted to measure investor sentiment by utilising imperfect proxies.

By far the most obvious and direct way of measuring sentiment is to directly approach market participants and question them on their views or perceptions of current market conditions. The most straightforward way of gathering this information is through the use of surveys. Brown and Cliff (2002) explored the relationship between investor sentiment and long horizon returns and asset valuations. They find that if prices are driven far above their intrinsic levels due to the overoptimism of investors, these periods of increased sentiment will be followed by a subsequent decrease in returns as prices revert back to their fundamental values.

One strand of literature that was the focus of several studies in the early 2000s was the effect of investor mood on asset prices. Many of these studies investigate the effect of a continuous variable that are not necessarily directly related to financial markets but have been proven to impact mood. Such variables include that of sunshine, daylight, temperature and lunar cycles (Saunders (1993); Hirshleifer and Shumway (2003); Kamstra, Kramer and Levi (2003); and Cao and Wei (2005).

Another approach to measure mood is that of an event-based approach which has the advantage of being able to clearly identify a sudden change of mood. Examples of such events include that of disruptions to sleep patterns caused by changes to and from daylight saving and results of soccer games (Kamstra, Kramer, and Levi, 2000; Garcia and Norli, 2007).

More in line with traditional finance literature several studies have attempted to utilise specific market features as proxies for sentiment. Such features include that of dividend premium liquidity amongst others. Stocks that pay dividends are often considered less risky than non-dividend paying stocks as they provide investors with a predictable stream of income thus offering an element of comfort and safety. Fama and French (2002) note that firms are more inclined or have an increased propensity to pay dividends when they are at a premium, thus firms appear to cater toward the general market sentiment when deciding to issue dividends. Liquidity or more specifically trading volume is another feature that has emerged as an apparent proxy for sentiment in the past. Scheinkman and Xiong (2003) find that trading volume or the "depth" of the market, exposes the differences in opinion amongst market participants and are subsequently related to levels of valuation.

Overall, the literature focussing on investor sentiment is incredibly broad and offers an in-depth analysis of the topic from various angles, with new proxies and measures frequently emerging.

2.3. The amalgamation of two concepts- the Adaptive Market Hypothesis

A combined understanding of evolutionary biology present in behavioural finance and classical financial economics yields an understanding of the Efficient Markets Hypothesis from an evolutionary perspective: the Adaptive Markets Hypothesis (AMH). Lo (2004) postulates that the Adaptive Markets Hypothesis (AMH) treats anomalies not as minor deviations from an otherwise, efficient market, but rather, as part of it. It is a formulation of market efficiency that is cognisant of the fact that the market is comprised of humans with predictable fallibilities. In highly efficient markets, investors are unwilling to gather information given the low levels of economic rent that can be accrued from doing so. In markets where "noise traders" (those who believe noise to be a signal and therefore use that to trade) exist, then the smart money is motivated to gather information and trade on it. (Black, 1986). The market therefore transfers wealth away from the

noise traders to the information gatherers, and thereby, becomes more efficient. The AMH states that price efficiency is a function of the number and nature of distinct groups of market participants who are seen to behave in a common manner, as well as the environmental pressures that act on these participants. As such, when there is an abundance of profit-making opportunities present in the market, the presence of competition is scarce, thus forming an inefficient market. In contrast, when the market shows signs of dwindling profits, competition amongst market participants rises, resulting in a more efficient market environment. *Ceteris paribus;* the market cycles between varying degrees of efficiency which are in turn based on the prevalence of economic profits within the market.

Undoubtedly the neoteric school of thought, the AMH has been thoroughly examined since its inception in the early two thousands. A few examples of these studies include that of Ito and Sugiyama (2009) who note the presence of time varying market inefficiency with the context of the United States. Further, Kim et al. (2011) find that the US. stock market is more efficient following the 1980s and that market conditions are primarily responsible for the change. Uruqhart and Hudson (2013) study the dynamics of the markets in the US, UK and Japan, their results corroborate that of the AMH. Additionally, Noda (2016) finds evidence in support of the AMH on the Japanese stock market. A more recent thread of literature is the applicability of the AMH to crypto markets. Khuntia and Pattanyak (2018) find that with regard to the return predictability of the bitcoin market, market efficiency progresses over time and thus validates the AMH in this context.

2.4. An overview of textual sentiment

Textual sentiment can be broadly defined as a text-based form of sentiment which describes the emotion present in text-based posts from online platforms (Li, van Dalen & van Rees, 2018). Traditionally sentiment has been measured exclusively via a survey-based approach. The issues inherent with this type of approach include the fact that they incredibly time consuming and the response rate of such surveys are relatively poor, which in turn leads to the sample being relatively tiny. Information broadcast via internet-based forums and social media has allowed researchers to augment this process by availing volumes of information that were previously unattainable thus

improving research in the social and computer science fields. (Mao, Counts and Bollen ,2011). Textual sentiment is a form of computational analysis of sentiment and as such may offer more increased accuracy and be both cost and time effective when compared to traditional methods (Mao, Counts and Bollen ,2011). Numerous studies have attempted to not only to measure textual sentiment but to unearth the relationship between this sentiment and features of the stock market. Beginning with Wuthrich et al., (1998), they noted that was possible to profit from fresh news using information contained in articles published on the internet. Antweiler and Frank (2004) find that posts on stock message boards aided in predicting market volatility. Schumacher and Chen (2009) noted that when comparing their machine learning model which used article terms and real time stock prices at the point articles were released performed far better than a rudimentary linear regression model in predicting stock prices in the following 20 minutes. Further, Mao, Counts and Bollen (2011) find that weekly volumes of Google Insights Search of financial terms have a significant correlation with Dow Jones Industrial Average closing prices.

2.5. The role of social media

Before the influence of social media can be explored, it is crucial to understand what exactly social media is. One definition is that social media are *"web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system"* (Boyd & Ellison, 2008, p. 211). In simpler terms, McIntyre (2014) defines social media as *"a form of computer-mediated communication"*.

The fourth industrial revolution has no doubt led to the widespread acceptance and utilisation of internet-based communication forums and social media platforms around the world. According to Statista the number of social media users has increased from 2.86 billion users in 2017 to approximately 3.6 billion in 2020 (an increase of almost 126 percent). In addition to these statistics, the number of social media users is forecast to increase to roughly 4.5 billion by 2025.

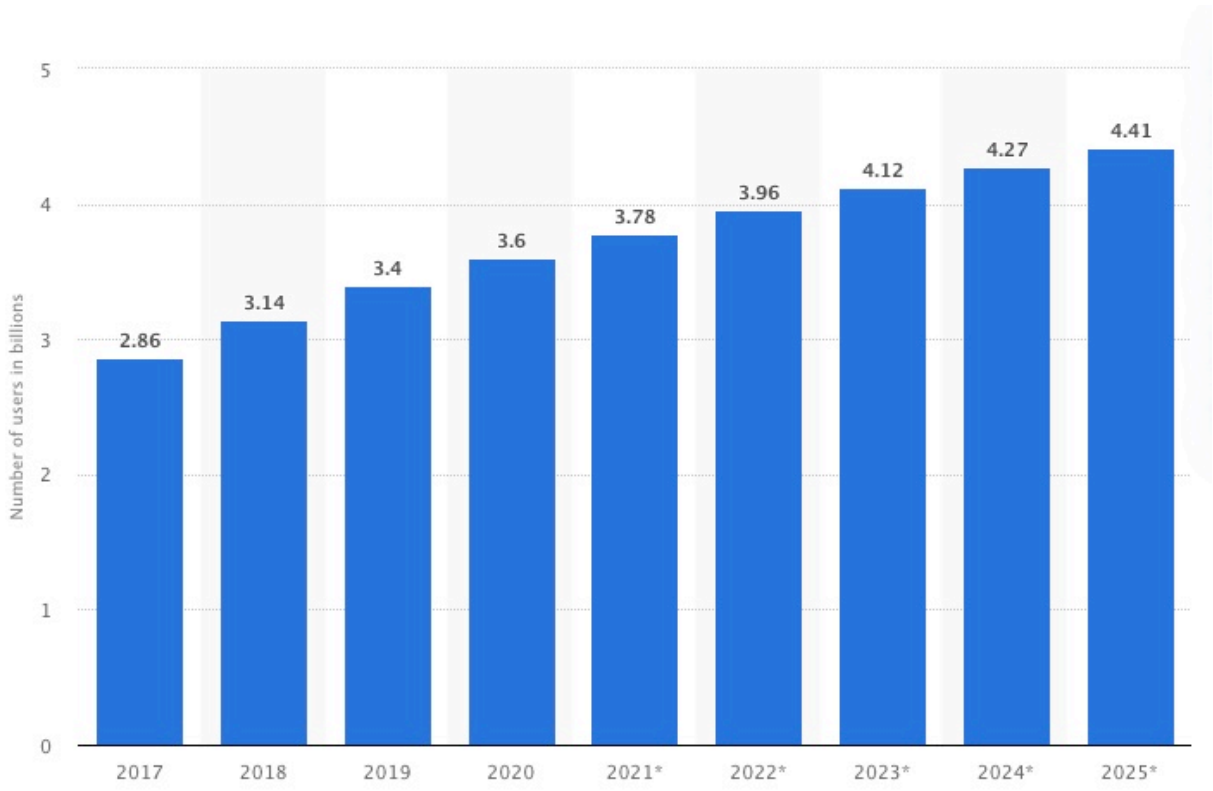


Figure 1: Number of social media users worldwide according to Statista

Note: * *Indicates estimates for the years 2021-2025*

Traditional finance literature such as Fama (1970) postulates that an increase in the rate at which information is made available and an increase in quantity of said information will make markets more efficient. Thus, it is logical to argue that the increase in the number of social media users and therefore an increase in the amount of information available, will enhance market efficiency as information should spread more rapidly compared to traditional news platforms. To support these claims Blankespoor, Miller and White (2014) find that the additional dissemination of firm initiated news via the Twitter platform is coherent with a decline in information asymmetry.

It is worth noting however, that this increased access to information may also assist in creating profit opportunities for investors by encouraging emotional responses to the news and being susceptible to noise trading.

2.5.1. Twitter

Twitter is a social media platform that was established in 2006. Twitter is defined as a service for friends, family, and colleagues to communicate and stay in touch via the exchange of quick, frequent messages (Twitter, 2021). The platform enables users to share information in real time using 280-character messages colloquially known as "tweets". The 280-character limit entails these tweets are usually concise and straightforward. Through its evolution, Twitter has become one the primary social media platforms on which firms decide to publish information for public consumption (Paul, 2015).

In an early study, Sprenger and Welpe (2010) analysed over 100, 000 tweets containing references to either a member of parliament or a political party to investigate, the ability of Twitter to accurately replicate offline political sentiment and election results. The authors' findings indicate that the overall disposition present in tweets plausibly reflects the offline political landscape. In corroboration with these findings, Goel and Mittal (2012) find the prevailing mood of the public can indeed be computed from large scale Twitter feeds using basic textual analysis.

Preliminary studies that analyse the textual sentiment present in Twitter postings include that of Bollen, Mao and Zeng (2011). The study investigates if public mood, which is measured from a large-scale collection of tweets posted on the Twitter platform is predictive of movements on the Dow Jones Industrial Average index (DJIA). The authors find that changes in public mood can in fact predict changes in the closing prices of the DJIA. Similarly, Mao, Counts and Bollen (2011) comment on the predictive ability of several types of online data sets with regard to financial market indices. Through their analysis, they find that the textual sentiment gathered from Twitter data is seen to be a statistically significant predictor of daily market log return. Gilbert and Karahalios (2010) show that the prevailing mood of the public, measured through the linguistic analysis of Twitter posts can be seen as indicator in anticipating changes on the S&P 500 index. Additonally, Sprenger and Welpe (2014) find the sentiment and volume of tweets to be related to abnormal returns and next day trading volume.

2.6. The change in investor behaviour preceding and during the COVID-19 pandemic

Prior to the spread of the Covid-19 pandemic in December of 2019, the financial markets had been relatively stable with little to no major forms of market disruption since the financial crisis of 2007/2008. Baker, Bloom, Davis, Kost, Sammon and Viratyosin (2020a) use text-based methods to analyse the significant daily stock market moves from 1900 to 2020. Their findings indicate that previous pandemics, including the Spanish flu have not possessed the ability to disrupt the stock market to the extent COVID-19. This result was likely caused by the government restrictions on commercial activity and voluntary social distancing, operating with powerful effects particularly in service-oriented economies.

The outbreak of the COVID-19 was an exogenous shock that was and is still seen to have drastic effects on not only financial markets and economies around the world, but the lives of every individual on the planet. The pandemic ultimately changed the perception of both retail and institutional investors. Ortman, Pelster and Wengerek (2020) state that the outbreak of the COVID-19 pandemic can be compared to that of a terrorist attack. In the aftermath of such an event, investor behaviour is associated with more risk averse choices. It is likely that there will be a reduced trading intensity and reduction in the flow toward risky assets. Levy and Galili (2005) find that terror is seen to have significant adverse effects on actual trade within the financial markets which can be explained by means of simple psychology. Amongst these explanations is the fact that terrorism is typically associated with feelings of deep uncertainty and ambiguity. Following this, there is a large body of theoretical research that presents findings of reluctance to trade when there is a significant amount of ambiguity present in financial markets (Dow and Werlang (1992), Epstein and Wang (1994), Murkeji and Tallon (2001)). Acts of terror are also associated with a spike in public fear. Slovic (2000) notes that terror is documented to instigate fear through its effect on "dread risk" and unknown risk. In the context of this study, dread risk can be loosely defined as a low probability, high risk event that induces a perceived lack of control and potentially fatal consequences. Unknown risk is defined as risk that is unfamiliar or new to an individual. Although unknown risk can be mitigated through an increase in the dissemination of information (which is essentially what social media provides for), dread risk is more difficult to control and therefore prevents individuals from overcoming their fears. When considering how investors participate in the market, one has to be cognisant of the fact that, ultimately, this depends

on the investors' appetite for risk. Raghunathan and Pham (1999) show that anxious subjects strictly prefer low risk, low reward options - investors are seen to be more risk averse in the event of a terrorist attack. Levy and Galili (2005) note that acts of terror are seen to subvert individuals' sense of safety and security and thus increase the likelihood of anxiety and depression. When individuals are in a depressive state their ability to consider, analyse and perform forward looking operations (as is the case when trading) is heavily impaired.

Following the subsequent increase in uncertainty brought about with the introduction of the pandemic, the market experienced significant declines and increased financial risk around the globe. Corroborating this point, Zhang, Hu and Ji (2020) find that the pandemic had a significant influence on stock markets. They noted that the risk levels of all countries in their data sample increased dramatically, further they attribute some of this increased volatility to market sentiment which is thought to be "amplified" through social media channels. Similarly, Baker, Bloom, Davis, Kost, Sammon and Viratyosin (2020a) state that no previous infectious disease (in their sample which spans 120 years) has impacted the financial markets to the extent of COVID-19. Other periods of infectious disease have seen large declines or increases in equity markets over periods of several weeks or months, the COVID-19 pandemic stands out for its extremely high frequency of swings in the market. In their 2020 study, Ortman, Pelster and Wengerek find that ,in contrast to acts of terrorism, average weekly trading intensity actually increased by 13.9%, as the number of positive COVID-19 cases began to rise rapidly.

This evidence leads one to question why the COVID-19 pandemic had such a powerful effect on markets in comparison to other infectious diseases. The first answer to this question attributes itself to the severity of COVID-19 which is a function of the apparent ease of transmission and the mortality rate among those who have been in the unfortunate position of contracting it. A second possible answer and one that aligns itself with the context of this study lies within the fact that information about pandemics is richer and diffuses much more rapidly in comparison to previous pandemics, thanks advancements in technology. With this explanation in mind, the impact of the COVID-19 pandemic is more likely to trigger stock market jumps and high stock market volatility more frequently than what has been observed in prior times. A third explanation relies on the role of cross-border flows of goods in the modern economic times, driven by decades of falling

transport costs and falling communication costs. These advancements have led to a reliance on relatively enormous geographical supply chains, which are very susceptible to sudden supply disruptions. Thus the ability of the pandemic disrupt these supply chains leads to major market disruptions especially within service-oriented economies.

2.7. The impact of COVID-19 on social media

As previously mentioned, social media plays a vital role when it comes to the dissemination of information in today's society. With the outbreak of the COVID-19 pandemic, and all the measures that were implemented to reduce the transmission of the virus (such as social distancing and working from home) individuals found themselves in a unique situation in which they were afforded more time in isolation and as such they turned to social media in an effort to communicate and alleviate boredom induced from isolation. Figure 2 below indicates the increase in time spent on social media over the period 2018-2020. As can be seen, all forms of social media experienced a significant increase in usage time over the sample period, none more so than Twitter which experienced a dramatic increase in usage time of approximately 68%.

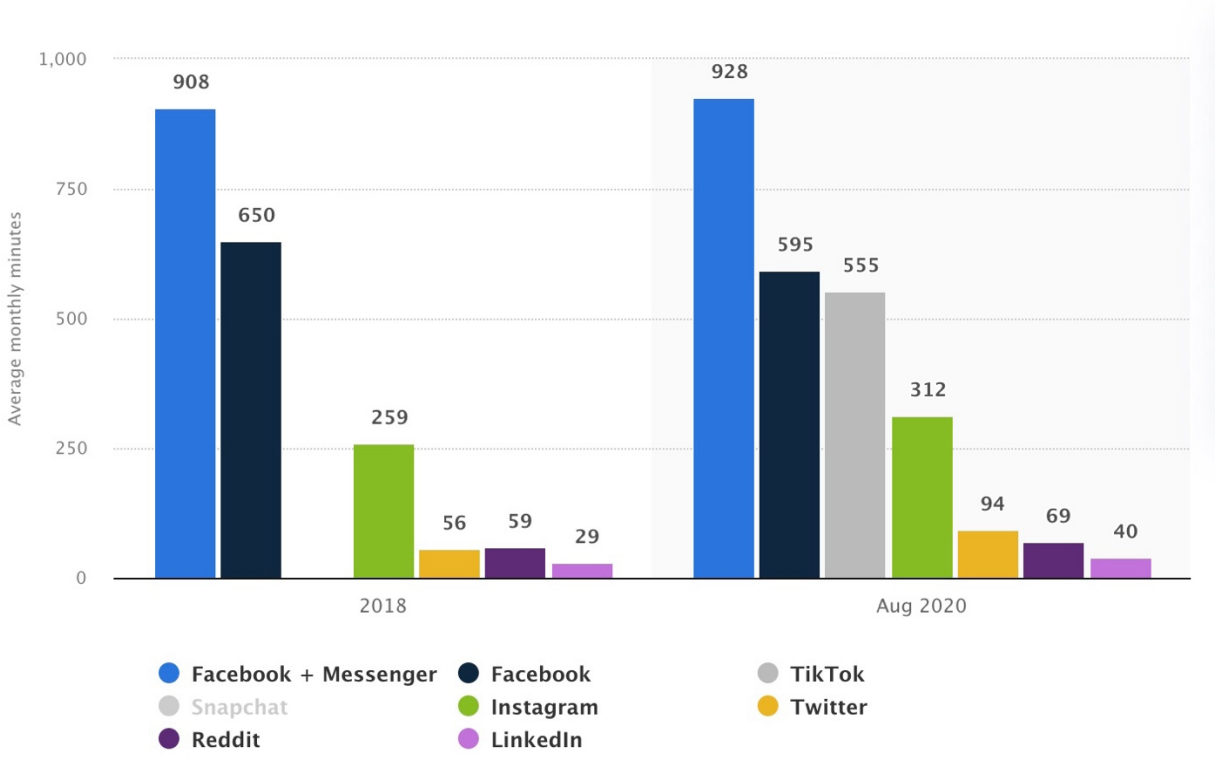


Figure 2: Average monthly time spent with leading social platforms in the United States from 2018 to 2020 (in minutes) According to Statista

2.8. Summary of literature review

After its creation in the 1960s, the EMH was viewed as one of the greatest theoretical and empirical findings in world of finance. People began to challenge the foundation on which the EMH was based. Theoretical critiques included the fact that it is naïve to assume that investors always act rationally, in fact many investors react to immaterial information in forming their portfolios. Another pitfall of the EMH was its inability to reasonably explain disruptions within financial markets. Throughout the 1980s and early 1990s academic literature set out with the focus of explaining these disruptions by arguing that investors, being human, have emotions which in turn have an influence on the decisions they make. These decisions are very often inefficient and irrational and could be the reason as to why disasters or disruptions in the stock market exist (Kapoor & Prosad, 2017). This new form of reasoning paved the path for what has become known today as behavioural finance. Kahneman and Tversky (1974) provide empirical evidence that

illustrates that the principles of expected utility theory do not necessarily hold when making decisions under risk. As such, by combining concepts from both psychology and economics, they introduced an alternative decision-making model, known as Prospect Theory. Prospect Theory suggests that individuals value gains and losses differently and that the dissatisfaction from losing a sum of money is greater than the satisfaction derived from gaining the same amount. In addition to these findings, Prospect Theory argues that are risk averse for gains and risk seeking for losses. Kahneman and Frederick (2002), suggest that cognitive heuristics work via a process called attribute substitution which happens without conscious awareness. A component of behavioural finance that has attracted significant amounts of attention in recent years is the influence that sentiment has on the stock market. Kearney and Liu (2014) differentiate between two forms of sentiment that each play fundamental roles in driving the stock market: investor sentiment and textual sentiment. Investor sentiment relates to the expectations of market participants with regard to both the expected return and risk when making financial decisions (De Long et al., 1990). In contrast, textual sentiment differs from investor sentiment in that it is concerned with the disposition of text-based posts with regard to financial markets. Textual sentiment is a form of computational analysis of sentiment and as such may offer more increased accuracy and be both cost and time effective when compared to traditional methods (Mao, Counts and Bollen ,2011). Numerous studies have attempted to not only to measure textual sentiment but to unearth the relationship between this sentiment and features of the stock market such as Wuthrich et al., (1998), Antweiler and Frank (2004), Schumacher and Chen (2009), Mao, Counts and Bollen (2011).

The fourth industrial revolution has no doubt led to the widespread acceptance and utilisation of internet-based communication forums and social media platforms around the world. According to Statista the number of social media users has increased from 2.86 billion users in 2017 to approximately 3.6 billion in 2020 (an increase of almost 126 percent). In addition to these statistics, the number of social media users is forecast to increase to roughly 4.5 billion by 2025. It is logical to assume that the increase in the number of social media users and therefore an increase in the amount of information available, will enhance market efficiency as information should spread more rapidly compared to traditional news platforms. To support these claims Blankespoor, Miller and White (2014) find that the additional dissemination of firm-initiated news via the Twitter

platform is coherent with a decline in information asymmetry. Twitter is a social media platform that was established in 2006. The platform enables users to share information in real time using 280-character messages colloquially known as “tweets”. The 280-character limit entails these tweets are usually concise and straightforward. Through its evolution, Twitter has become one the primary social media platforms on which firms decide to publish information for public consumption (Paul, 2015). Preliminary studies that analyse the textual sentiment present in Twitter postings find that changes in public mood can in fact predict changes in the closing prices of the DJIA (Bollen, Mao and Zeng ,2011). Similarly, Mao, Counts and Bollen (2011)find that Twitter data is seen to be a statistically significant predictor of daily market log return. Gilbert and Karahalios (2010) show that the prevailing mood of the public, measured through the linguistic analysis of Twitter posts can be seen as indicator in anticipating changes on the S&P 500 index. Additionally, Sprenger and Welpe (2014) find the sentiment and volume of tweets to be related to abnormal returns and next day trading volume. With the outbreak of the COVID-19 pandemic, and all the measures that were implemented to reduce the transmission of the virus (such as social distancing and working from home) individuals found themselves in a unique situation in which they were afforded more time in isolation and as such they turned to social media in an effort to communicate and alleviate boredom induced from isolation. All forms of social media experienced a significant increase in usage time over the sample period, none so more than Twitter which experienced a dramatic increase in usage time of approximately 68%.

Prior to the spread of the COVID-19 pandemic in December of 2019, the financial markets had been relatively stable with little to no major forms of market disruption since the financial crisis of 2007/2008. The outbreak of the COVID-19 was an exogenous shock that was and is still seen to have drastic effects on not only financial markets and economies around the world, but the lives of every individual on the planet. The pandemic ultimately changed the perception of both retail and institutional investors. In the aftermath of such an event, investor behaviour is associated with more risk averse choices. It is likely that there will be a reduced trading intensity and reduction in the flow toward risky assets. Following the subsequent increase in uncertainty brought about with the introduction of the pandemic, the market experienced significant declines and increased financial risk around the globe. Corroborating this point, Zhang, Hu and Ji (2020) find that the pandemic had a significant influence on stock markets. They noted that the risk levels of all

countries in their data sample increased dramatically, further they attribute some of this increased volatility to market sentiment which is thought to be "amplified" through social media channels. Similarly, Baker, Bloom, Davis, Kost, Sammon and Viratyosin (2020a) state that no previous infectious disease (in their sample which spans 120 years) has impacted the financial markets to the extent of COVID-19. Other periods of infectious disease have seen large declines or increases in equity markets over periods of several weeks or months, the COVID-19 pandemic stands out for its extremely high frequency of swings in the market. In their 2020 study, Ortman, Pelster and Wengerek find that ,in contrast to acts of terrorism, average weekly trading intensity actually increased by 13.9%, as the number of positive COVID-19 cases began to rise rapidly. This evidence leads one to question why the COVID-19 pandemic had such a powerful effect on markets in comparison to other infectious diseases. The first answer to this question attributes itself to the severity of COVID-19 which is a function of the apparent ease of transmission and the mortality rate among those who have been in the unfortunate position of contracting it. A second possible answer and one that aligns itself with the context of this study lies within the fact that information about pandemics is richer and diffuses much more rapidly in comparison to previous pandemics, thanks advancements in technology. With this explanation in mind, the impact of the COVID-19 pandemic is more likely to trigger stock market jumps and high stock market volatility more frequently than what has been observed in prior times. A third explanation relies on the role of cross-border flows of goods in the modern economic times, driven by decades of falling transport costs and falling communication costs. These advancements have led to a reliance on relatively enormous geographical supply chains, which are very susceptible to sudden supply disruptions. Thus the ability of the pandemic disrupt these supply chains leads to major market disruptions especially within service-oriented economies.

3. Description of data and research methodology

This section provides an analysis of the data and methodology utilised in addressing the main objective of this study.

3.1. Data

The study makes use of secondary data, sourced from the Bloomberg database. The study utilises the historical stock features of 10 technology firms over the period of February 2016 to December 2021 such as daily trading volume, volatility and raw returns, closing price, intraday high, intraday low as well as the Bloomberg Social Velocity measure[3] attached to the stocks. The firms observed include that of Apple Inc (AAPL), Microsoft (MSFT), Alphabet (Google) (GOOG), Amazon (AMZN), Facebook (FB), Tesla (TSLA), TSMC (TSM), Tencent (TCEHY), NVIDIA (NVDA) and Alibaba (BABA). The reason these technology stocks were chosen is due to the fact that they are globally the largest tech companies by market cap, and historic pricing data is readily available. Data on the NASDAQ 100 technology sector index was also obtained to serve as a benchmark. The rationale behind using this sample period is to account for the fact that prior to October 2015, Bloomberg had not made the BSV data available to users.

Additionally, the fact that the sample period spans over 5 years (which is far larger than that of previous studies), may enable further analysis and therefore greater understanding regarding the relationship between sentiment on social media (Twitter) and movements within the technology sector. In addition, this relatively long time period was chosen as it affords one the opportunity to analyse data in both a pre- (Feb 2016- December 2019) and intra-pandemic (January 2020- December 2021) context. This follows a similar methodology to that of Li, Shen and Zhang (2019) who also split their sample to gauge the magnitude and significance of investor sentiment in different market conditions. This sample period thus provides for a unique analysis of not only how the markets responded to an anomaly such as the COVID-19 pandemic but the relationship between stock market features and the use of social media during unprecedented times.

---

[3] The Bloomberg social velocity index is a measure that employs the utilises sophisticated tools in an attempt measure a tweet's sentiment. The system relies on algorithms developed by a team of PhDs, product managers and R&D professionals incorporating statistical models and years of Twitter history applied to financial markets (Bloomberg, 2014).

3.2. Variables

3.2.1. *Tweet* Features

As in Sprenger (2014), three *tweet* features are used as the explanatory variables: namely; message volume (M_t), bullishness/overall sentiment (B_t), and message agreement (A_t). The process of calculating the sentiment (bullishness) index used by Bloomberg Inc. starts with manually analysing large datasets of *tweets* using human experts. The manually classified feeds are then fed into machine learning models that are taught to imitate language experts in analysing text-based messages. The completed machine learning models are subsequently used to analyse new *tweets* tagged with tickers and assigns each *tweet* a story-level sentiment score ranging from -1 to +1 in real-time. The average firm-level daily sentiment is then extracted from the weighted average story-level sentiment scores in the last 24 hours collected from Twitter and StockTwits and updated every day 10 minutes before the market opens and is calculated as:

$$B_{i,t} = \frac{\sum k \in P(i,T) S_i^k C_i^k}{N_{i,T}}, \qquad T \in [t-24h, t] \quad (1)$$

Where:

$B_{i,t}$ is the bullishness score for firm i at time t;

S_i^k is the sentiment polarity score for *tweet* k that references firm i;

C_i^k is the confidence of *tweet* k that references firm i;

$P(i,T)$ is the set of all non-neutral *tweet* feeds that reference firm i in the 24 hour-period, *T;*

$N_{i,T}$ is firm $i's$ total number of positive or negative *tweets* during period *T.*

Bullishness ranges from -1, the most negative sentiment to +1, the most positive sentiment. Hence, a bullishness score of 0 denotes neutral sentiment.

Bloomberg provides the daily total number of posts uploaded to Twitter *(tweets)* for each firm i aggregated at the end of the day. Message volume ($M_{i,t}$) in this study is calculated as the natural logarithm [$\ln(1 + aggregate\ tweets)$] of the aggregate *tweets* for stock i at time interval t.

Consistent with other studies (such as Antweiler & Frank; Kim & Kim, 2014), 1 is added in the log-transformed equation as indicated above to cater for firms with no *tweets* at any point in time since the natural logarithm of zero is undefined.

The message agreement index ($A_{i,t}$) reflects the extent of the consensus among microbloggers on the prospects of each stock i at time-variable t. Following Antweiler and Frank (2004), the following is used to measure message agreement:

$$A_{i,t} = 1 - \sqrt{1 - \left(\frac{Pos_{i,t} - Neg_{i,t}}{Pos_{i,t} + Neg_{i,t}}\right)^2} \in [0,1] \quad (2)$$

Where $Pos_{i,t}\ and\ Neg_{i,t}$ indicate the number of messages which are respectively categorised as positive and negative. If all the messages at a given time are equally dispersed between positive and negative, it means that there is absolute disagreement among Twitter users and therefore the value of A_t will be equivalent to 0. In a situation where all messages are either positive or negative, then it means Twitter users are in total agreement and the agreement index will therefore be 1. Thus, the closer the calculated numerical value of the agreement index is to 0, the greater the disagreement among stock Twitter users while a value close to 1 indicates greater agreement. Consider the case, when presented with 3 messages, if all three messages are positive, $Pos_{i,t} = 3, Neg_{i,t} = 0$ then, $A_{i,t} = 1$.

The primary benefit of utilising the above metric for agreement is that it measures the distribution of investor opinions directly as opposed to alternative metrics which rely on more indirect measures such as volatility and analyst forecast dispersion. In addition to this, the agreement measure is computed on a daily basis, in contrast to alternative metrics which are usually measured at lower monthly and quarterly frequencies (Diether, Malloy & Scherbina, 2002). Consistent with Antweiler and Frank (2004) and Sprenger *et al.,* (2014), this study assigns a value of zero to *tweets* features for all "quiet" periods. If the previous value is maintained instead of assigning a zero value during a period without messages this may induce a bias toward bullishness, as noted by Antweiler

and Frank (2004). Imputing zero values to companies that are not mentioned on the Twitter and StockTwits platforms on any day is done on the presumption that when investors do not mention a specific counter, this means that they are neutral on the prospects of the counter.

3.2.2. Market Features

Three stock market features are used in the study, namely; Trading Volume ($TV_{i,t}$),Volatility ($V_{i,t}$) and Raw returns ($R_{i,t}$). Raw returns ($R_{i,t}$) for stock i at time interval t are defined as:

$$R_{i,t} = ln\left(\frac{P_t}{P_{t-1}}\right) \quad (3)$$

Where P_t denotes the closing price at time interval t and P_{t-1} denotes the closing price at time interval $t-1$. Continuously compounded returns are used instead of discrete returns to overcome any issues of non-normality since stock returns are usually assumed to be lognormally distributed (Brooks, 2019). The continuously compounded formula for computing returns in Equation [4] is therefore taken to constitute a non-linear transformation of the data. For holidays when there is no trading, missing values are imputed using the average of the closing price a day before and a day after the missing value day in conformity with previous studies (such as Sprenger *et al*., 2014). The missing values of the closing prices are therefore imputed using the following formula:

$$m_{i,t} = \frac{P_{i,t-1} + P_{i,t+1}}{2} \quad (4)$$

Where:

$m_{i,t}$ represents the imputed value of the closing price of stock i at time t,

$P_{i,t-1}$ is the closing price of stock i the day before the missing value.

$P_{i,t+1}$ is the closing price of stock i the day after the missing value.

This study will employ a volatility estimation model that captures drops and recoveries of financial markets daily instead of the classical close-to-close volatility models to overcome the estimation

errors inherent with the GARCH family of models . Volatility ($V_{i,t}$) in this study will be estimated using intraday price highs and lows in line with the PARK volatility measure (Parkinson, 1980). Assuming that the stock price follows a simple Brownian model without a constant term, the PARK statistic is calculated as follows:

$$VOL^{PARK} = \frac{(\ln (H_t - \ln(L_t)))^2}{4\ln (2)} \quad (5)$$

Where $H_t \, and \, L_t$ stand for the intraday highs and lows of a stock price respectively. PARK volatility has also been used in academic articles examining the role of textual sentiment in capital markets (such as Sprenger *et al*., 2014; Li, van Dalen & van Rees, 2018) and is, therefore, useful for comparing the findings from this study with previous studies.

Trading volume ($TV_{i,t}$) is calculated as the natural log of the traded volume of stock i at time t, which is standard practice within the finacial world as the raw volume is typically too large to detect significant changes. The log-transformed trading volume is computed using $ln\,(1 + trading\ volume)$ in an effort to account for firms that are subject to a zero-trading volume at any time[4].

3.3. Econometric model estimation

3.3.1. The relationship between *tweet* features and stock market features

This subsection presents the tests for the primary hypotheses of the study. Following previous studies (such as Sprenger *et al*., 2014), panel regressions with both firm fixed-effects are used to analyse the contemporaneous link between *tweet* features and market features as shown in Equation [6]. This model follows that of Sprenger *et al.,* (2014); where all the *tweet* features are used as covariates and the market index is included as a control variable as shown below:

[4] This is done to prevent a value of 0, due to the fact that the natural logarithm of 0 is undefined

$$Y_{i,t} = \beta_1 B_{i,t} + \beta_2 M_{i,t} + \beta_3 A_{i,t} + \beta_4 R^m{}_t + \delta_i + \epsilon_{i,t} \quad (6)$$

Where:

$Y_{i,t}$ represents the three market features (firm-level stock returns, volatility and trading volume) of firm i at time t;

$B_{i,t}$ is bullishness (sentiment);

$M_{i,t}$ represents message volume of firm i at time interval t;

$A_{i,t}$ represents message agreement for firm i at time interval t;

$R^m{}_t$ is the market return calculated as the natural logarithm of the closing value of the NASDAQ 100 Technology Sector Index at time t divided by the value at time $t-1$;

$\epsilon_{i,t}$ is an error term that is clustered by firm;

δ_i is the unknown intercept for every firm.

3.3.2. Pooled quantile regressions

To examine the link between the magnitude of *tweet* features and stock market features, the study uses quantile regression[5]. Quantile regressions are used in this study to test the secondary hypotheses of the study. This type of model was first propositioned by Koenker and Bassett (1978) and has been utilised extensively in research examining investor sentiment as well as in asset pricing studies (such as Swamy & Dharani, 2019). Quantile regressions are often used as alternative measures to the typical ordinary least squares (OLS) regression and related methods. This method relaxes the assumption that associations between the dependent and independent variables are constant across all levels. Utilising a quantile regression allows one to relax the common regression slope assumption. The quantile regression differentially weights the distances between values predicted by the regression line and the observed values and then attempts to

[5] The pooled quantile regressions are estimated using the Rstudio Package *"quantreg"*

minimise the weighted distances (Lê Cook & Manning, 2013). Quantile regression necessitates the approximation of conditional quantiles of the regressand given a range of predictor variables without splitting the sample (Koenker & Bassett, 1978). The quantile regression estimator also permits the effect of the explanatory variable to fluctuate across quantiles of the dependent variable. A few of the documented advantages of quantile regression estimators include the fact that they are robust or less sensitive to outliers (Koenker & Machado, 1999) and they deal with non-linearity without presuming a specific form of the model (Koenker & Bassett, 1978). Quantile regression is also useful because it allows the comparison of coefficients across quantiles using interquartile regression (Koenker & Bassett, 1978). Using quantile regression, the conditional quantile function of $\gamma_{i,t}$ at quantile τ given explanatory variable $x_{i,t}$ is defined as follows:

$$Q_{\tau}\left(\gamma_{i,t}\middle|x_{i,t}\right) = c_{\tau} + \beta_{\tau}x_{t} + F_{\varepsilon_t}^{-1}(\tau) \quad (7)$$

Where F_{ε} stands for the distribution of errors and β_{τ} and c_{τ} are the parameters. The coefficients of the τ^{th} conditional quantile regression are approximated as follows:

$$\hat{\beta}^{\tau} = \underset{c_{\tau},\ \beta_{\tau} \in \mathbb{R}}{\arg\min} \sum_{t=1}^{T-1} \rho_{\tau}\left(\gamma_{i,t} - \left(c_{\tau} + \beta_{\tau}x_{i,t}\right)\right) \quad (8)$$

Where T indicates the sample size and ρ_{τ} is the check function defined as $\rho_{\tau}(\varepsilon) = \tau\varepsilon$ if $\varepsilon \geq 0$ and $\rho_{\tau}(\varepsilon) = (\tau - 1)\varepsilon$ otherwise. This study uses equation (6) as the baseline equation for the quantile regression model specification as follows:

$$Y_{it,\tau} = \alpha_{\tau} + \beta_1 B_{it,\tau} + \beta_2 M_{it,\tau} + \beta_3 A_{it,\tau} + \beta_4 R^{m}{}_{t,\tau} + \epsilon_{it} \quad (9)$$

Where:

τ is the τ^{th} quantile in the conditional distribution of the regressand.

Equation (9) is the quantile regression model which is estimated to ascertain whether the magnitude of sentiment is linked to the magnitude of stock return. Akin to that previous literature such as Swamy & Dharani (2019), five quantile intervals representing the different market conditions as shown by the firm-level stock returns within the technology sector are used as follows: $\tau \in (0.1, 0.25, 0.5, 0.75, 0.9)$

Where:

$\tau \in (0.1, 0.25)$ represent unfavourable market conditions;

$\tau \in (0.5)$ represents mediocre or normal market conditions and;

$\tau \in (0.75, 0.9)$ represent favourable market conditions.

3.4. Granger Non-Causality Analysis

To test the causality of the relationship between Twitter posting volume and the stock market data, this study will utilise the Granger non-causality technique. This analysis will determine whether the Twitter predictor provides information that could be beneficial in predicting stock market return within the technology sector . Granger non-causality analysis assumes that if a variable X_t causes Y_t then changes in X_t will steadily occur before changes in Y_t (Bollen *et al*., 2011).

The Granger non-causality test will be employed in the same fashion as Bollen *et al.,* (2011). The stock return time series, D_t, reflects daily changes in stock price value. This test will compare the variance explained by two linear models as shown in Equations (10) and (11) to examine whether Twitter volume predicts changes in stock market values over the sample period. The first model (L_1) uses *n* lagged values of D_t for prediction. The second model (L_2) uses the *n* lagged values of both D_t and the Twitter predictor denoted X_t. The models are as follows:

$$L_1 : D_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \epsilon_t \quad (10)$$

$$L_2 : D_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \epsilon_t \quad (11)$$

Although this technique is used for causation, it can only identify whether one-time series can explain information about another, which is not necessarily true causation. Granger causality is sensitive to numerical properties which vary over time, in other words non-stationary time series. A further limitation to this technique is that it models information linearly (Gilbert & Karahalios, 2010). The relationship between Twitter volume and stock market data is not always linear.

3.5. Robustness checks

The pooled quantile regression approach outlined in Section 3.3.2 might be susceptible to the problem of unobserved heterogeneity. Koenker (2004) proposes a fixed-effects quantile regression model that addresses the prospect of endogenous independent variables. Consistent with Koenker (2004), the following model for the conditional quantile functions of the dependent variable of firm i at time t is used:

$$Q_{yi,t}(\tau|x_{i,t}) = \alpha_i + x'_{i,t}\beta(\tau) \quad (12)$$

Where:

α_i indicates the firm fixed effects;

$x'_{i,t}$ is a vector of predictor variables and;

The $\tau - dependent$ vector β is the vector of predictor variables of parameters to be estimated.

In the model, the firm fixed effects α_i imply a pure location shift on the conditional quantiles of the regressand. Essentially, the effects of the independent variables are conditoned on the quantile τ while the effects of α_i are not. To estimate the model in Equation (12), Koenker (2004) proposes solving the following using linear programming:

$$\min_{\alpha,\beta} \sum_{k=1}^{q} \sum_{t=1}^{T} \sum_{i=1}^{n} \omega_k \, \rho_{T_k}(y_{i,t} - \alpha_{i,t} - x'_{i,t}\beta(\tau_k)) \quad (13)$$

Where:
ω_k denotes the weights controlling the relative influence of the quantiles on the estimation of the α_i parameters and;
$\rho_\tau(.)$ is the quantile loss function.

The fixed effects panel quantile regression methodology outlined herein has been widely used in various studies in the area of finance, particularly studies looking at investor sentiment and other behavioural finance topics (such as Swamy & Dharani, 2019). This study estimates the fixed effects panel quantile regression model in Equation (12) using the *"rqpd"* Rstudio package to optimally solve Equation (13).

In addition to the robsutness check mentioned above, to further guarantee the robustness of the results, two additional quantile regression are estimated using the conditional quantiles τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] and $\tau = \epsilon[0.05, 0.2, 0.4, 0.6, 0.8, 0.95]$ which is in line with that of previous studies such as Ni and Xu (2015) and Ma, Xiao and Ma (2018) respectively

3.6. Limitations

As with all studies of this nature the common limitations include that of time constraints access to information. Furthermore, this studies relies on information sourced from the Bloomberg Social Velocity index which is an aggregated daily metric, thus the most in-depth possible analysis will take place on a daily level.

4. Empirical analyses

This section explores the analysis of the data, and reports findings on the same. Section 4.1 places emphasis on an analysis of the full sample, section 4.2 shifts the focus to an analysis that precedes the COVID-19 pandemic. Finally, section 4.3 focuses on an intra-pandemic analysis of the data.

4.1. Overall sample

4.1.1. Descriptive statistics

The summary of descriptive statistics regarding the data utilised in this study are presented in Table 1 below. These statistics provide detail on the 10 technology companies used in this analysis over the entire sample period. The panel data includes observations of both the tweet features and market features.

As shown in Table 1 below, bullishness ranges from -0.09163 to a maximum value of 0.9898 which is indicative that textual sentiment as measured by the calculated metric is relatively dispersed across the entire sample. The mean value of the Bullishness score is approximately 0.0099 which indicates that the overall or average textual sentiment of individuals, although very close to neutral (0), is in fact slightly positive over the entire sample period. When considering the messages or in this case the *tweets* over the entire sample, the average number of daily *tweets* referencing each of the firms under observation is approximately 1947, the minimum number of *tweets* referencing at least one of the firms is 1, meaning that at least one firm was mentioned every day over the duration of the period under observation. The cumulative maximum number of daily messages mentioning the firms in the study is approximately 142267.

Table 1: Summary Statistics of the Variables (Full Sample)

Statistic	N	Mean	St. Dev	Median	Min	Max
Tweet Features						
Bullishness	15670	0.0099	0.0912	0.0006	-0.09163	0.9898
Messages	15670	1946.3724	3080.8665	1123	1	142267
Agreement	15628	0.0982	0.1275	0.0499	0	1
Market Features						

Statistic	N	Mean	St. Dev	Median	Min	Max
Stock Returns	21920	0.0008	0.0189	0	-0.2365	0.2609
Trading Volume	15670	29924716	42211530	15672136	1587	5334786696
Volatility	15670	0.0008	0.0001	0.0001	0	0.051
Market Return	12530	0.0009	0.00153	0.0011	-0.1043	0.0971

Figure 3 below shows the average daily distribution of tweets. As can be seen in this figure, and confirmed in the Table above, the average daily messages regarding each of the firms under observation is approximately 1947 over the sample with an unexpected decrease in the number of messages in more recent years. The period with the highest number of daily *tweets* occurs during the latter half of 2016 which saw a number of firms in the sample release revolutionary new products, an example of this includes the release of the highly anticipated Apple Airpods during December 2016. The period with the lowest number of *tweets* is observed to be in the beginning of 2018. This period saw a decline in smartphone sales perhaps due a perceived lack of innovation in comparison to that of previous years.

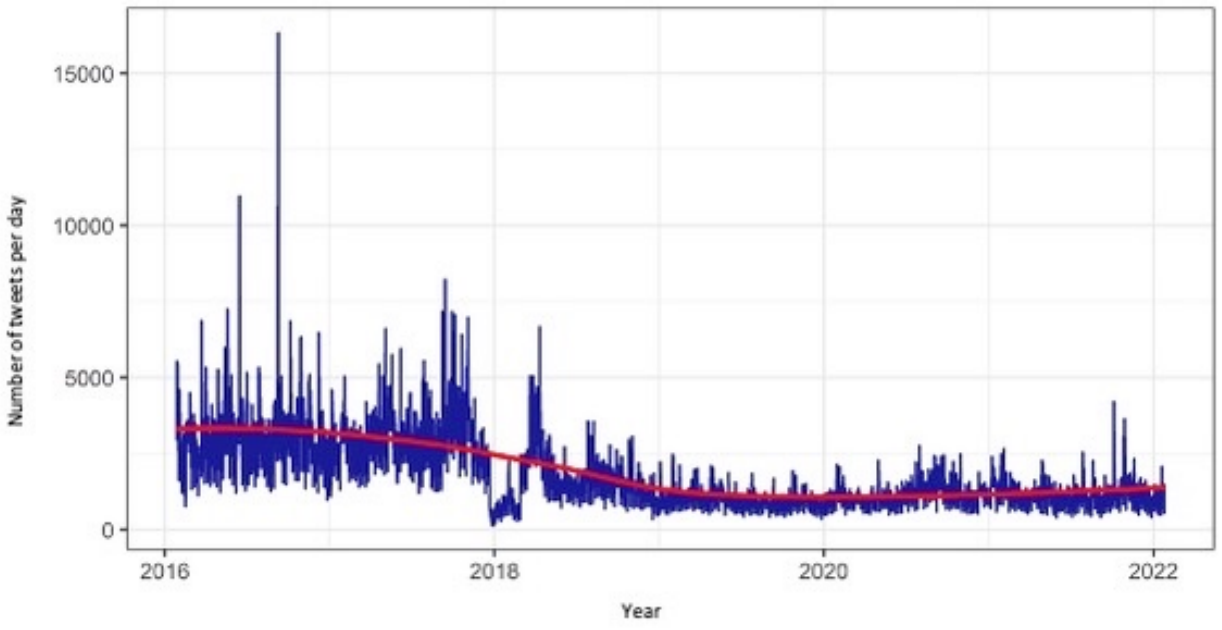


Figure 3: Daily Distribution of aggregate tweets

Note: *The y-axis in the Figure represents the amount of average daily tweets across all firms, whilst the x-axis shows the timeline. The blue area of the graph shows the time series of the average daily messages (tweets), whilst the red line indicates the smoothed conditional means with a 95% confidence band.*

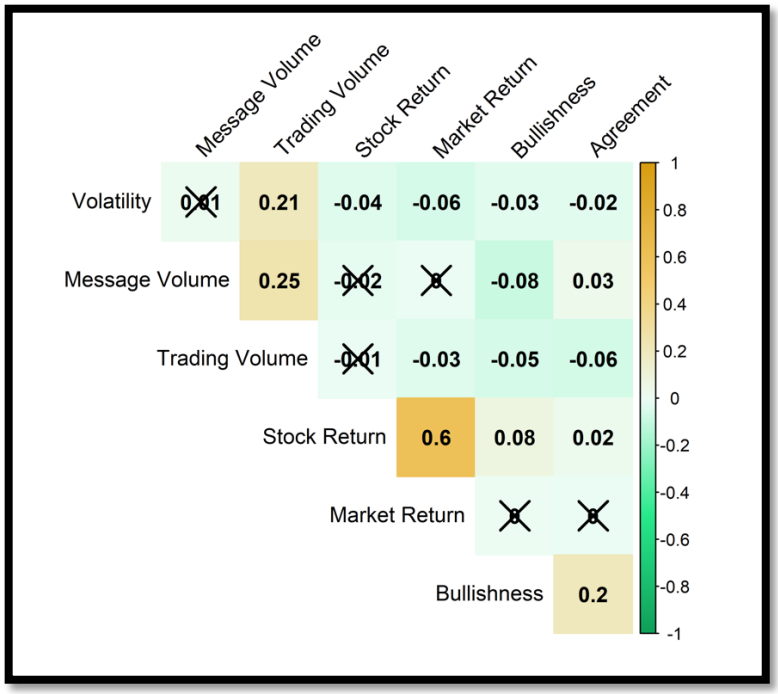
As expected, the message agreement index is seen to have a maximum value of 1, indicating total agreement and a minimum value of 0 indicating total message disagreement. With a mean value of approximately 0.01 signifying a general disagreement in textual sentiment over the period and thus displaying the difference in user opinions.

When observing the summary statistics of the market features, the average return over the period is seen to be approximately 0.08%, with a minimum value of -23.65% and a maximum value of 26.09%. In comparison, the market average return, which in this case is based on the performance of the NASDAQ 100 Technology sector index, is 0.009% with a maximum of 9.71% and a minimum of -10.43%. The average volatility over the period is an estimated 0.08% which is relatively small.

4.1.2. The correlation between *tweet* features and market features

Table 2 below represents the pairwise correlation matrix that attempts to show the correlation between tweet features (message volume, bullishness and agreement) and market features (stock return, trading volume volatility and market return. As shown in the matrix, the correlations relatively low and more often than not are negative. These low to moderate correlation figures (<0.7) is signifies the absence of potential multicollinearity with the variables.

Table 2: Pairwise correlation matrix (full sample)



Note: Blocks with an X/cross denote a case in which correlation is insignificant at the 5% level. The shading and numerical value indicate the relative strength of the correlation. A value closer to 1 within a block shaded toward the dark orange side of the spectrum is assumed to represent a highly positive correlation, whilst a value closer to -1 within a block shaded toward green in the colour spectrum is assumed to represent a highly negative correlation. A value of 0 and a more neutral colour is associated with a weaker correlation amongst variables.

Bullishness is found to have a weakly positive correlation with stock return (0.08). In contrast to this bullishness and trading volume are weakly negatively correlated (-0.05), indicating that as bullishness increases trading volume decreases. In addition to this, bullishness is found to have no effect on market returns. Message volume and stock return is observed to have a correlation which

deemed insignificant (-0.02) at the 5% level , thus there is no relationship between message volume and stock return. Similarly, there is no indication of a relationship between message volume and market return. Message volume and trading volume is seen however to have weakly positive relationship (0.25). Message Agreement and stock returns are found to have weak positive correlation with one another (0.02). Surprisingly, message agreement and trading volume are found to have a weak negative relationship with one another (-0.06). Message agreement is found to have no correlation with market return with a value of (0) which is deemed insignificant at the 5% level. Although correlation does not necessarily imply causation, the significance of majority of the figures in Table 2 require further investigation by utilising more meticulous tests in the form of comprehensive econometric models. These results may uncover additional relationships that are unobservable at a surface level.

4.1.3. The contemporaneous association between market and tweet features

Following the correlation results in the preceding section, this section aims to shed additional light on the relationship between tweet features and market features. More specifically this section aims to address the primary hypothesis as outlined in the introduction, that is the ability of tweet features to explain market features of the firms under observation. Both a fixed effect model and random effect model can be utilised to achieve this. Gelmon (2004) notes that there are several different definitions of both fixed and random effects, however, in the context of this study the following definitions apply: In the random effects model, effects are estimated with shrinkage or in the words of Robinson (1991, p.22) “Best linear unbiased prediction (BLUP)”. In contrast, the fixed effects model estimates effects using a method of least squares or more commonly maximum likelihood. A Hausman test is utilised in an effort to best select the most appropriate model to be employed for the remainder of the analysis. The Hausman test is often regarded as test for model misspecification, in other words, the model may not account for all it that it should. Mis-specified models often have biased coefficients and error terms and, in some cases, permit biased parameter estimations Glen (2017). The Hausman test has the null hypothesis that the preferred model is random effects. The alternate hypothesis is thus, the preferred model is fixed effects. This test essentially examines if there is a correlation between the regressors and the unique errors in the model.

Table 3: Hausman Test (Full Sample)

	Stock Return	Volatility	Trading Volume
χ^2	10.141	40.202	0.082
p- value	0.038	0.000	0.999

Note: *1. The null hypothesis that the preferred model is random effects.*

2. The alternate hypothesis is thus; the preferred model is fixed effects.

Interpreting the results presented in Table 3 is fairly straightforward: if the p-value is less than 0.05 one should reject the null hypothesis in favour of the alternate hypothesis. Thus with this in mind, it is clear to see that the preferred model for both stock return (0.038<0.05) and volatility (0.000<0.05) is that of the fixed effects model at a 5% level of significance. The preferred model for when considering trading volume (0.999>0.05) is that of the random effects. Thus, fixed effect models are chosen to measure both stock return and volatility and their association with the various *tweet* features which is akin to that of previous studies (Antweiler & Frank, 2004; Twedt & Rees, 2011). In contrast, the random effects model is deemed the most appropriate model to measure the relationship between the trading volume and various tweet feature and will therefore be employed.

Table 4: Contemporaneous OLS regressions (full sample)

	Dependant Variable		
	Return	**Volatility**	**Log(Volume)**
Bullishness	0.0018^{***}	-0.0001	-0.190^{***}
	(0.002)	(0.0001)	(0.053)
Message Volume	-0.00000^{**}	0.000^{**}	0.00003^{***}
	(0.0000)	(0.000)	(0.0000)
Message Agreement	0.002^{*}	-0.00001	0.087^{**}
	(0.001)	(0.0001)	(0.040)
Market Return	0.860^{***}	-0.003^{***}	-2.237^{***}

	Dependant Variable		
	(0.010)	(0.0004)	(0.305)
Observations	12497	12497	12497
R^2	0.371	0.004	0.031
Adjusted R^2	0.371	0.003	0.030
F Statistic	1844.541^{***}	12.240^{***}	98.508^{***}

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. *= p<0.1; **=p<0.05; ***=p<0.01

Table 4 presents the three models which show the regression of stock returns, volatility and trading volume on the *tweet* features (bullishness, message volume and message agreement). The results of model 1 (stock returns) indicate that bullishness is in fact positively related to stock returns at a 1% level of significance ($\beta = 0.0018, p < 0.01$), this implies that a greater level of sentiment will likely lead to an increase in the stock price of the firm. Message volume is seen to have inconsequential relation to stock returns at a 5% level of significance ($\beta = -0.0000), p < 0.05$ which suggests that as the number of *tweets* increases or in this case decreases, stock returns remain unchanged. These findings are coherent with that of Sprenger et al. (2014) yet inconsistent with the those of both Antweiler & Frank, 2004 and Li, van Dalen and van Rees (2018) who note that message volume had both a positive and significant effect on stock returns. Similarly, in contrast to that of previous studies (Antweiler & Frank, 2004; Twedt & Rees, 2011), the message agreement index is seen to have a small yet positive relation to stock returns that is significant at a 10% level ($\beta = 0.002, p < 0.1$) which implies that agreement among Twitter users is related to an increase in the stock return of the respective firm. These results reveal that bullishness has a greater effect on stock returns than that of both message agreement and message volume. This model is also seen to both a R^2 and adjusted R^2 value of 0.371. Thus, 37.1% of the movement in stock returns is thought to be explained by the variables in the regression, which is a relatively large number in comparison to that of other studies utilising this same methodology

The findings in model 2 (volatility) indicate, contrary to that of previous studies (Kim & Kim, 2014; Sprenger et al., 2014), that bullishness is associated with a relatively tiny negative effect on volatility ($\beta = -0.0001$), this result is however found to be statistically insignificant at any conventional level. Message volume is seen to have a net zero effect on that of volatility, this finding is also significant at a 5% level ($\beta = 0.0000, p < 0.05$), meaning that as the number of *tweets* increase there no observed effect on the volatility across any of the firms in the sample. Although significant, these results oppose that of those found in previous research papers (Antweiler & Frank, 2004 and Sprenger et al., 2014). The message agreement index is observed to have a very small negative effect on volatility ($\beta = -0.00001$), which logically is in line with what one would expect, meaning that an increase in the disagreement amongst micro-bloggers who utilise the Twitter platform will lead to a subsequent increase in volatility. Differing from that of earlier studies this finding is statistically insignificant at any meaningful level. Both the R^2 and the adjusted R^2 are seen to be relatively small (0.004 and 0.003 respectively) which initially is thought of as cause for concern, however, prior studies are found to have similar values for this measure.

The third and final model in Table 4 examines the relationship between that of trading volume (in this case the natural logarithm of trading volume) and the various *tweet* features. Surprisingly, the results indicate that bullishness has a negative relation with trading volume ($\beta = -0.190, p < 0.01$),this finding is also statistically significant at the 1% level which is not coherent with the findings of previous studies (Li, van Dalen and van Rees (2018)). This suggests that as overall sentiment increases or is positive, the number of traded shares of the firms under observation decreases. There is however a positive association between message volume and the volume of traded shares, a finding that is statistically significant at the 1% level of significance ($\beta = 0.00003, p < 0.01$). These findings are akin to that of Sprenger et al., (2014) and since both the message volume and the trading volume are variables which have undergone a log transformation, these results can be interpreted as elasticities. Thus a 1% increase in message volume will lead to 0.003% change in trading volume. This result indicates the inelastic property of these two variables. In short, an increase in message volume leads to a relatively tiny increase in the number of traded shares. Observing the findings on the association between the message agreement index and trading volume, one will note a positive association between the two variables which is

statistically significant at the 5% level ($\beta = 0.087, p < 0.05$). These findings are coherent with that of Sprenger et al., (2014) that note that difference in information need to interact with other forms of heterogeneity, such as heterogeneous beliefs or biases that were mentioned in section 2 of this study, to induce trading.

4.1.4. The association between the size of stock returns and that of *tweet* features

The intention of this section is to address the secondary hypothesis of the study which is to examine if the there is a monotonic association between the size of market returns and *tweet* features within the technology sector. To examine this relationship, this study makes use of pooled quantile regressions. Quantile regressions are often used as alternative measures to the typical ordinary least squares (OLS) regression and related methods, this method relaxes the assumption that associations between the dependent and independent variables are constant across all levels. Utilising a quantile regression allows one to relax the common regression slope assumption. The quantile regression differentially weights the distances between values predicted by the regression line and the observed values and then attempts to minimise the weighted distances (Lê Cook & Manning, 2013).

Table 5 presents the results from the quantile regressions at each of the given stock return quantiles ($\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$).

Table 5: Pooled Quantile Regression Estimates (full sample)

	Dependent Variable				
	Stock Return				
	0.1	**0.25**	**0.5**	**0.75**	**0.9**
Bullishness	0.011***	0.006***	0.003***	0.008***	0.017***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Message Volume	-0.00000	0.0000***	-0.00000	-0.00000***	-0.00000
	(0.00000)	(0.000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.005**	0.0001	0.0004	0.003*	0.011***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)
Market Return	0.871***	0.845***	0.822***	0.835***	0.879***
	(0.017)	(0.008)	(0.006)	(0.010)	(0.018)
Constant	-0.015***	-0.007***	-0.0001	0.007***	0.016***
	(0.0004)	(0.0002)	(0.0001)	(0.0002)	(0.00004)
Psuedo R^2	0.518	0.4704	0.4044	0.4656	0.4834
Observations	12497	12497	12497	12497	12497

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

*2. *= p<0.1; **=p<0.05; ***=p<0.01*

The results observed in Table 5 indicate that bullishness has a positive and significant association with stock returns, across all quantiles. Under closer examination of these results, it is found that bullishness has a positive relationship with stock returns in times where stock returns are in their lowest quantile ($\beta = 0.011, \tau \in [0.1]$). This relationship is seen to weaken from the second ($\tau \in [0.25]$) to the third quantile ($\tau \in [0.5]$) and strengthen again slightly in the fourth quantile ($\tau \in [0.5]$) which is an indicator that in times where returns are more representative of normal market conditions (neither terrible nor fantastic), the effect of bullishness present in Twitter posts is seen to relatively small effect. On the other hand, similar to that of the lowest stock return quantile, in the highest stock return quantile ($\tau \in [0.9]$) the relation between stock returns and bullishness is

seen to be the greatest ($\beta = 0.017, \tau \in [0.9]$). It is important to note that these results are significant at a 1% level across all of the specified quantiles. These findings imply that bullishness is assumed to have a stronger effect on stock returns during good times than both mediocre and bad times. Furthermore, all coefficients are observed to be statistically significant at a 99% confidence interval. Figure 4 provides a visual representation of the relationship between bullishness and stock returns at the 5 specified levels which are included in Table 5.

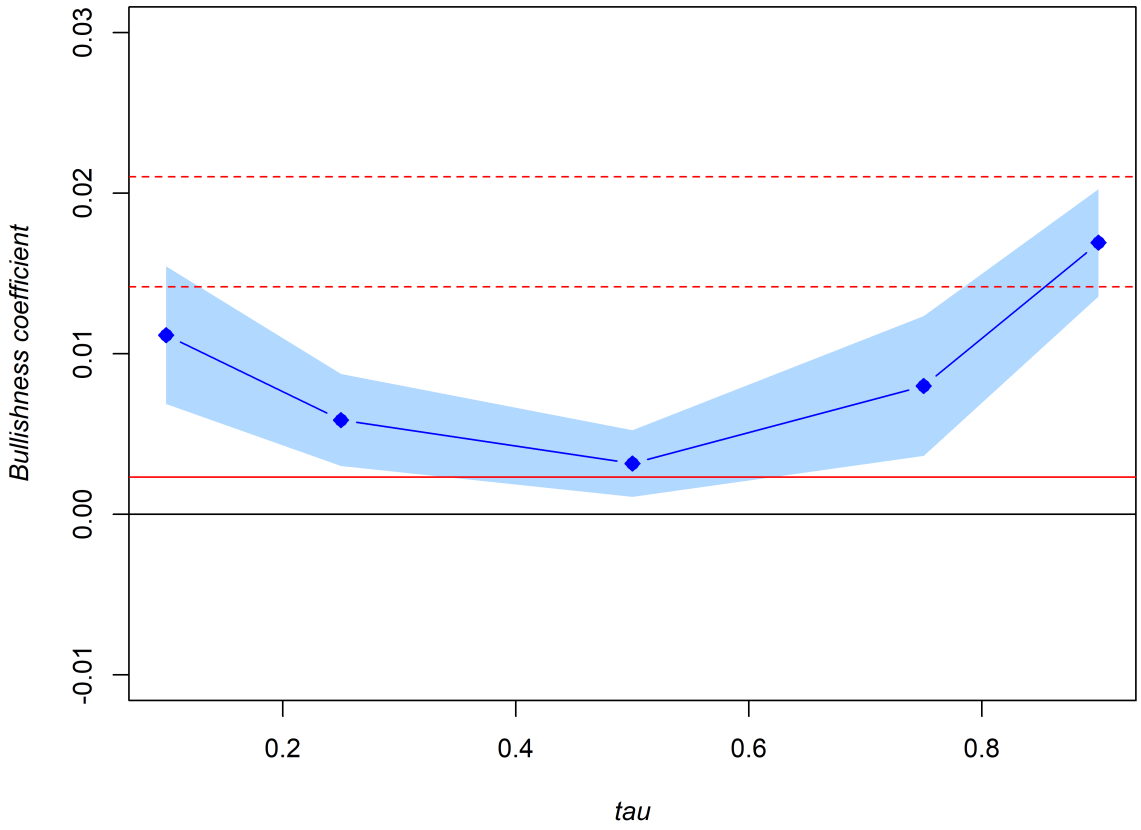


Figure 4: Visualisation of QR and OLS estimates with regard to bullishness at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* ($\tau \epsilon [0.1, 0.25, 0.5, 0.75, 0.9]$) *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a*

specified quantile measurement[6]. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated $\beta(\tau)$ parameters.

Figure 4 corroborates the results presented in Table 5. The convex nature of the graph implies that bullishness has a stronger relationship with stock returns during times of extreme downturn or extreme prosperity ($\tau\epsilon[0.1, 0.9]$) in which stock return margins are increased. Under normal market conditions however ($\tau\epsilon[\ 0.25, 0.5,0.75]$), this relationship is seen to weaken significantly. These results are expected as individuals are more likely to seek and receive more information during times of extreme underperformance and overperformance in the market. Figure 4 also indicates that coefficient estimate of bullishness calculated in the more common ordinary least squares (OLS) regression are likely to be biased, based on the fact that the OLS estimator concurs with that of the quantile regressions only at the highest quantile ($\tau\epsilon[\ 0.9]$). The remainder of the estimated coefficients of the bullishness index in the quantile regression lie outside the 95% confidence interval of the OLS regression whilst the coefficients of bullishness at all quantiles remain within the 5% level of significance for the quantile regression as indicated by the shaded blue area. Figure 4 does not imply the existence of a monotonic relationship between the bullishness index and stock returns.

---

[6] In figures 4 to 7, five points are visualised, representing the specified quantiles as stated in Table 5, is done for ease of interpretation and coherency. To guarantee the robustness of the results in this section, an additional quantile regression is estimated using the conditional quantiles τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] which is in line with that of previous studies (Ni and Xu ,2015). Furthermore, the regression above is re-estimated using $\tau = \epsilon[0.05, 0.2, 0.4, 0.6, 0.8, 0.95]$ which is akin to that of Ma, Xiao and Ma (2018) .The results of these quantile regression are attached in appendix A.

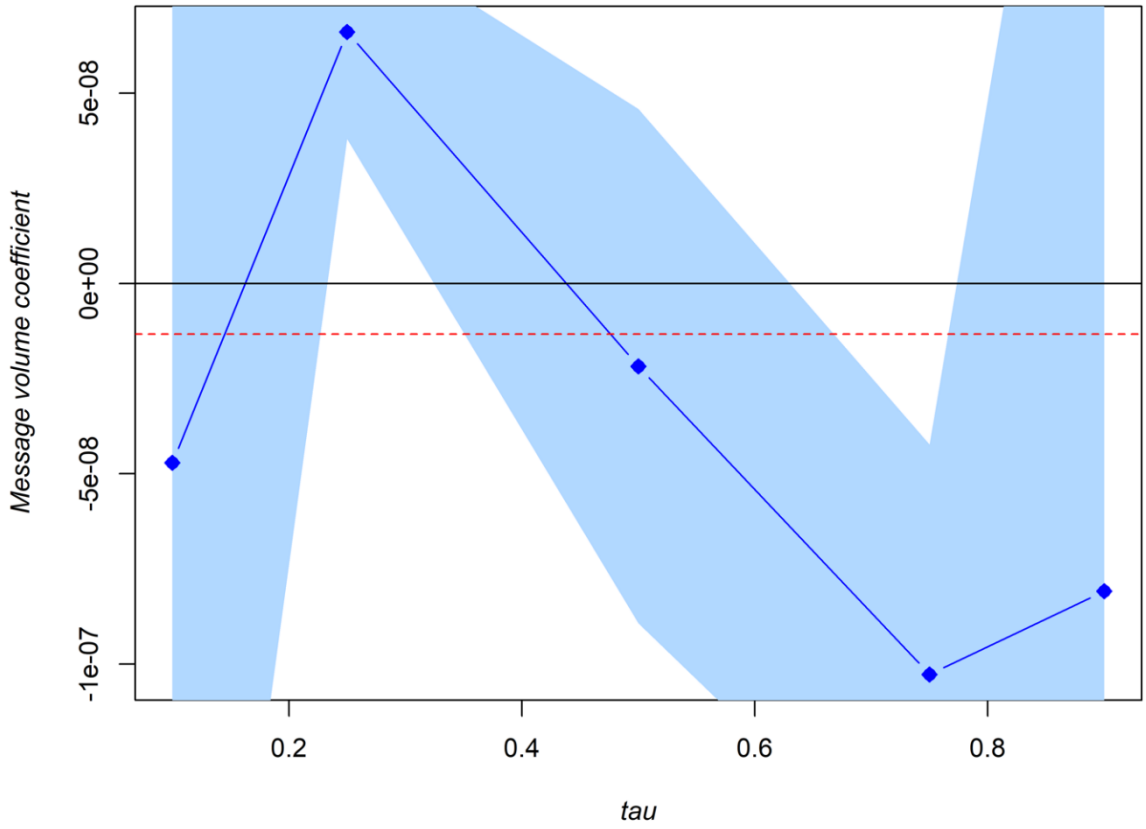


Figure 5: Visualisation of QR and OLS estimates with regard to message volume at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

Figure 5 depicts the relation between message volume and stock market returns at the specified quantiles found in Table 5. The results in Table 5 indicate that there is a minuscule relation between message volume and stock market return at each quantile. As the quantiles increase there is no obvious relation between the two variables which leads to an inconclusive description.

Majority of the quantile regression estimates lie within the band of the 95% confidence interval of the OLS regression which indicates that the estimate in the OLS in the OLS regression captures majority of the relation between message volume and stock returns under a wide variety of situations. The coefficient estimate of message volume is observed to reach its minimum in the upper most quantiles ($\tau\epsilon[\ 0.75; 0.9]$), whilst the coefficient estimate of bullishness is seen to reach its maximum within these same quantiles. These findings infer the presence of an inverse relationship between message volume and bullishness during times in which returns are observed to be at their peak. These results are consistent with that of Tetlock (2007), who postulated that information posted on social media is likely to already be known and taken into account. Thus, there would be little impact on the overall bullishness as this is "old news".

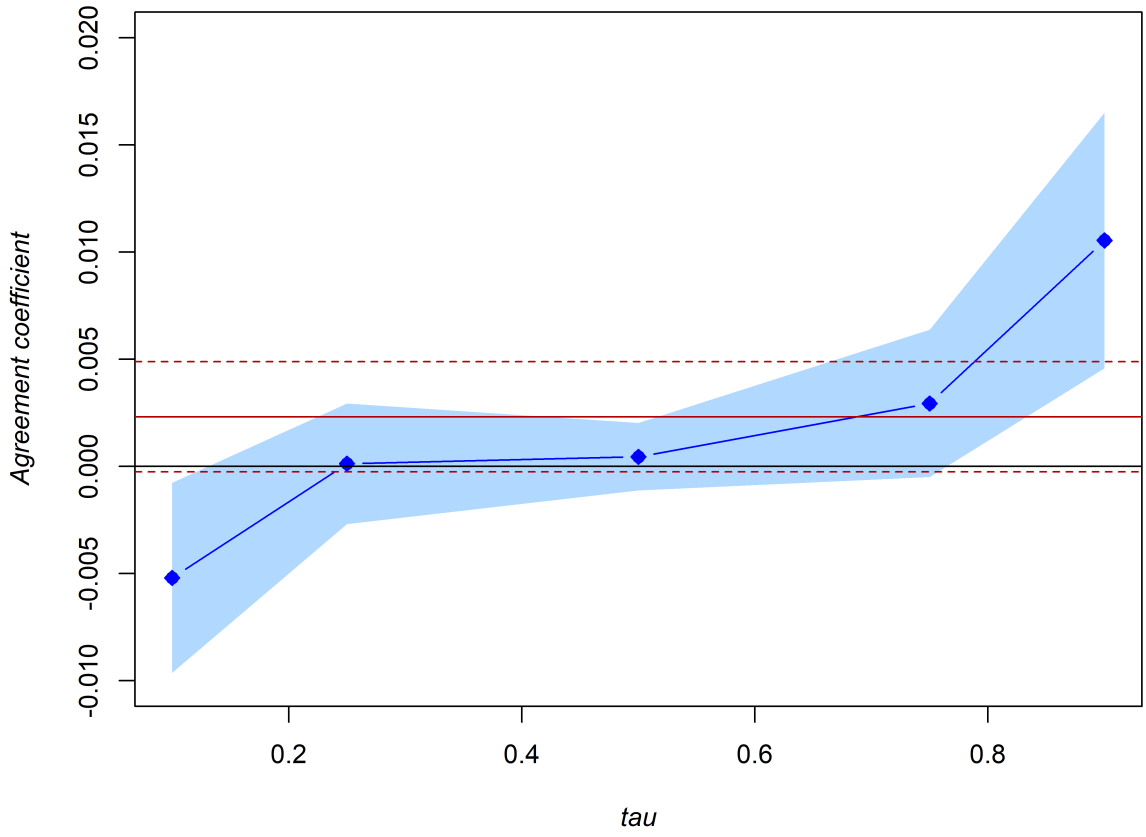


Figure 6:Visualisation of QR and OLS estimates with regard to the agreement index at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

Figure 6 presents the visualisation of the relationship between the message agreement index and stock returns at the 5 quantiles referenced in Table 5. As can be seen in both Figure 6 and Table 5, the strength of the relationship (β) between the message agreement index and the stock return quantiles is observed strengthen as the quantiles increase implying the existence of a monotonic relationship between the two variables. These findings are consistent with that of Diether, Malloy and Scherbina (2002), who describe the relationship between message agreement and stock returns as negative during times when overpricing is mitigated or corrected, thus eroding excess returns. in addition they note that this relation is positive during times in which overpricing exists. At ($\tau\epsilon[0.1]$) the estimated coefficient of the agreement index (β) is approximately -0.005, however this coefficients lies outside the 95% level of confidence of the ordinary least squares regression which yet again implies that the coefficients within the OLS regression may be subject to some form of bias. The strength of coefficients are observed to flatten out or plateau between ($\tau\epsilon[0.25]$) and ($\tau\epsilon[0.50]$) ,with an absolute change of approximately + 0.0003. These estimates and that of the 4th quantile ($\tau\epsilon[0.75]$) are seen to overlap with the 95% confidence interval of the OLS estimate as they lie between the two dotted red lines. In the upper most quantile ($\tau\epsilon[0.90]$) the estimated coefficient of the message agreement index is found to be the greatest indicating that at the positive extreme there is a relatively strong relation between message agreement and stock returns these findings indicate the presence of a monotonic relationship between the magnitude of message agreement and stock market returns. Much like the value of the estimated coefficient in the first quantile ($\tau\epsilon[0.1]$) The value in the final quantile lies outside the 95% confidence interval of the OLS regression estimate which is indicative that the OLS estimate does not accurately capture the relation between message agreement and stock returns in extreme cases. Furthermore, as the coefficient of the message agreement index reaches a maximum in the upper return quantile ($\tau\epsilon[0.9]$), so too does the estimated coefficient of bullishness which corroborates the findings of Cujean and Hasler (2017). These results suggest the presence of a positive relationship between message agreement and bullishness. In times of downturn, such as that presented in the first quantile ($\tau\epsilon[0.1]$), it is likely that information posted on social media is more likely to split opinion and lead to an increase in disagreement as opposed to during periods of high returns, where individuals are observed to agree to a higher extent.

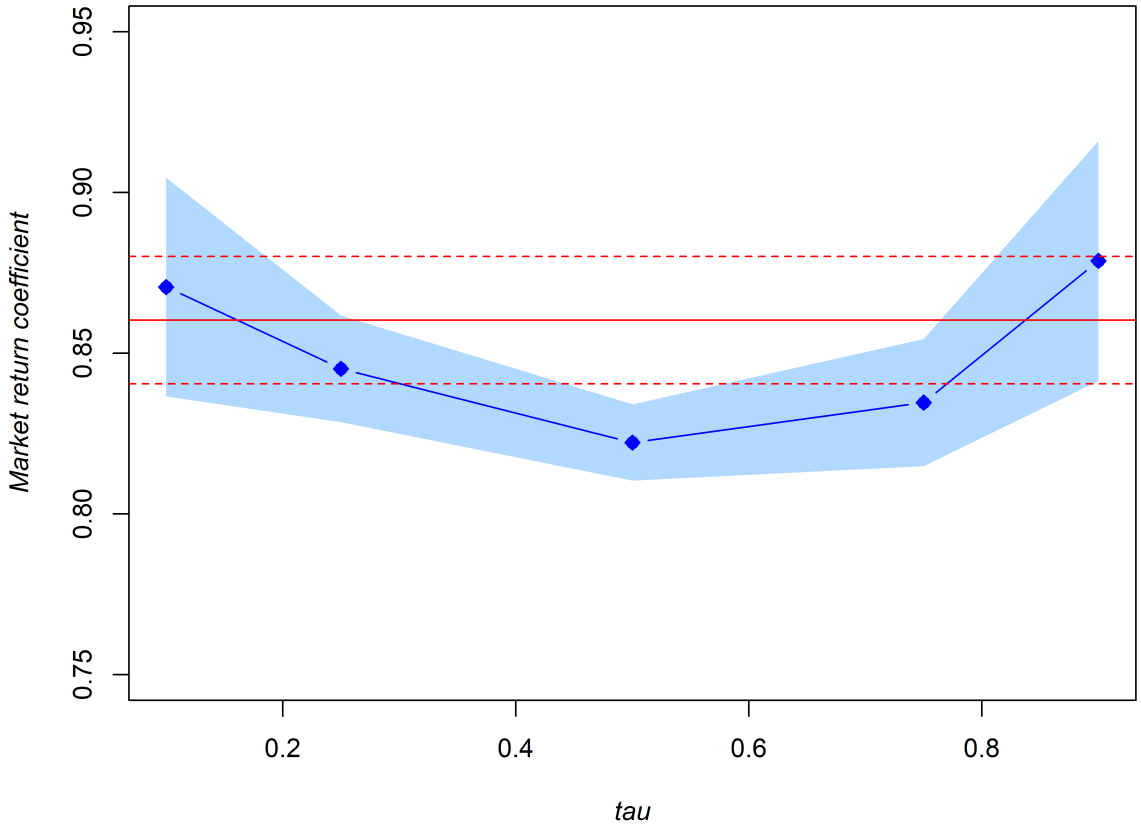


Figure 7:Visualisation of QR and OLS estimates with regard to market return at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

Last but by no means least, is the analysis of the relationship between the control variable, market return and that of stock return at the generic quantiles outlined in Table 5. Market return is observed to be statistically significant a 99% confidence interval across all of the quantiles. The estimated coefficient reaches a minimum during the median return quantiles $\tau\epsilon[\,0.25, 0.5, 0.75])$, and is

greatest at the extreme quantiles $\tau\epsilon[\,0.1, 0.9])$, with a numerical value of 0.871 and 0.879 respectively. These findings suggest that the relationship between stock returns and the market is strongest during periods of downturn and upturn, whilst this relationship is observed to weaken during periods of normal economic conditions. In addition to these findings, Figure 7 indicates that majority of the estimated coefficients in the quantile regression sit with the 95% confidence interval of the OLS regression, suggesting that the estimated coefficient in the OLS regression account for majority of the relation between market returns and stock returns although it is seen to weaken during periods of typical performance.

It is worth noting that the presence of a significant constant in both the upper and lower quantiles in table 5 indicates that additional variables are required to fully explain the association between the size of stock returns and that of *tweet* features during times of both bad times and prosperous times.

4.1.5. The relationship between the past value of *tweet* features and stock returns

Sections 4.1.2 to 4.1.3 play a vital role in identifying and quantifying the relationship between *tweet* features and market features, however they do not fully address the primary hypothesis of this study which is whether or not the value of **past** information content present in *tweet* features has any bearing or effect on stock returns. Bollen, Mao & Zeng (2011) note that the most appropriate test for this measure is that of a Granger non-causality test, as this test has the ability to indicate if the past value of a variable (*tweet* features) holds information that can be used to predict another variable (stock returns). An often-overlooked prerequisite regarding the Granger non-Causality test, is to make sure that the panel data is stationary. To accomplish this this study employs two methods (Augmented Dickey Fuller (ADF) and Phillips Perron (PP)) to test for unit roots. If it is concluded that the data is non-stationary and thus has a unit root, it may be subject to a systematic pattern that is unpredictable otherwise known as a random walk. Both the ADF and PP test utilise the Dickey Fuller (DF) test as a foundation. The ADF was created in order to rectify the issue of serial correlation which is an inherent flaw in the normal DF test, it can also deal with bigger more complex models than that of the DF. Similarly the PP test aims to correct for both

autocorrelation and heteroscedasticity in the errors (Glen 2016). The results of the panel unit root tests are presented in Table 6:

Table 6: summary of panel unit root tests (full sample)

Method	*Tweet Feature*	Statistic
ADF	Bullishness	-16.337**
PP	Bullishness	-15997**
ADF	Message Agreement	-10.621**
PP	Message Agreement	-11692**
ADF	Message Volume	-14.906**
PP	Message Volume	-16096**
ADF	Stock Return	-25.39**
PP	Stock Return	-16908**

Note: *1. The null hypothesis is that the variable has a unit root. The alternate hypothesis is that the variable does not have a unit root.*

*2. *= p<0.1; **=p<0.05; ***=p<0.01*

The results in Table 6 indicate that one should reject the null hypothesis that the variable(s) has a unit root at the 95% confidence interval based on the output for both the ADF test and PP test for all variables in the sample. Thus the bullishness, message agreement, message volume and stock return variables are shown to be stationary ($I(0)$) processes. These results ensure that the Granger non-causality test will not produce bias results. Table 7 presents the results of the Granger non-causality tests with a lag of 1 day.

Table 7: Summary of Granger Non-Causality Tests (full sample)

Null Hypothesis	Observations	Lag	F	Pr(>F)	Causality?
Bullishness →Returns	12518	1	0.034	0.853	NO
Agreement →Returns	12484	1	2.520	0.112	NO
Message Volume →Returns	12518	1	3.617	0.057	NO
Returns →Bullishness	12518	1	164.457	0	YES
Returns →Agreement	12484	1	11.843	0.001	YES
Returns → Message Volume	12518	1	0.093	0.761	NO

Note: *1. The null hypotheses are: (1) that the information contained in tweet features do not cause returns and (2) stock returns do not cause the information contained in future tweet features.*

2. The alternate hypotheses are: (1) that the information contained in tweet features do cause returns and (2) stock returns do cause the information contained in future tweet features.

The rejection criteria is set at the 5% level which is akin to that of previous studies in this field (Kim and Kim ,2014). Based on the results in Table 7 the first null hypothesis for (the information contained in *tweet* features cause returns) is accepted as the p-values for each *tweet* feature is greater than that of the rejection level (0.853>0.05; 0.112>0.05; 0.057>0.05). Thus, there is observed to be no causal relationship between the *tweet* features and stock returns in the sample. In contrast, Table 7 reveals that the second null hypothesis (stock returns do not cause the information contained in *tweet future* features) with regard to bullishness and message agreement is rejected at the 5% criterion (0<0.05; 0.001<0.05) in favour of the alternate hypothesis. This implies the existence of a causal relationship between stock returns and bullishness and stock return and message agreement; thus the lagged value of stock returns encompass information that can be useful in predicting future bullishness and message agreement metrics. The p-value regarding the relationship between stock returns message volume is greater than that of the rejection criteria (0.761>0.05) which means the null hypothesis is accepted. Therefore, stock returns are observed to have no causal effect on message volume. Overall, the results outlined in

this table are largely analogous to that of previous literature such as Kim and Kim (2014) who, using the same methodology, note that there is no evidence that suggests investor sentiment proxies are predictive regarding future stock returns.

4.2. Comparison of Pre-COVID and Intra-COVID data

Section 4.1 presented the results for the entire sample period, these results would generally be more than adequate during times of normal trading, however, it is not often that the market is subjected to an exogenous shock the with the scale and severity of the COVID-19 pandemic. This section aims to address the tertiary hypothesis put forward in section 1 , utilising a more granular analysis by splitting and comparing the pre-COVID data and intra-COVID data which is likely to be overlooked due to the aggregation present in the original sample.

4.2.1. Comparison of descriptive statistics

Table 8 exhibits the descriptive statistics over the pre-COVID period, which ranges from February 2016 up to an including December 2019. Table 9 presents the descriptive statistics for the intra-COVID period, which spans over the period of January 2020 up to and including December 2021.

Table 8: Summary Statistics of the Variables (Pre-COVID sample)

Statistic	N	Mean	St. Dev	Median	Min	Max
Tweet Features						
Bullishness	10250	0.0126	0.1076	0.0004	-0.09163	0.9898
Messages	10250	2312.0277	3639.2117	1193.5	1	142267
Agreement	10250	0.101	0.1285	0.053	0	0.9043
Market Features						
Stock Returns	14340	0.0007	0.0162	0	-0.2102	0.2609
Trading Volume	10250	29507835	42000358	15073638	153638	5334786696
Volatility	10250	0.0002	0.0004	0.0001	0	0.0099
Market Return	8167	0.0011	0.0123	0.0012	-0.055	0.0539

Table 9: Summary Statistics of the Variables (Intra-COVID sample)

Statistic	N	Mean	St. Dev	Median	Min	Max
Tweet Features						
Bullishness	5410	0.0048	0.0457	0.0017	-0.6394	0.6109
Messages	5410	1255.56	1291.4633	1044	5	29554
Agreement	5368	0.0927	0.1254	0.0442	0	1
Market Features						
Stock Returns	7570	0.0009	0.0232	0	-0.2365	0.1814
Trading Volume	5410	30729562	426186147	16837622	1587	426884920
Volatility	5410	0.0004	0.0012	0.0002	0	0.051
Market Return	4330	0.0006	0.0199	0.001	-0.1043	0.0971

In Table 8 (pre-COVID period), the bullishness index is seen to have minimum value of -0.09163 and a maximum value of 0.9898. Table 9 shows that during the COVID-19 pandemic bullishness ranges from a minimum value of -0.6394 and a maximum value of 0.6109. These results indicate that overall sentiment (bullishness) in a pre-COVID context was greater than that during the COVID-19 pandemic. This is further reiterated by the fact that in Table 8 the bullishness index has a mean value of 0.0126 whilst in Table 9 (intra-COVID period), the average magnitude this *tweet* feature is seen that is approximately 2,625 times smaller with a value of approximately 0.0048. Similarly, message volume is seen to be significantly smaller in the COVID period (with a mean value approximately 1256) as opposed to that of the pre-COVID period (with a mean value of 2312). This suggests that the aggregate number of posts referencing at least one of the firms in the study decreased during the pandemic. Additionally individuals were found to agree more under more normal or attractive market conditions than in more unfavourable times, which corroborates the findings of Cujean and Hasler (2017), who note that periods in which market conditions are unfavourable such as that of the COVID-19 pandemic it is plausible that information posted on social media is more likely to split opinion and lead to an increase in disagreement as opposed to during favourable market conditions, where individuals are observed to agree to a higher extent.

Figures 8 and 9 display the aggregate stock returns for both the pre-COVID sample and the intra-COVID sample, respectively. As seen in Table 8, the pre-sample period the maximum return was 26.09% whilst the minimum return was a low -21.02%. The maximum and minimum returns of the intra-COVID period were seen to be 18.14% and -23.65% respectively. Surprisingly, aggregate returns are approximately 0.07% over the pre-COVID period which is 0.02% less than that of the aggregate returns figure in the intra-COVID period. These findings corroborate those of Ortman, Pelster and Wengerek (2020). The standard deviation of the pre-COVID sample is an estimated 0.0162, whilst that of the intra-COVID period is 0.0232, a relative increase of approximately 43%. These findings are illustrated in figures 8 and 9 respectively. In Figure 9 (intra-COVID sample) the distribution of aggregate returns observed to have more frequent spikes and deviations from the mean value than that of the distribution of returns in Figure8 (pre-COVID sample) which is akin to the findings of previous studies (Baker, Bloom, Davis, Kost, Sammon and Viratyosin, 2020a & Zhang, Hu and Ji, 2020).

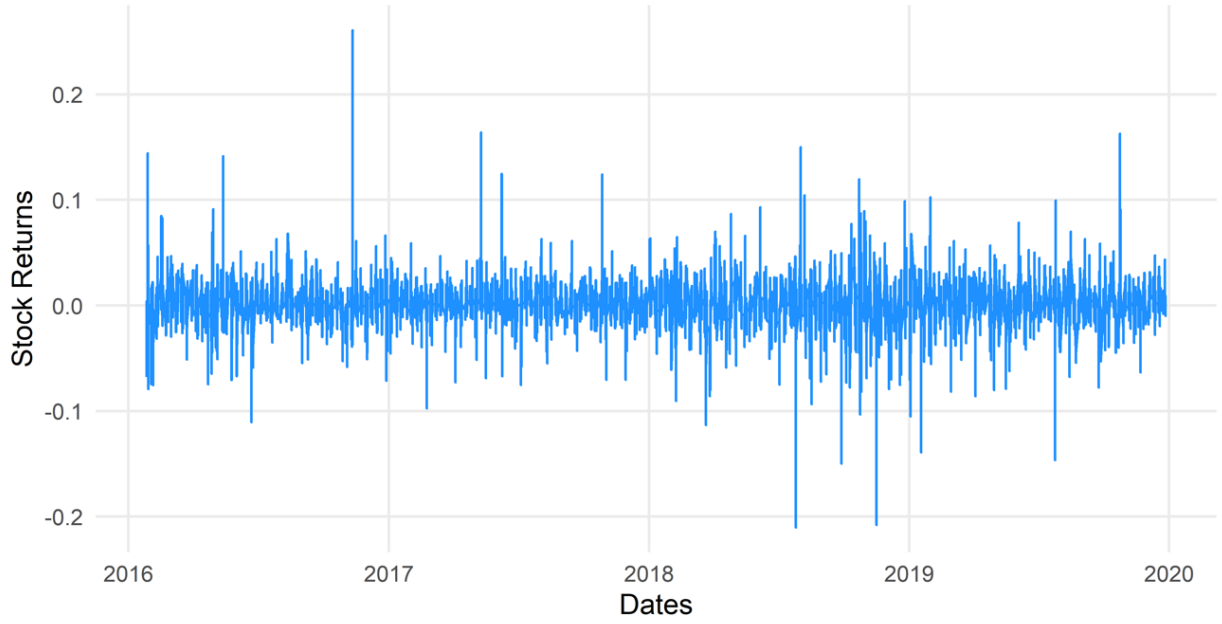


Figure 8: Distribution of Aggregate Returns (Pre-COVID sample)

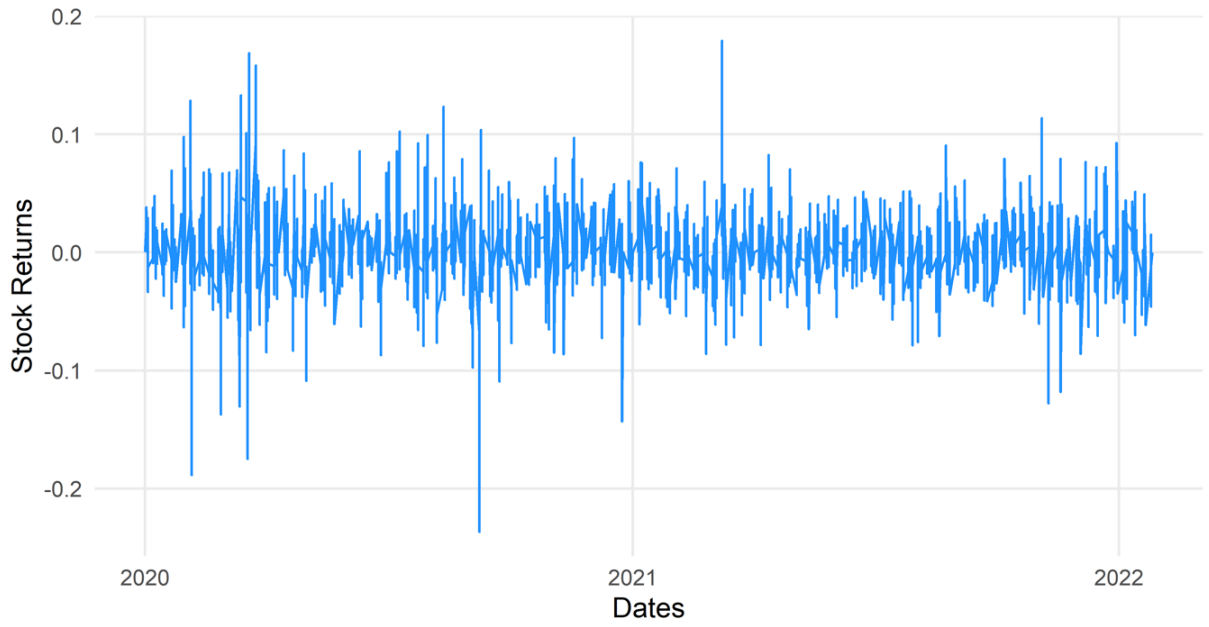


Figure9: Distribution of Aggregate Returns (Intra-COVID sample)

The aggregate trading volume is observed to have increased over the intra-pandemic period as compared to that of pre-pandemic levels (30729562>29507835) which coincides with the findings of Ortman, Pelster and Wengerek (2020) but contradicts the findings of Dow and Werlang (1992), Epstein and Wang (1994), Murkeji and Tallon (2001), who found that in the aftermath of an exogenous event such as a terrorist attack or in this case a global pandemic trading volume is seen to decrease. It is worth noting however that the minimum trading volume is dramatically smaller in the intra-COVID period (1587) in comparison to that of the minimum trading volume of the pre-COVID sample (153638). The aggregate volatility is seen to increase significantly between the two periods, with the average volatility of the pre-COVID sample measuring approximately 0.002 as opposed to that of the intra-COVID sample which is observed to have an average volatility of approximately 0.004, a figure that is two times greater than that of the previous period. These findings corroborate that of previous studies such as Zhang, Hu and Ji (2020) who note a subsequent increase in market volatility following the discovery of the COVID-19 pandemic. Additionally, in line with previous studies such as Baker, Bloom, Davis, Kost, Sammon and Viratyosin (2020a) market return decreased drastically in the intra-COVID period relative to that of the pre-COVID period (0.0006<0.0011).

4.2.2. Comparison of the contemporaneous association between market and tweet features before and during the COVID-19 pandemic

Utilising the same methodology as that of section 4.1.3, this sections aims to address the tertiary hypothesis put fourth at the beginning of the study. That is, that *Twitter is an investment tool that is utilised to a greater extent during times in which an exogeneous shock is introduced to the market.* As before a Hausman test is performed in an effort to aid in choosing the most appropriate regression model. The results are found to be almost identical to that of section 4.1.3 that is that the fixed effect models are chosen to measure both stock return and volatility and their association with the various *tweet* features which is in line with that of previous studies (Antweiler & Frank, 2004; Twedt & Rees, 2011). In contrast, the random effects model is deemed the most appropriate model to measure the relationship between the trading volume and various tweet feature and will therefore be employed.

Table 10: Contemporaneous OLS regressions (Pre-COVID sample)

	Dependant Variable		
	Return	**Volatility**	**Log(Volume)**
Bullishness	0.015***	-0.0001	-0.157***
	(0.002)	(0.00004)	(0.053)
Message Volume	-0.00000*	0.000***	0.0002***
	(0.0000)	(0.000)	(0.00001)
Message Agreement	0.003**	0.00004	0.068
	(0.001)	(0.00003)	(0.045)
Market Return	0.902***	-0.003***	-2.906***
	(0.014)	(0.0003)	(0.449)
Observations	8190	8190	8190
R^2	0.343	0.016	0.034
Adjusted R^2	0.342	0.014	0.033
F Statistic	1065.834***	32.764***	72.067***

Note*: 1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. **= p<0.1; **=p<0.05; ***=p<0.01*

Table 11: Contemporaneous OLS regressions (COVID sample)

	Dependant Variable		
	Return	**Volatility**	**Log(Volume)**
Bullishness	0.043***	-0.01*	0.584***
	(0.007)	(0.0004)	(0.187)
Message Volume	-0.00000*	0.00000***	0.0001***
	(0.0000)	(0.000)	(0.00001)
Message Agreement	0.001	-0.00000	-0.024
	(0.003)	(0.0002)	(0.072)
Market Return	0.829***	-0.002**	-1.622***
	(0.016)	(0.0001)	(0.394)
Observations	4297	4297	4297
R^2	0.404	0.015	0.079
Adjusted R^2	0.403	0.012	0.076
F Statistic	727.218***	16.841***	92.179***

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. **= p<0.1; **=p<0.05; ***=p<0.01*

In model 1, the bullishness index is observed to be positively related to returns in both the pre-COVID and intra-COVID data sets ($\beta = 0.015, p < 0.01$) & ($\beta = 0.0043, p < 0.01$) respectively. This Figure is significant at a 99% confidence interval in both instances; however, the relationship is observed to be nearly 3 times stronger in the intra-COVID period. This finding suggests that the bullishness sentiment (bullishness) has a greater effect on stock returns during the COVID-19 pandemic as opposed to before the COVID-19 pandemic and that an increase in bullishness is associated with a subsequent increase in stock returns. Message volume is seen to have any negligible effect on stock returns during both periods ($\beta = -0.0000, p < 0.1$). which

suggests that the number of posts regarding the firms on Twitter has no meaningful effect on stock returns. In the pre-COVID sample message agreement is seen to be positively related to stock returns and significant at the 5% level($\beta = 0.003, p < 0.05$), this Figure is 3 times greater than that of the intra-COVID period which is found to be non-significant at any conventional level $\beta = 0.001$). Furthermore, the findings of the pre-COVID sample are noted to be inconsistent with that of previous literature (Antweiler & Frank, 2004; Twedt & Rees, 2011).

In both cases the bullishness index is seen to have a negative relationship with volatility, in the intra-COVID case however, this relationship is far stronger ($\beta = 0.01, p < 0.1$), and significant at the 10% level of significance. These findings suggest that increase in sentiment (bullishness) is associated with a decrease in volatility to a much greater extent in during the pandemic. Message volume is once again observed to have a negligible effect on volatility, a finding that is significant at the 1% level of confidence in both periods. This suggests that the number of messages regarding the respective firm has no impact on the volatility of that firm. Message agreement is noted to have a miniscule yet positive effect on volatility in the pre-COVID sample ($\beta = 0.0004, p < 0.1$), which is consistent with that of previous literature (Antweiler & Frank, 2004 and Sprenger et al., 2014) and a negligible effect on volatility in the intra-COVID sample, however in both cases these results are non-significant.

Interestingly, level of bullishness is seen to have the opposite effect on trading volume in both periods. In the pre-COVID period bullishness has a negative effect on trading volume, whilst in the intra-COVID period bullishness has a positive effect on message volume which is coherent with the findings of previous studies (Li, van Dalen and van Rees , 2018). Additionally, message volume is seen to have a positive and statistically significant effect on trading volume in both cases ($\beta = 0.0002, p < 0.01$) & ($\beta = 0.0003, p < 0.01$), which is in line with that of Sprenger et al., (2014). The magnitude of this relationship is seen to be greater in the pre-COVID period, indicating that the number of Twitter message regarding the firms under observation had a greater effect on trading volume during this period than that of COVID period. The message agreement index is seen to have a positive relation with trading volume during the pre-COVID period $\beta = 0.068$ which is consistent with Sprenger et al., (2014) and a negative relation to trading volume

in the intra-COVID period $\beta = -0.024$. It is worth noting that these results are found to be insignificant at any conventional level.

4.2.3. Comparison of the association between the size of stock returns and that of *tweet* features before and during the COVID-19 pandemic

Similar to that of section 4.1.4, this section employs the use of pooled quantile regressions to provide a better understanding of the relation between the magnitude of stock returns and tweet features before and during the COVID-19 pandemic. Table 12 provides the pooled quantile regression estimates for the pre-COVID period and Table 13 provides the pooled quantile regression estimates for the COVID period. Furthermore, the results from the quantile regressions are given at each of most commonly specified stock return quantiles ($\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$).

Table 12: Pooled Quantile Regression Estimates (Pre-COVID sample)

	Dependent Variable				
	Stock Return				
	0.1	**0.25**	**0.5**	**0.75**	**0.9**
Bullishness	0.007***	0.004***	0.003***	0.007***	0.011***
	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
Message Volume	-0.00000	0.0000	-0.00000	-0.00000***	-0.00000
	(0.00000)	(0.000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.003	0.001	0.001	0.003**	0.011***
	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
Market Return	0.921***	0.872***	0.839***	0.858***	0.904***
	(0.023)	(0.014)	(0.008)	(0.014)	(0.022)
Constant	-0.014***	-0.007***	-0.0001	0.006***	0.014***
	(0.0004)	(0.0002)	(0.0001)	(0.0002)	(0.0004)
Psuedo R^2	0.506	0.4531	0.387	0.4534	0.4697
Observations	8190	8190	8190	8190	8190

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. *= p<0.1; **=p<0.05; ***=p<0.01

Table 13:Pooled Quantile Regression Estimates (COVID sample)

	Dependent Variable				
	Stock Return				
	0.1	**0.25**	**0.5**	**0.75**	**0.9**
Bullishness	0.014	0.015*	0.013***	0.056***	0.071***
	(0.011)	(0.008)	(0.004)	(0.008)	(0.012)
Message Volume	-0.0000***	0.0000**	-0.00000	-0.00000*	-0.00000**
	(0.00000)	(0.000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.007	-0.003	-0.003	0.006**	0.012*
	(0.005)	(0.003)	(0.002)	(0.003)	(0.007)
Market Return	0.847***	0.815***	0.797***	0.821***	0.855***
	(0.028)	(0.014)	(0.010)	(0.016)	(0.022)
Constant	-0.016***	-0.008***	-0.0001	0.007***	0.017***
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)
Psuedo R^2	0.5397	0.4959	0.4281	0.4863	0.5068
Observations	4297	4297	4297	4297	42497

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

*2. *= p<0.1; **=p<0.05; ***=p<0.01*

The bullishness index is observed to have a positive relation to stock returns across all quantiles in both the pre-COVID period and the intra-COVID period, however, this relation is far greater in the COVID period especially at the upper quantiles ($\beta = 0.0071, \tau \in [0.9]) > (\beta = 0.011, \tau \in [0.9]$). Additionally all of the estimates in the pre-COVID sample are statistically significant at a 99% confidence interval whilst in comparison, only the median and upper quantiles of the intra-COVID period are significant at this level. These findings suggest that the bullishness index has a stronger relation to stock returns in the intra-COVID period as opposed to the pre-COVID period, especially at the upper quantiles. Figures 10 and 11 reiterate these findings, suggesting that during the pre-COVID period the bullishness index's relation with stock returns is observed to be stronger

at the extremes ($\tau\epsilon[0.1, 0.9]$) and weaker during normal conditions ($\tau\epsilon[0.25,0.5,0.75]$) which accounts for the convex shape of the graph. In the intra-COVID period the graph suggests the existence of a non-linear monotonic relationship between bullishness and stock returns. It is noted that the relation between bullishness and stock returns in this period is far greater at the upper quantiles ($\tau\epsilon[0.75, 0.9]$) than that of the lower and median quantiles ($\tau\epsilon[0.1, 0.25, 0.5]$) indicating that bullishness has a much stronger association with stock returns in times where returns are high. In Figure 10 all quantile regression estimates lie outside the 95% confidence interval of OLS regression which implies that the OLS regression estimate for the pre COVID sample may be largely biased. Figure 11 indicates a similar finding, with only one of the quantile regression estimates ($\tau\epsilon[0.75]$) corresponding to OLS regression estimate for the intra-COVID period.

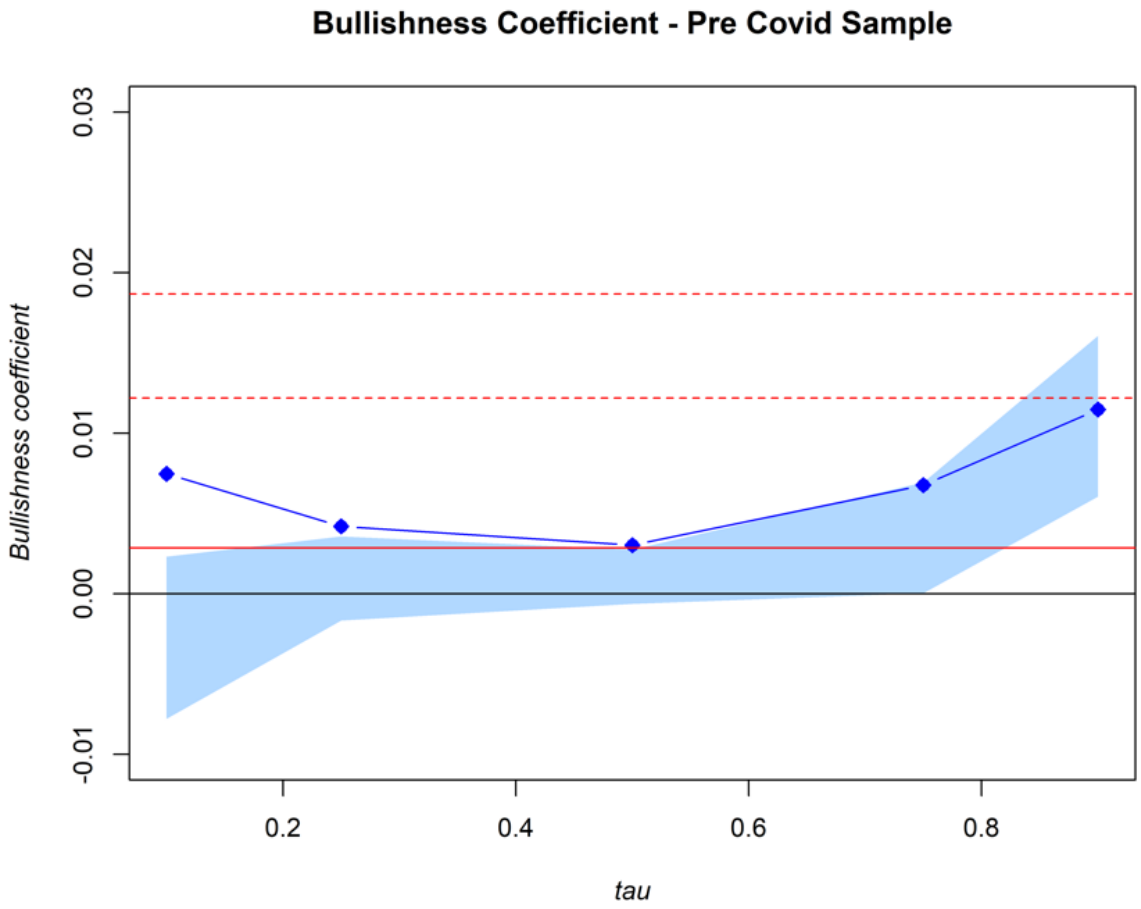


Figure 10:Visualisation of QR and OLS estimates with regard to pre-COVID bullishness at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement*[7]*. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

[7] In figures 10 to 17, five points are visualised, representing the specified quantiles as stated in Table 5, is done for ease of interpretation and coherency. To guarantee the robustness of the results in this section, an additional quantile regression is estimated using the conditional quantiles τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] which is in line with that of previous studies (Ni and Xu ,2015). Furthermore, the regression above is re-estimated using $\tau = \epsilon[0.05, 0.2, 0.4, 0.6, 0.8, 0.95]$ which is akin to that of Ma, Xiao and Ma (2018) .The results of these quantile regression are attached in appendix B.

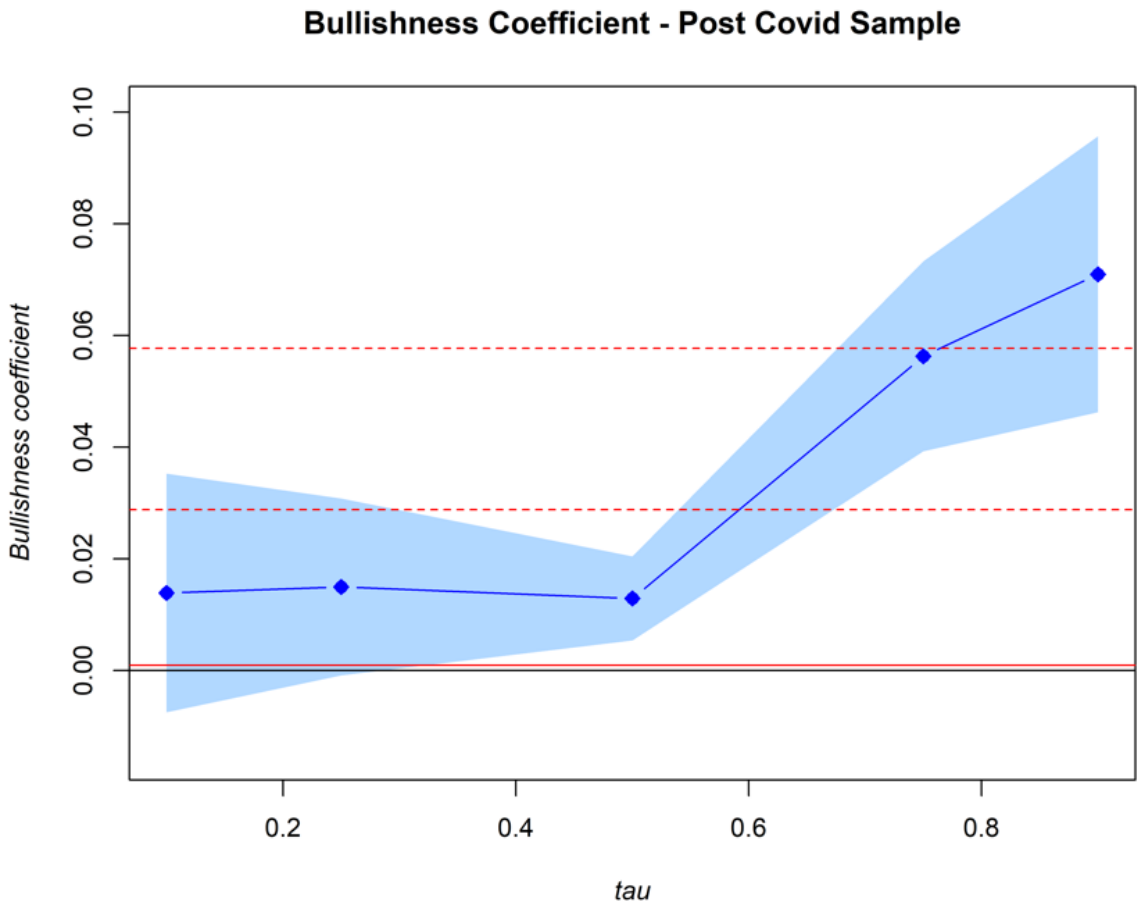


Figure 11: Visualisation of QR and OLS estimates with regard to COVID bullishness at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

Figures 12 and 13 depict the relation between message volume and stock market returns at the specified quantiles for the pre-COVID sample and intra-COVID sample respectively. Message volume is observed to have an infinitesimal association with stock returns in both periods. These results are consistent with that of previous studies. In the intra-COVID sample only two of the five

quantile regression estimates($\tau\epsilon[0.25, 0.5]$) lie within the 95% confidence interval of the OLS regression which indicates that the OLS regression estimate is biased and likely to accurately capture the association between message volume and stock returns during times of normality. It is worth noting that during the intra-COVID period, both the bullishness index and message volume are seen to reach their highest values in the uppermost quantile indicating the presence of a direct relationship which is inconsistent with that of previous studies such as Tetlock (2007). In contrast there is no apprehensible relation between message volume and stock returns in Figure12, however message volume is seen to be at its lowest in the uppermost quantiles which suggests the existence of an inverse relationship between bullishness and message volume which is consistent with the findings of Tetlock, (2007).

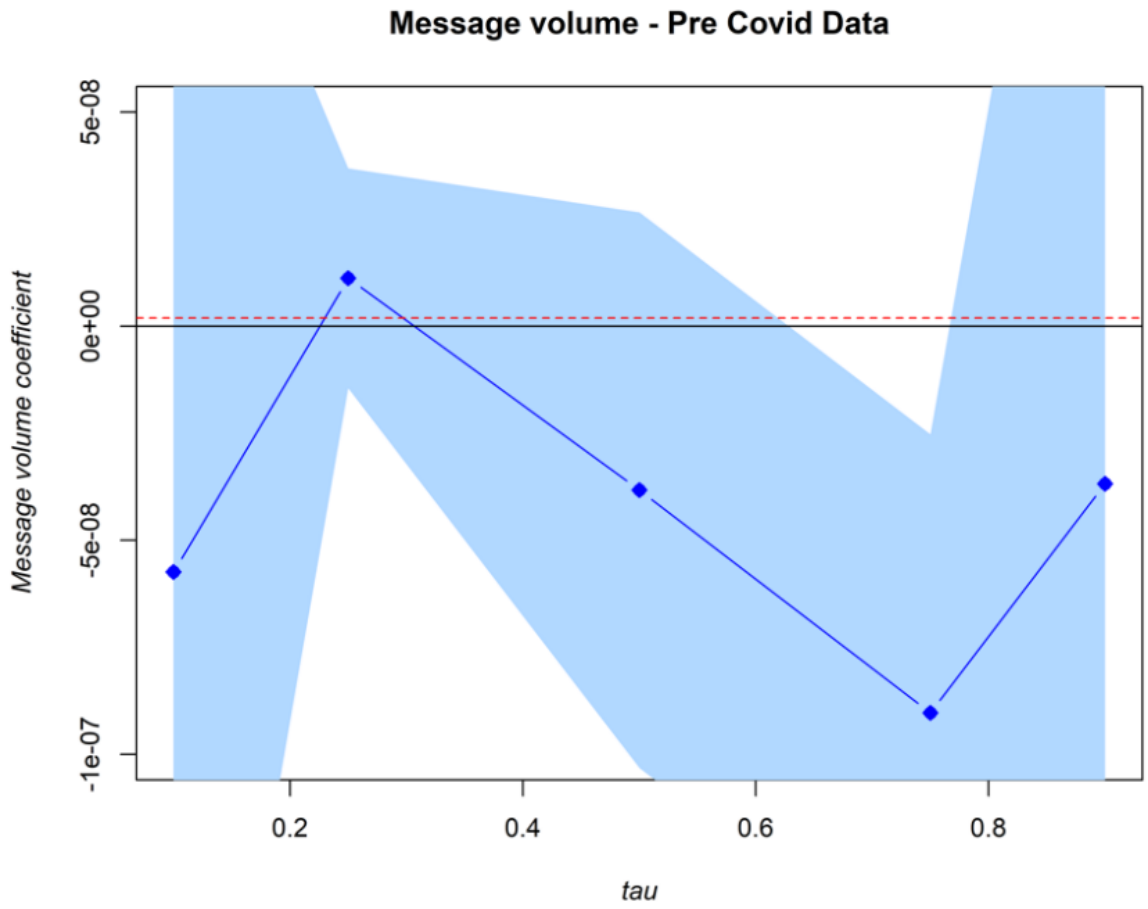


Figure 12:Visualisation of QR and OLS estimates with regard to pre-COVID message volume at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

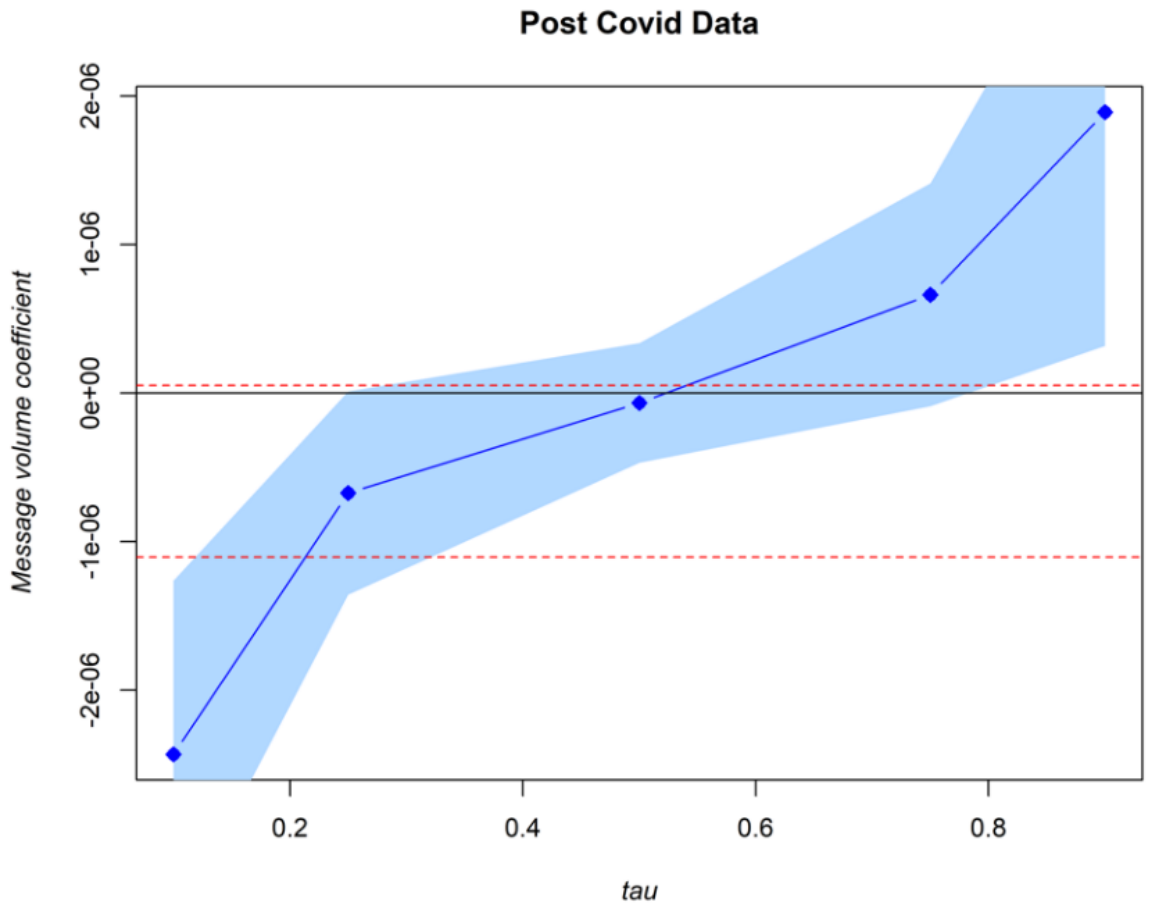


Figure 13:Visualisation of QR and OLS estimates with regard to COVID message volume at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters*

Figures 14 and 15 illustrate the relationship between the message agreement index and stock returns across 5 quantiles for the pre-COVID and intra-COVID periods respectively. In the pre-COVID period the relationship between stock the agreement index and stock returns is observed to strengthen as the quantiles increase, this leads to an overall positively sloped graph as seen in Figure 14. The estimated coefficient in the lower quantile is for this period negative ($\beta = -0.003, \tau \in [0.1]$) and positive in the uppermost quantile ($\beta = 0.011, \tau \in [0.9]$), which is line

with previous studies (Diether, Malloy and Scherbina ,2002) indicating the presence of a monotonic relationship between the two variables. These findings suggest that in times of poor performance message agreement is seen negative impact on stock returns however, this Figure is not statistically significant at any conventional level. In contrast to this message agreement is observed to have the strongest relation with stock returns during prosperous times and this finding is seen to be statistically significant at 1% level of significance. A similar result is noted with regard to the intra-COVID data, that is that message agreement is seen to have an increasingly strong relation to stock returns as the quantiles increase. The estimated coefficients are negative in the lower and median quantiles $[(\beta = -007, \tau \in [0.1])$, $(\beta = -0.003, \tau \in [0.25])$, $(\beta = -0.003, \tau \in [0.5])]$ which indicates that message agreement is seen to negatively impact stock returns in bad and mediocre times. In contrast, the uppermost quantiles have positive and statistically significant coefficients which is in line with that of Diether, Malloy and Scherbina (2002). As indicated by Figures 14 and 15 majority of the estimated coefficients of the quantile regressions lie within the 95% confidence interval of the OLS regressions which indicates that the estimated message agreement coefficients of the OLS regressions capture most of relationship between the message agreement index and stock returns for both the pre-COVID and COVID periods. It is worth noting that estimated coefficients of the quantile regressions at the uppermost and lowermost quantiles lie outside the 95% confidence interval of the OLS regressions, suggesting that the estimates in the OLS regressions are unable to account for extreme events.

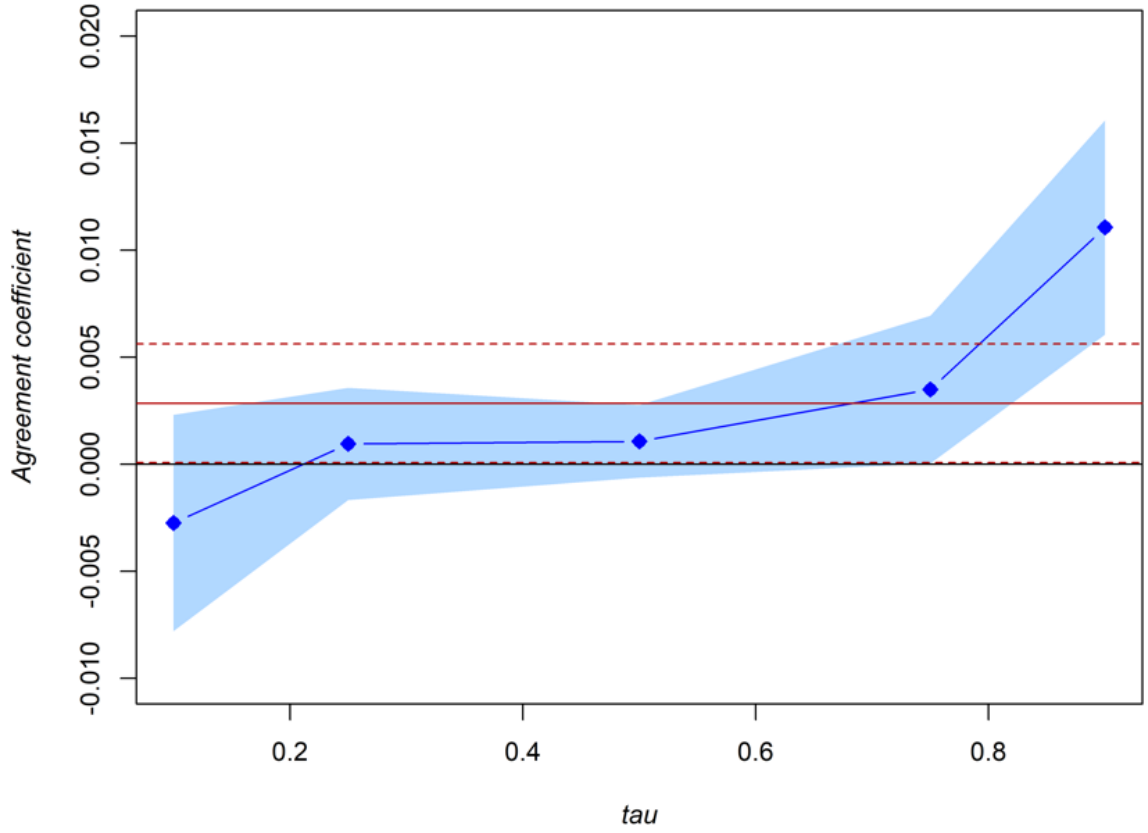


Figure 14:Visualisation of QR and OLS estimates with regard to the pre-COVID agreement index at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

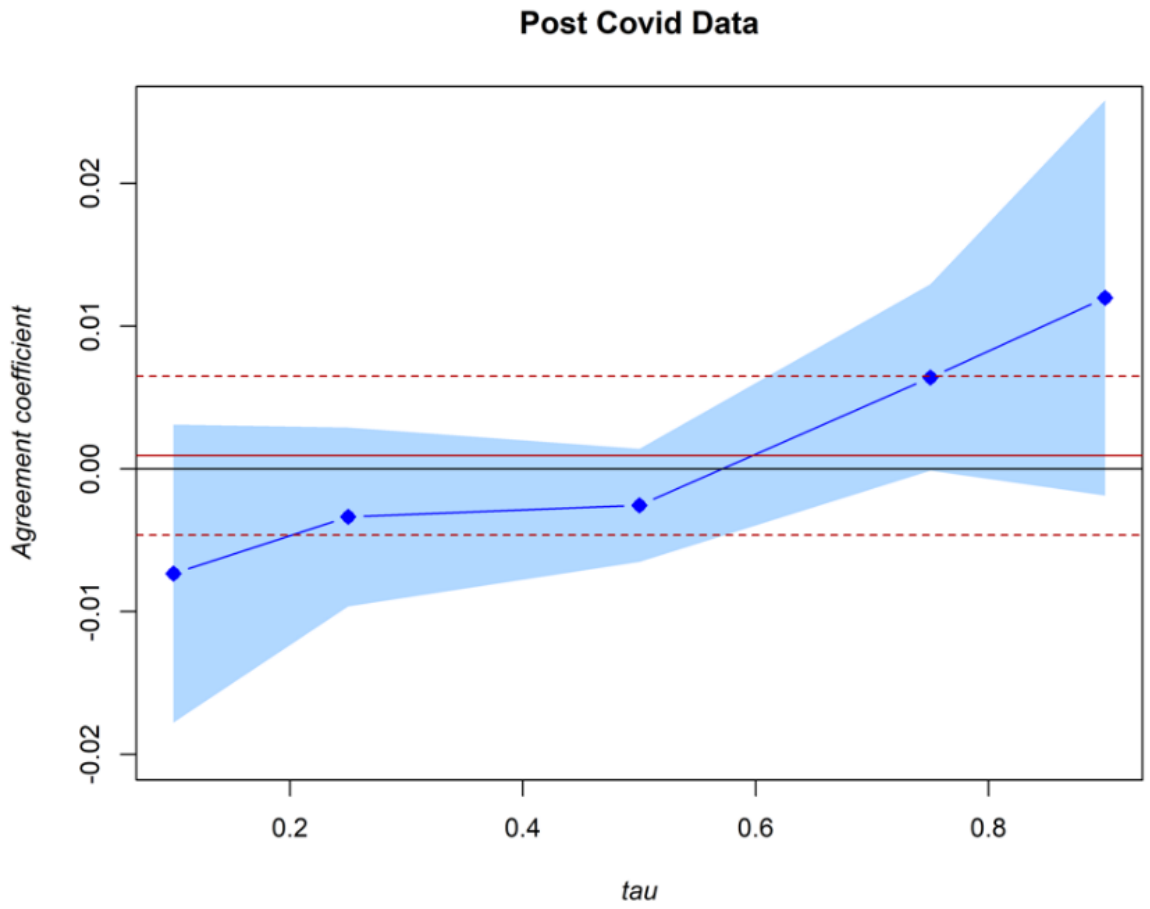


Figure 15: Visualisation of QR and OLS estimates with regard to the COVID agreement index at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

As shown in Tables 12 and 13 and further illustrated in Figures 16 and 17, market return is observed to be positive and statistically significant at a 99% confidence interval across all of the quantiles in both periods. The estimated coefficients reach a minimum during the median return quantiles $\tau\epsilon[\,0.25, 0.5, 0.75])$, and are greatest at the extreme quantiles $\tau\epsilon[\,0.1, 0.9])$. These findings suggest that the relationship between stock returns and the market is strongest during periods of downturn and upturn, whilst this relationship is observed to weaken during periods of

normal economic conditions. In addition to these findings, Figure 16 indicates that majority of the estimated coefficients in the pre-COVID quantile regression sit with the 95% confidence interval of the OLS regression, suggesting that the estimated coefficient in the OLS regression account for majority of the relation between market returns and stock returns, although it is seen to weaken during periods of typical performance. Figure 17 indicates that all of the estimated coefficients in the intra-COVID quantile regression sit with the 95% confidence interval of the OLS regression, suggesting that the estimated coefficient in the OLS regression account for all of the relation between market returns and stock returns.

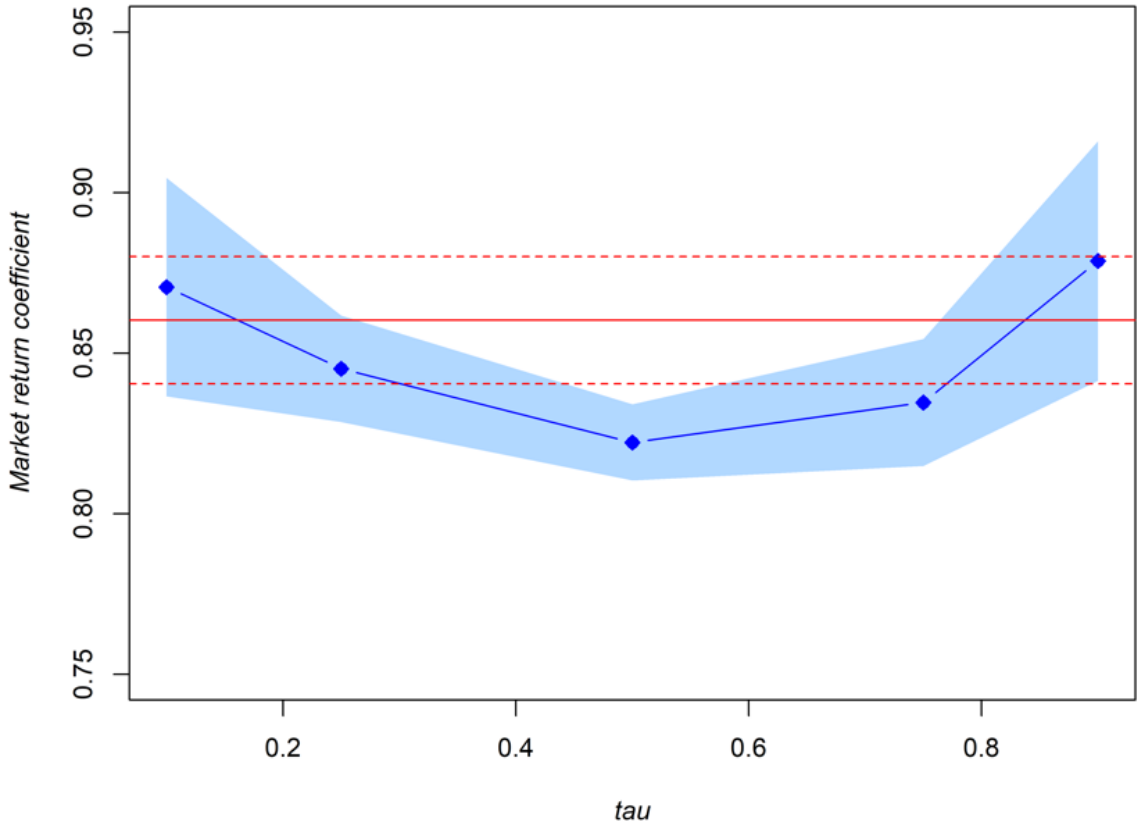


Figure 16: Visualisation of QR and OLS estimates with regard to pre-COVID market return at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

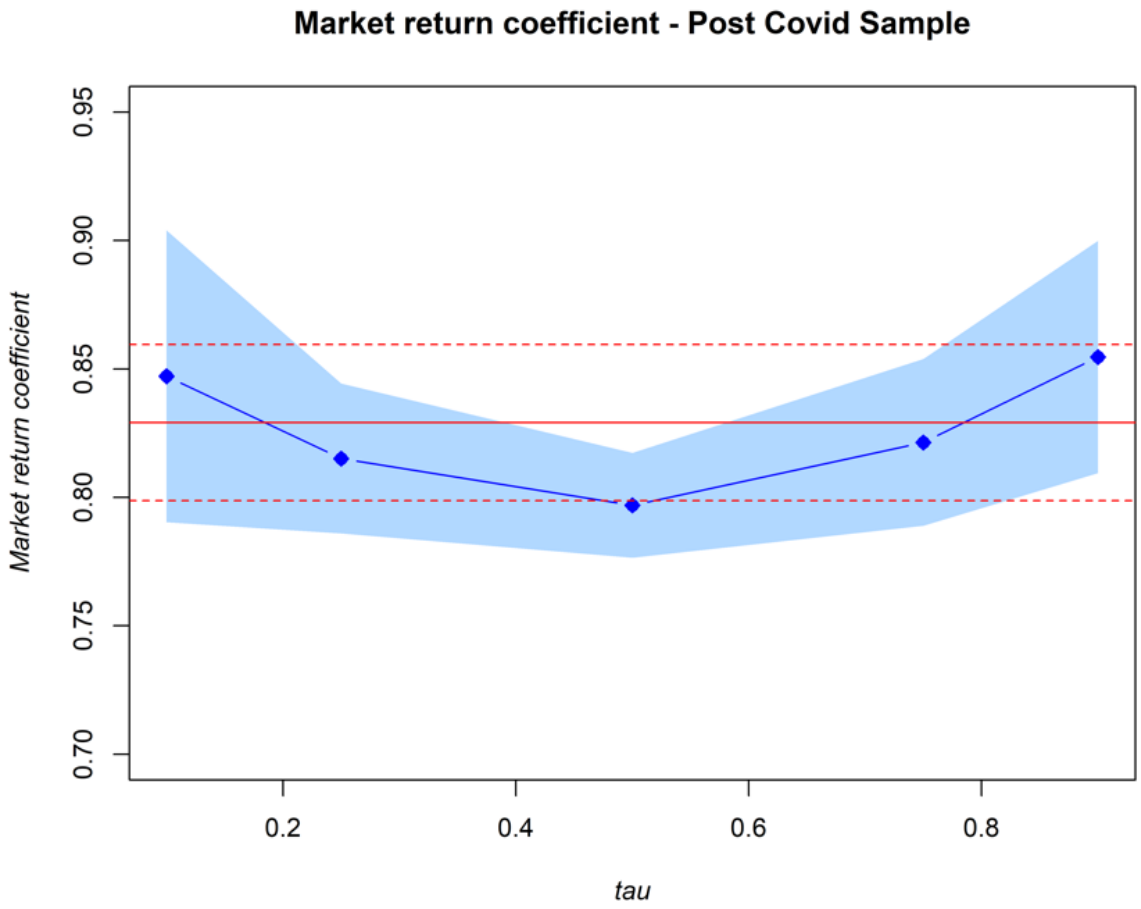


Figure 17: Visualisation of QR and OLS estimates with regard to COVID market return at a 5% level of significance

Note: *In the Figure above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile* $(\tau\epsilon[0.1, 0.25, 0.5, 0.75, 0.9]$ *of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression that was conducted in the preceding section. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement thus. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

4.2.4. Comparison of the relationship between the past value of *tweet* features and stock returns before and during the COVID-19 pandemic.

Following section 4.1.5 and that of Bollen, Mao and Zeng (2011), a Granger non-causality test is utilised to indicate if the past values of *tweet* features hold information that can be used to predict stock returns in both a pre-COVID setting and an intra-COVID setting. In section 4.1.5 panel unit root tests are performed to make sure that the data is stationary. For the sake of concision these results are carried over to apply to each of the individual data sets. Table 14 and 15 display the results of the granger causality tests for the pre-COVID period and intra-COVID period respectively.

Table 14: Summary of Granger Non-Causality Tests (Pre-COVID sample)

Null Hypothesis	Observations	Lag	F	Pr(>F)	Causality?
Bullishness →Returns	8178	1	0.097	0.755	NO
Agreement →Returns	8178	1	1.264	0.261	NO
Message Volume →Returns	8178	1	2.281	0.131	NO
Returns →Bullishness	8178	1	117.741	0	YES
Returns →Agreement	8178	1	1.197	0.274	NO
Returns → Message Volume	8178	1	0.163	0.686	NO

Note: *1. The null hypotheses are: (1) that the information contained in tweet features do not cause returns and (2) stock returns do not cause the information contained in future tweet features.*

2. The alternate hypotheses are: (1) that the information contained in tweet features do cause returns and (2) stock returns do cause the information contained in future tweet features.

Table 15: Summary of Granger Non-Causality Tests (COVID sample)

Null Hypothesis	Observations	Lag	F	Pr(>F)	Causality?
Bullishness →Returns	4237	1	0.038	0.846	NO
Agreement →Returns	4292	1	0.903	0.342	NO
Message Volume →Returns	4237	1	5.392	0.015	YES
Returns →Bullishness	4237	1	148.926	0	YES
Returns →Agreement	4292	1	14.169	0.002	YES
Returns → Message Volume	4237	1	5.978	0.015	YES

Note: *1. The null hypotheses are: (1) that the information contained in tweet features do not cause returns and (2) stock returns do not cause the information contained in future tweet features.*

2. The alternate hypotheses are: (1) that the information contained in tweet features do cause returns and (2) stock returns do cause the information contained in future tweet features.

As before the rejection criteria is set at the 5% level which is akin to that of previous studies in this field (Kim and Kim, 2014). The results in Table 14 suggest that the first null hypothesis (the information contained in *tweet* features cause returns) is accepted as the p-values for each *tweet* feature is greater than that of the rejection level (0.755>0.05; 0.261>0.05; 0.131>0.05). Thus, there is observed to be no causal relationship between the *tweet* features and stock returns in the pre-COVID sample. The second null hypothesis (stock returns do not cause the information contained in *future tweet* features) with regard to bullishness is rejected at the 5% criterion (0<0.05;

0.001<0.05) in favour of the alternate hypothesis. This implies the existence of a causal relationship between stock returns and bullishness; thus the lagged value of stock returns encompass information that can be useful in predicting future bullishness metrics. The P-value regarding the relationship between stock returns message agreement and stock returns and message volume is greater than that of the rejection criteria (0.274>0.05, 0.686>0.05) which means the null hypothesis is accepted. Therefore, stock returns are observed to have no causal effect on message agreement nor message volume. Overall, the results outlined in this Table are largely analogous to that of previous literature such as Kim and Kim (2014) who, using the same methodology, note that there is no evidence that suggests investor sentiment proxies are predictive regarding future stock returns.

Based on the results in Table 15 the first null hypothesis for (the information contained in *tweet* features cause returns) with regard to bullishness and message agreement is accepted as the p-values for these *tweet* features is larger than that of the rejection criteria (0.846>0.05; 0.342>0.05). Thus, there is observed to be no causal relationship between the bullishness index and stock returns and message agreement and stock returns in the sample. In contrast to the two *tweet* features above and the results in the pre-COVID period, there is observed to be a causal relation between message volume and stock returns in the intra-COVID sample. Furthermore, Table 15 reveals that the second null hypothesis (stock returns do not cause the information contained in *future tweet* features) with regard to all *tweet* features is rejected at the 5% criterion (0<0.05; 0.0002<0.05; 0.015>0.05) in favour of the alternate hypothesis. This implies the existence of a causal relationship between stock returns and *tweet* features; thus, the value of stock returns in the intra-COVID period encompass information that can be useful in predicting future bullishness, message agreement and message volume metrics. The results outlined in this Table are found to oppose that of previous literature such as Kim and Kim (2014). A potential explanation for this may lie within a concept postulated by Shiller (2003), known as price-to-price feedback theory. This theory states that as speculative prices rise, thus, generating profits for some investors, this may attract public attention and increase word of mouth or in this case text-based enthusiasm/positive sentiment. Measures introduced during the height of the COVID-19 pandemic may have contributed to this. The implementation of social distancing, self-isolation and working from home, meant that individuals found themselves in a unique situation in which they were afforded more time in

isolation and as such they were more likely to utilise social media as a platform on which they could add their opinion or commentary on the events that transpired around them.

5. Conclusion

The primary purpose of this study was to examine the ability of textual analysis to determine the information content of Twitter posts in predicting returns within the technology sector. Furthermore, the extent of the relationship between *tweet* features and stock market returns the technology sector was examined during a period of unrivalled volatility across markets which was brought about by the coronavirus disease (COVID-19). With this in mind, three hypotheses were postulated, namely; *Past values of tweet features contain useful information that could be used to predict future returns; the size of firm-level tweet features is monotonically related to the size of market features within the technology sector;* and finally *Twitter is an investment tool that is utilised to a greater extent during times in which an exogeneous shock is introduced to the market.*

Utilising methodologies akin to that of previous studies in this field, this study finds that over the entire 5-year sample period, most of the *tweet* features are observed to be contemporaneously associated to the stock returns of the ten biggest technology firms by market capitalisation. Both bullishness and message agreement features are observed to have a positive relation to stock returns within the technology sector whilst message volume is observed to have an inconsequential effect on stock market returns.

Although the results above are useful in determining whether a relationship exists between *tweet* features and stock returns, they provide only surface level detail that is insufficient to address the proposed hypotheses. To provide additional detail on the relationship between *tweet* features and the stock market returns of the chosen technology firms and to ultimately address the secondary hypothesis, pooled quantile regressions are utilised. The findings from these regressions indicate that over the entire sample, there is no evidence of the existence of a monotonic relationship between the magnitude of the bullishness index and stock returns. A similar finding is noted with regard to message volume and stock returns. In contrast to the aforementioned *tweet* features, an analysis of the pooled quantile regression results regarding message agreement and stock returns reveal the existence of a non-linear relationship between the two variables. Thus, the results show the that for the most part there is no monotonic relationship between *tweet* features and technology sector stock market returns with exception of the relationship between the magnitude of message

agreement and stock returns. These findings are largely similar to those of previous studies (Diether, Malloy and Scherbina , 2002; Tetlock ,2007; Cujean and Hasler ,2017)

To address the primary hypothesis of this study, that is whether Twitter posts contain any information that could be utilised to accurately forecast future stock returns within the technology sector, panel granger non-causality tests are performed with a one-day lag. The results from this test are largely consistent with previous studies such as Kim and Kim (2014) and indicate that there is no evidence that the past values of *tweet* features contain any useful information that could be used to predict future stock returns. Interestingly, there is evidence of the existence of a causal relationship between stock returns and bullishness and stock returns and message agreement; thus the lagged value of stock returns encompass information that can be useful in predicting future bullishness and message agreement metrics. In contrast, stock returns in the technology sector are observed to have no causal relationship with message volume.

To directly address the tertiary hypothesis, this study utilises a more granular analysis by splitting and comparing the pre-COVID data and intra-COVID data that is likely overlooked due to the aggregation present in the original sample. For the sake of continuity, the same approach is used as in the full sample analysis. The descriptive statistics reveal that average message volume decreased and average trading volume increased during the COVID period. These findings corroborate those of Ortman, Pelster and Wengerek (2020) who note that an exogenous event of this nature should in fact lead to a increase in trading intensity. Furthermore, the strength of the contemporaneous relationship between bullishness and stock returns is nearly three times higher during the COVID period in comparison to the pre-COVID period. This finding suggests that the sentiment (bullishness) has a greater effect on stock returns during the COVID-19 pandemic as opposed to before the COVID-19 pandemic and that an increase in bullishness is associated with a subsequent increase in stock returns. Additionally, the strength of the contemporaneous relationship between message agreement and stock returns is three times higher during the pre-COVID period in comparison to the intra-COVID period however this result is not significant at any conventional level.

To provide additional detail on the relationship between *tweet* features and the stock market returns of the chosen technology firms preceding and during the COVID period quantile regressions are used as before. These findings suggest that the bullishness index figure has a stronger relation to stock returns in the intra-COVID period as opposed to the pre-COVID period, particularly in extreme conditions. There is no evidence of a monotonic relationship between bullishness and stock returns in either period. Message volume is observed to have an infinitesimal association with stock returns in both periods, this finding suggests that there is once again no monotonic relationship between message volume and firm specific technology sector returns. In the pre-COVID period the relationship between stock the agreement index and stock returns is observed to strengthen as the quantiles increase, indicating the presence of a monotonic relationship between the two variables. A similar result is noted with regard to the intra-COVID data, that is that message agreement is seen to have an increasingly strong relation to stock returns as the quantiles increase. Thus, the results show the that for the most part there is no monotonic relationship between *tweet* features and technology sector stock market returns in both periods with exception of the relationship between the magnitude of message agreement and stock returns.

There is observed to be no causal relationship between the *tweet* features and stock returns in the pre-COVID sample. As with the pre-COVID sample there is observed to be no causal relationship between the bullishness index and stock returns and message agreement and stock returns in the sample. In contrast to the two *tweet* features above and the results in the pre-COVID period, there is observed to be a causal relation between message volume and stock returns in the intra-COVID sample. This finding suggests that the quantity of Twitter posts relating to firms in the sample may hold information that could potentially be used to predict future stock returns. Additionally, there is evidence of the existence of a causal relationship between stock returns and all *tweet* features during the intra-COVID period; thus the lagged value of stock returns in the intra-COVID period encompass information that can be useful in predicting future bullishness, message agreement and message volume metrics. Therefore, based on these results, there is evidence that the use of Twitter as an investment tool increases following an exogenous shock to the market.

These findings align themselves with both the semi strong form of the Efficient Market Hypothesis (EMH) and the Adaptive Market Hypothesis. With regard to the semi-strong form of the EMH, it is assumed that prices encompass all publicly available information, thus attempting to beat the

market using technical or fundamental analysis will not yield any reward. Thus, information posted on social media is likely to already be known and taken into account. Thus, there would be little impact on the overall bullishness as this is “old news”. A more granular analysis of the data reveals that the market cycles between varying degrees of efficiency which entails that investors seek information during specific periods such as in the aftermath of an exogenous shock, which is validated by the increase in trading activity and subsequent increase in returns during this period.

The results in this study may prove to be beneficial to academics, institutional investors and retail investors by offering a better understanding of the use of social media as a sentiment indicator and its correlation with the technology sector on a global scale. This may lead to the enhancement of investment decisions and could provide a basis for new investment strategies. As always, there are however a few recommendations for future research.

Given the range of economic sectors that operate within any economy across the globe, future researchers may add beneficial information to this field of study by exploring the effect of social media (Twitter and others) on other sectors of the economy. Furthermore, this study observed the effect of specific *tweet* features and their relationship with stock market features on a global scale, it may be valuable to focus this analysis on specific markets to provide commentary on details that may have been overlooked in this study. Finally, although not critical to this study the issue of endogeneity was not directly addressed and may be beneficial for future academics to consider.

6. References

Antweiler, W. & Frank, M.Z. 2004. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*. 59(3):1259–1294. DOI: 10.1111/j.1540-6261.2004.00662.x.

Baker, M. & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance*. 61(4):1645–1680.

Baker, M. & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, 21*(2), 129-152. DOI: 10.1257/jep.21.2.129

Baker, S., Bloom, N., Davis, S., Kost, K., Sammon, M. & Viratyosin, T. (2020). The Unprecedented Stock Market Reaction to COVID-19. *The Review of Asset Pricing Studies, 10*(2020), 742-758.

Bachelier, L. 1901. Théorie mathématique du jeu. Annales scientifiques de l'École normale supérieure. 18:143–209. DOI: 10.24033/asens.493.

Banerjee, A. 1992. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, *107*(3), 797–817.

Black, F. (1986). Noise. *Journal of Finance, 41*(3), 529-543.

Blankespoor, E., Miller, G., & White, H. (2014). The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter. *The Accounting Review, 89*. doi:10.2139/ssrn.1657169

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

Boyd, D. M. & Ellison, N. B. (2007). Social Network Sites: Definition, History and Scholarship. Journal of Computer-Mediated Communication, 13(1), article 1.

Brown, R. (1828) Philosophical Magazine N. S., 4, 161-173.

Brown, G., & Cliff, M. (2005). Investor Sentiment and Asset Valuation. *The Journal of Business, 78*(2), 405-440. DOI:10.1086/427633

Cao, M. & Wei, J. (2005). An Expanded study on the Stock Market Temperature Anomaly. *Research in Finance, 22,* 73-112. https://doi.org/10.1016/S0196-3821(05)22003-8

Cowles, A. (1933), Can stock market forecasters forecast?. Econometrica 1(3), 309–324.

Cropper, A. (2011). *Modelling Stock Volume Using Twitter.* Submitted to Oxford University for the degree of MSc Computer Science. Retrieved from https://newshedge.com/pdf/Modelling_Stock_Volume_Using_Twitter.pdf

Cujean, J., & Hasler, M. (2017). Why does Return Predictability Concentrate in Bad Times? *The Journal of Finance,72*(6), 2717-2758.

De Bondt, W.F.M. & Thaler, R. (1985), Does the Stock Market Overreact?. *The Journal of Finance, 40,* 793-805. Retrieved from https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

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

Diether, K.B., Malloy, C.J. & Scherbina, A. (2002). Differences of Opinion and the Cross-Section of Stock Returns. *The Journal of Finance, 57*(5), 2113–2141.

Dow, J., Werlang, S. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica 60* (1), 197–204.

Epstein, L., Wang, T. (1994). Intertemporal asset pricing under Knightian uncertainty. *Econometrica 62* (3), 283–322.

Fama, E. F. (1965b). Random walks in stock-market prices. *Financial Analysts Journal, 21*(5), 55-59. Retrieved from https://www.jstor.org/sTable/4469865

Fama, E. F. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review, 10*(1), 1-21. Retrieved from http://links.jstor.org/sici?sici=0020-6598%2819690 ... O%3B2-P&origin=repec

Fama, E.F. & French, K.R. (2002). The Equity Premium. *The Journal of Finance, 57*, 637-659. Retrieved from https://doi.org/10.1111/1540-6261.00437

Fama, E.F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*. 25(2):383–417. DOI: 10.2307/2325486.

Garcia, D. & Norli, Ø. (2007). Sports Sentiment and Stock Returns. *The Journal of Finance, 62*, 1967-1998. Retrieved from https://doi.org/10.1111/j.1540-6261.2007.01262.x

Gilbert, E., & Karahalios, K. (2010). Widespread Worry and the Stock Market. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 58-65.

Glen, S. (2016). Hausman Test for Endogeneity (Hausman Specification Test). Retrieved from https://www.statisticshowto.com/hausman-test/

Glen, S. (2017). Unit Root: Simple Definition, Unit Root Tests. Retrieved from https://www.statisticshowto.com/unit-root/

Goel, A. & Mittal, A (2012). Stock prediction using Twitter sentiment analysis. Stanford University. Retrieved from http://cs229. stanford. edu/proj2011/GoelMittalStockMarketPredictionUsingTwitterSentimentAnalysis. pdf) (2012).

Hirshleifer, D. and Shumway, T. (2003), Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance, 58*, 1009-1032. Retrieved from https://doi.org/10.1111/1540-6261.00556

Ito, M. & Sugiyama, S. (2009). Measuring the degree of time varying market inefficiency. *Economic Letters, 103*(1), 62-64. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.1731&rep=rep1&type=pdf

Kahneman, D. & Tversky, A. (1974). Judgement Under Uncertainty: Heuristics and Biases, *Science, 185*(4157), 1124-113. Retrieved from https://apps.dtic.mil/sti/pdfs/AD0767426.pdf

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica. 47*(2), 263–91.

Kahneman, D. & Tversky, A. (1989). Rational Choice and the Framing of Decisions. *NATO ASI Series (Series F: Computer and Systems Sciences), 56.* Retrieved from https://doi.org/10.1007/978-3-642-74919-3_4.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute Substitution in Intuitive Judgement. *Cambridge University press.* Retrieved from https://www.researchgate.net/profile/Shane-Frederick/publication/229071271_Representativeness_Revisited_Attribute_Substitution_in_Intuitive_Judgment/links/54087a8c0cf2c48563bd6c75/Representativeness-Revisited-Attribute-Substitution-in-Intuitive-Judgment.pdf

Kamstra, M., Kramer, L., & Levi, M. (2000.) Losing Sleep at the Market: The Daylight Saving Anomaly. *American Economic Review, 90*(4), 1005-1011.DOI: 10.1257/aer.90.4.1005

Kamstra, M., Kramer, L., & Levi, M. (2003.) Winter Blues: A SAD Stock Market Cycle. *American Economic Review, 93*(1), 324-343.DOI: 10.1257/000282803321455322

Kapoor, S., & Prosad, J.M. (2017). Behavioural Finance: A Review. *Procedia Computer Science 122* (2017), 50–54.

Kearney, C. & Liu, S. (2014). Textual Sentiment in Finance: A Survey of Methods and Models. *International Review of Financial Analysis, 33*, 171–185. DOI: 10.1016/j.irfa.2014.02.006.

Kendall, M. G. (1953), The analysis of economic time-series— Part I: Prices, *Journal of the Royal Statistical Society. Series A (General) 116*(1), 11–25.

Khuntia, S. & Pattanayak, J.K. (2018). Adaptive market hypothesis and evolving predictability of bitcoin. *Economics Letters,167*, 26-28.

Kim, J.A., Shumsuddin, A. & Lim, K.P. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long U.S. data. *Journal of Empirical Finance, 18*(5), 868-879.

Kim, S.H. & Kim, D. (2014). Investor Sentiment from Internet Message Postings and the Predictability of Stock Returns. *Journal of Economic Behaviour & Organization*, *107*, 708–729. DOI: 10.1016/j.jebo.2014.04.015.

Koenker, R. (2004). Quantile Regression for Longitudinal Data. *Journal of Multivariate Analysis*, *91*(1), 74–89. DOI: 10.1016/j.jmva.2004.05.006.

Koenker, R.W. & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R. & Machado, J.A.F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association*, *94*(448), 1296–1310. DOI: 10.1080/01621459.1999.10473882.

Konstantinidis, A., & Katarachia, A., & Borovas, G., & Maria Eleni Voutsa, M.A. 2012. From Efficient Market Hypothesis To Behavioural Finance: Can Behavioural Finance Be The New Dominant Model For Investing?, *Scientific Bulletin - Economic Sciences, University of Pitesti, 11*(2), 16-26.

Landman, J. (1993). Regret: The persistence of the possible. Oxford University Press.

Lê Cook, B., & Manning, W. (2013). Thinking Beyond The Mean: a practical guide for using quantile regression methods for health services research. *Shanghai Archives of Psychiatry, 25*(1), 55-59.

Levy, O., Galili, I. (2005). Terror and Trade of Individual Investors. *The Journal of Socio Economics, 35*(2006), 980-991.

Li, X., Shen, D. & Zhang, W. (2019). How Does Foreign Sentiment Affect the Chinese Stock Markets? Some Empirical Evidence. *China Accounting and Finance Review, 20*(3), 25.

Li, T., van Dalen, J. & 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.

Lo, A.W. (2004).The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. *Journal of Portfolio Management, 30*, 15-29.

Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, Twitter and search engine data. Retrieved from http://arxiv.org/abs/1112.1051

Ma, C., Xiao, S. & Ma, Z. 2018. Investor Sentiment and the Prediction of Stock Returns: A Quantile Regression Approach. *Applied Economics*. 50(50):5401–5415. DOI: 10.1080/00036846.2018.1486993.

McIntyre, K.E. (2014). The Evolution of Social Media from 1969 to 2013: A Change in Competition and a Trend Toward Complementary, Niche Sites. *The Journal of Social Media in Society, 3*(2), 6-25. Retrieved from https://thejsms.org/index.php/JSMS/article/view/89/43

Mukerji, S., Tallon, J.M. (2001). Ambiguity aversion and incompleteness of financial markets. *Review of Economic Studies*
68 (4), 883–904

Noda, A. (2016). A test of the adaptive market hypothesis using a time-varying AR model in Japan. *Finance Research Letters, 17*, 66-71.

Ortman, R., Pelster, M., & Wengerek, S. (2020). COVID-19 and Investor Behaviour. *Finance Research Letters.* 2020;37:101717. doi: 10.1016/j.frl.2020.101717.

Paul, T. (2015). The Effect of Social Media on Trading Behavior: Evidence From Twitter. Retrieved from https://repository.upenn.edu/cgi/viewcontent.cgi?article=1134&context=wharton_research_scholars

Raghunathan, R., Pham, M.T. (1999). All negative moods are not equal: motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Processes 79* (1), 56–77.

Regnault, J. (1863). Calcul des Chances et Philosophie de la Bourse, Mallet-Bachelier et Castel, Paris. https://archive.org/details/calculdeschances00regn/page/n1.

Robinson, G. (1991). That BLUP is a good thing: The Estimation of Random Effects. *Statistical Science, 6*(1), 15-32

Saunders, E. (1993). Stock Prices and Wall Street Weather. *The American Economic Review, 83*(5), 1337-1345. Retrieved from http://www.jstor.org/sTable/2117565

Scheinkman, J.A. & Xiong, W. (2003). Overconfidence and Speculative Bubbles. *Journal of Political Economy, 111*(6).

Shefrin, H. (2001), Behavioral Corporate Finance. *Journal of Applied Corporate Finance, 14*, 113-126. https://doi.org/10.1111/j.1745-6622.2001.tb00443.x

Shiller, R. J. (2003) From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perpectives, 17*(1).

Slovic, P. (2000). The Perception of Risk. *Earthscan Publications*, London.

Sprenger, T. O., & Welpe, I. M. (2010). Tweets and Trades: the Information Content of Stock Microblogs. European Financial Management, 20(5). doi:10.2139/ssrn.1702854

Sprenger, T.O., Tumasjan, A., Sandner, P.G. & Welpe, I.M. 2014. Tweets and Trades: The Information Content of Stock Microblogs: Tweets and Trades. *European Financial Management*, *20*(5),926–957. DOI: 10.1111/j.1468-036X.2013.12007.x.

Swamy, V. & Dharani, M. (2019) Intensity and the Investor Attention Effect: A Quantile Regression Approach. *Journal of Quantitative Economics*. 18:403–423. DOI: 10.1007/s40953-019-00185-9

Tetlock, P.C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance. 62(3):1139–1168. DOI: 10.1111/j.1540-6261.2007.01232.x.

Thaler, R. (1985). Mental Accounting and Consumer Choice. *Marketing Science, 4*(3), 199-214. Retrieved from http://www.jstor.org/sTable/183904

Thaler, R. (1999). The End of Behavioural Finance. *Financial Analysts Journal, 55*(6), 12-17. DOI: 10.2469/faj.v55.n6.2310

Uruqhart, A. & Hudson R. (2013). Efficient or adaptive markets? Evidence from major stock markets using very long run historic data. *International Review of Financial Analysis, 28*, 130-142.

Wuthrich, B., Permunetilleke, D., Leung, S., Cho, V., Zhang, J., & Lam, W. (1998). Daily prediciton of major stock indices. *KDD: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, (pp. 364-368). Retrieved from https://dl.acm.org/doi/10.5555/3000292.3000361

Wysocki, P. D. (1998). *Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards.* University of Michigan Business School Working Paper No. 98025. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=160170.

Zhang, D., Hu, M., & Ji, Q. (2020). Financial Markets Under the Global Pandemic of COVID-19. Finance Research Letters, 36(2020). Retrieved from https://doi.org/10.1016/j.frl.2020.101528

7. Appendix A

7.1. Full sample robustness checks

7.1.1.1. Quantile regressions using alternative conditional quantiles

The results presented in Table 16, Table 17 and Figure 18 below uphold the findings of the original quantile regression as presented in section 4.1.4 of the study. The results observed in both Table 16 and Table 17 indicate that bullishness has a positive and significant association with stock returns, across all quantiles. The convex nature of the graph (upper left quartile of Figure 18) implies that bullishness has a stronger relationship with stock returns during times of extreme downturn or extreme prosperity in which stock return margins are increased which is the case in the original quantile regression. There is observed to be a miniscule relation between message volume and stock market return at each quantile. These findings validate the presence of an inverse relationship between message volume and bullishness during times in which returns are observed to be at their peak as was noted in the original quantile regression. Furthermore, the strength of the relationship (β) between the message agreement index and the stock return quantiles is observed strengthen as the quantiles increase, implying the existence of a monotonic relationship between the two variables once again. Market return is observed to be statistically significant a 99% confidence interval across all of the quantiles. The estimated coefficient reaches a minimum during the median return quantiles and is greatest at the extreme quantiles which is consistent with that of original quantile regression. These findings suggest that the relationship between stock returns and the market is strongest during periods of downturn and upturn, whilst this relationship is observed to weaken during periods of normal economic conditions.

Table 16: Pooled Quantile Regression Estimates (full sample) with $\tau = \epsilon[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9]$

	Dependent Variable Return								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Bullishness	0.11***	0.007***	0.005***	0.004***	0.003***	0.005***	0.007***	0.010***	0.0017***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Message Volume	-0.00000	0.00000	0.00000	-0.000	-0.00000	-0.0000*	-0.00000**	-0.0000**	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.005**	-0.003*	0.0002	0.001	0.0004	0.002	0.002**	0.005***	0.011***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Market Return	0.871***	0.857***	0.829***	0.821***	0.822***	0.819***	0.826***	0.834***	0.879***
	(0.017)	(0.010)	(0.008)	(0.008)	(0.006)	(0.008)	(0.008)	(0.012)	(0.018)
Constant	-0.015***	-0.009***	-0.006***	-0.003***	-0.0001	0.002***	0.005***	0.009***	0.016***
	(0.0004)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0004)
Psuedo R^2	0.518	0.4895	0.4494	0.4119	0.4044	0.4206	0.4539	0.4739	0.4834
Observations	12497	12497	12497	12497	12497	12497	12497	12497	12497

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. $^{*}= p<0.1$; $^{**}=p<0.05$; $^{***}=p<0.01$

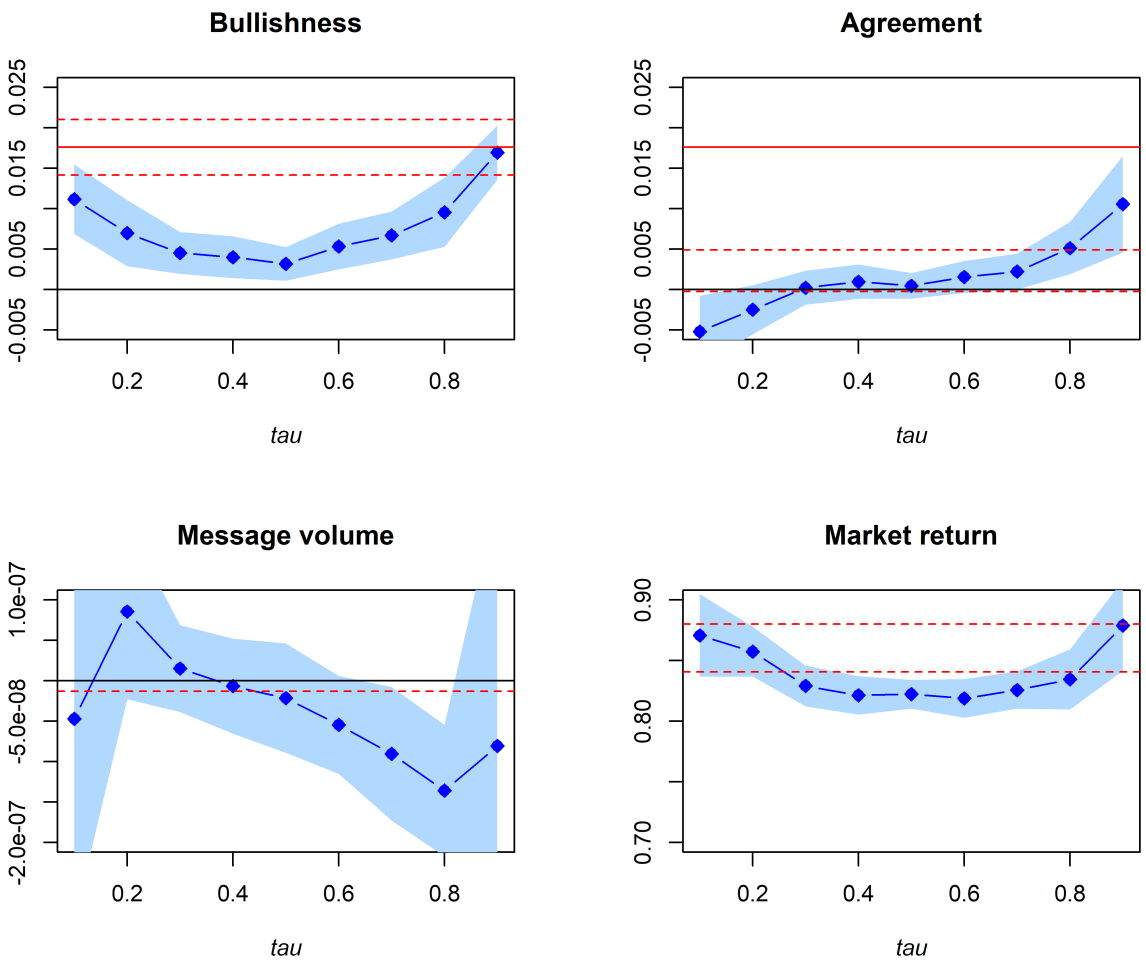


Figure 18: Visualisations of full sample quantile estimators (τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9]

Note: *In the Figures above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile (*τ*= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

Table 17: Pooled Quantile Regression Estimates (full sample) with τ=ϵ[0.05,0.2,0.4,0.6,0.8,0.95]

	Dependent Variable					
	Stock Return					
	0.05	**0.2**	**0.4**	**0.6**	**0.8**	**0.95**
Bullishness	0.016***	0.007***	0.004***	0.005***	0.010***	0.030***
	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)
Message Volume	-0.00000***	0.0000	-0.000	-0.00000*	-0.00000***	0.0000*
	(0.00000)	(0.000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.008**	0.003*	0.001	0.002	0.005***	0.021***
	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)
Market Return	0.885***	0.857***	0.821***	0.819***	0.834***	0.892***
	(0.027)	(0.010)	(0.008)	(0.008)	(0.012)	(0.032)
Constant	-0.021***	-0.009***	-0.003***	0.002***	0.009***	0.022***
	(0.001)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.001)
Psuedo R^2	0.522	0.4895	0.4119	0.4206	0.4739	0.4872
Observations	12497	12497	12497	12497	12497	12497

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

*2. *= p<0.1; **=p<0.05; ***=p<0.01*

8. Appendix B

8.1. Pre-COVID and intra-COVID robustness checks

8.1.1. Pre-COVID and intra-COVID quantile regressions using alternative conditional quantiles

The results presented in Tables 18-21 and Figures 19 and 20 below, substantiate the findings of the original quantile regression as presented in section 4.2.3 of the study. The bullishness index is observed to have a positive relation to stock returns across all quantiles in both the pre-COVID period and the intra-COVID period, however, this relation is far greater in the COVID period especially at the upper quantiles as noted before. Additionally all of the estimates in the pre-COVID sample are statistically significant at a 99% confidence interval whilst in comparison, only the median and upper quantiles of the intra-COVID period are significant at this level. These findings suggest that the bullishness index figure has a stronger relation to stock returns in the intra-COVID period as opposed to the pre-COVID period, especially at the upper quantiles. Figures 19 and 20 reiterate these findings, suggesting that during the pre-COVID period the bullishness index's relation with stock returns is observed to be stronger at the extremes and weaker during normal conditions which accounts for the convex shape of the graph. In the intra-COVID period the graph suggests the existence of a non-linear monotonic relationship between bullishness and stock returns. Message volume is, once again, observed to have an infinitesimal association with stock returns in both periods. In the pre-COVID period the relationship between stock the agreement index and stock returns is observed to strengthen as the quantiles increase, this leads to an overall positively sloped graph as seen in Figure19. The estimated coefficients are negative in the lower quantiles for this period and are positive in the uppermost quantiles which is line with the aforementioned findings indicating the presence of a monotonic relationship between the two variables. These findings suggest that in times of poor performance message agreement is seen negative impact on stock returns however, this figure is not statistically significant at any conventional level. Market return is observed to be positive and statistically significant at a 99% confidence interval across all of the conditional quantiles in both periods. The estimated coefficients reach a minimum during the median return quantiles and are greatest at the extreme quantiles .These findings suggest that the relationship between stock returns and the market is strongest during periods of downturn and upturn, whilst this relationship is observed to weaken

during periods of normal economic conditions - a finding that is identical to that of the original quantile regression in the study.

Table 18: Pooled Quantile Regression Estimates (Pre-COVID sample) with $\tau = \epsilon[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9]$

	Dependent Variable Return								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Bullishness	0.007***	0.005**	0.003**	0.003**	0.003***	0.004***	0.006***	0.008***	0.0011***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Message Volume	-0.00000	0.00000	-0.00000	-0.0000**	-0.00000	-0.00000	-0.00000***	-0.0000***	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.003	-0.001	0.001	0.002**	0.001	0.001	0.003*	0.006***	0.011***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Market Return	0.921***	0.890***	0.862***	0.850***	0.839***	0.842***	0.851***	0.856***	0.904***
	(0.023)	(0.015)	(0.012)	(0.011)	(0.008)	(0.010)	(0.013)	(0.016)	(0.022)
Constant	-0.014***	-0.009***	-0.005***	-0.002	-0.0001	0.002***	0.005***	0.008***	0.014***
	(0.0004)	(0.0003)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0004)
Psuedo R^2	0.5060	0.4735	0.4307	0.3929	0.3870	0.4058	0.4409	0.4616	0.6970
Observations	8190	8190	8190	8190	8190	8190	8190	8190	8190

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. $^{*}= p<0.1$; $^{**}=p<0.05$; $^{***}=p<0.01$

Table 19: Pooled Quantile Regression Estimates (COVID sample) with $\tau = \epsilon[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9]$

	Dependent Variable Return								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Bullishness	0.014	0.017**	0.012**	0.009*	0.013***	0.028***	0.045***	0.069***	0.071***
	(0.011)	(0.008)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)	(0.012)
Message Volume	-0.000***	-0.0000***	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000**	-0.0000***	-0.00000**
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.007	-0.006*	-0.003	-0.001	-0.003	0.001	0.002	0.005	0.012*
	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.007)
Market Return	0.847***	0.822***	0.807***	0.798***	0.797***	0.784***	0.815***	0.825***	0.855***
	(0.028)	(0.017)	(0.013)	(0.012)	(0.010)	(0.012)	(0.011)	(0.019)	(0.022)
Constant	-0.016***	-0.010***	-0.006***	-0.003***	-0.0001	0.002***	0.005***	0.010***	0.017***
	(0.001)	(0.001)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.001)	(0.001)
Psuedo R^2	0.5397	0.5133	0.4771	0.4389	0.4281	0.4418	0.4745	0.4947	0.5068
Observations	4297	4297	4297	4297	4297	4297	4297	4297	4297

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this figure in brackets represents the estimated standard error.*

2. $^{*}= p<0.1$; $^{**}=p<0.05$; $^{***}=p<0.01$

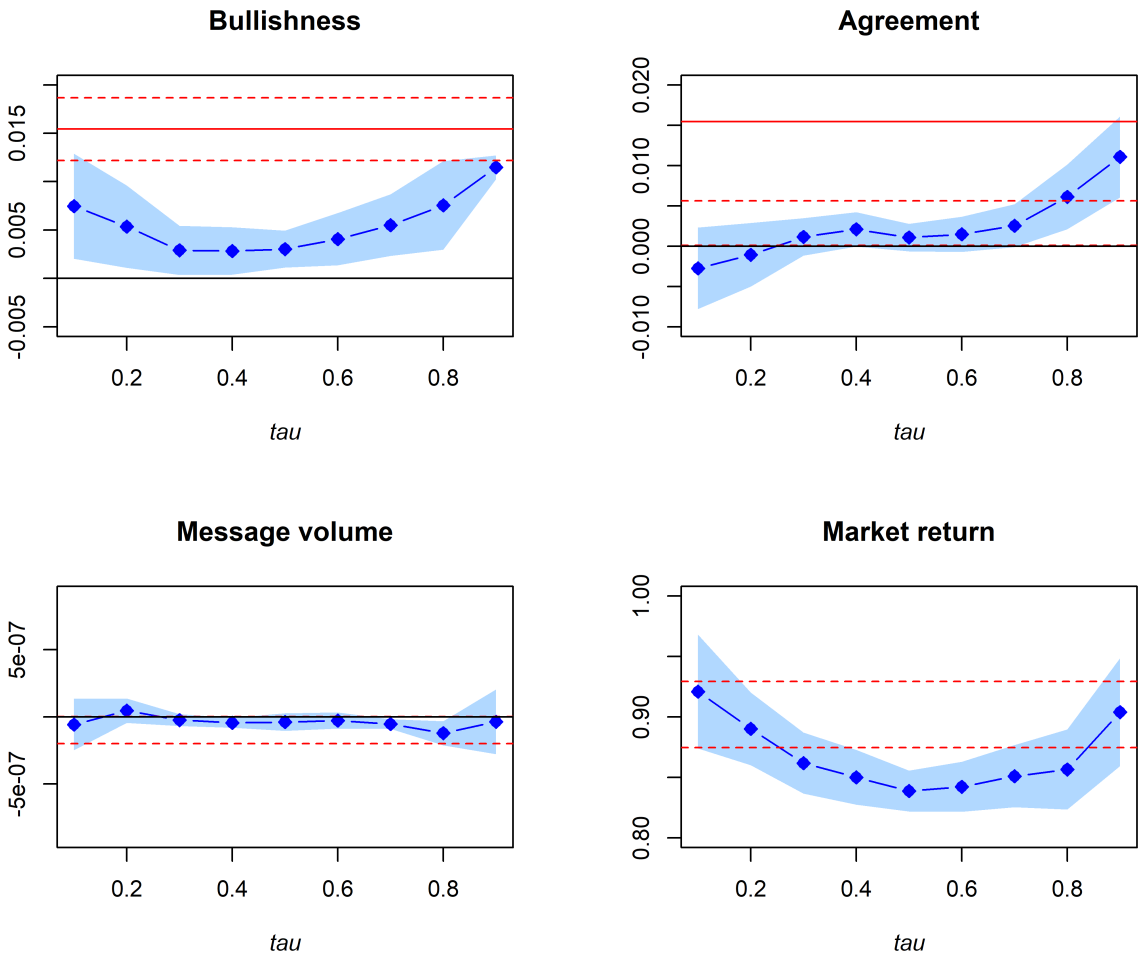


Figure 19:Visualisations of Pre-COVID sample quantile estimators (τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9])

Note: *In the Figures above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile (τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

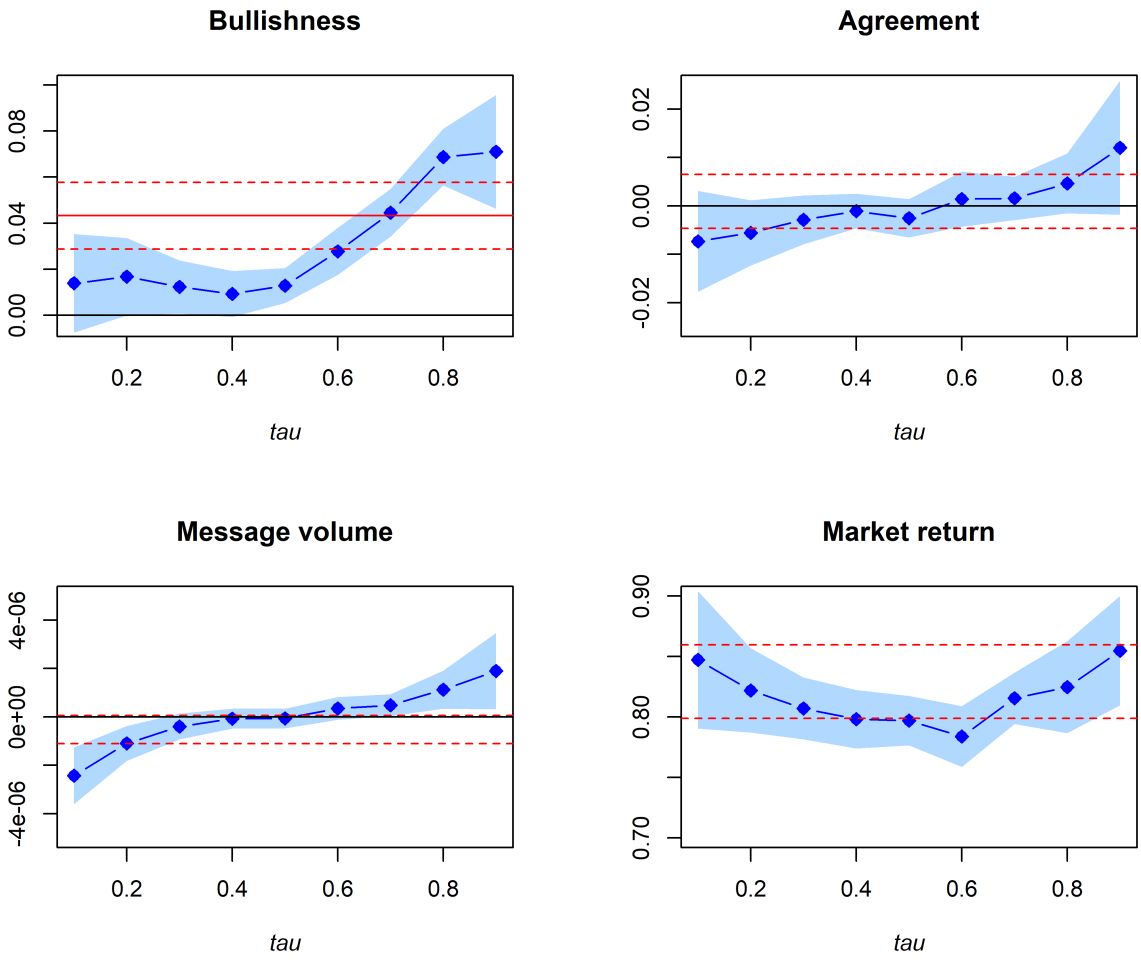


Figure 20:Visualisations of Pre-COVID sample quantile estimators (τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9])

Note: *In the Figures above, the y-axis represents numerical value of the coefficients* (β) *of the bullishness score at the specified quantile. The x-axis indicates the numerical value of the specified quantile (τ= ϵ[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9] of stock returns. The dotted red lines represents the 95% confidence interval of the ordinary least squares regression. The solid black line indicates where the vertical axis is equal to zero (0) and the solid red line signifies the actual ordinary least squares (OLS) regression. The values of the estimated* $\beta(\tau)$ *parameters from the quantile regression (QR) are connected by the solid dark blue line (each point represents a specified quantile measurement. Finally, the area of the graph shaded in light blue indicates the 5% level of significance of the estimated* $\beta(\tau)$ *parameters.*

Table 20: Pooled Quantile Regression Estimates (Pre-COVID sample) with $\tau=\epsilon[0.05,0.2,0.4,0.6,0.8,0.95]$

	Dependent Variable					
	Stock Return					
	0.05	**0.2**	**0.4**	**0.6**	**0.8**	**0.95**
Bullishness	0.013***	0.005**	0.003**	0.004***	0.008***	0.023***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)
Message Volume	-0.00000*	0.0000	-0.00000**	-0.00000	-0.00000***	0.0000*
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.009**	-0.001	0.002**	0.001	0.006***	0.021***
	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)
Market Return	0.967***	0.890***	0.850***	0.842***	0.856***	0.924***
	(0.035)	(0.015)	(0.011)	(0.010)	(0.016)	(0.033)
Constant	-0.019***	-0.009***	-0.002***	0.002***	0.008***	0.018***
	(0.001)	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.001)
Psuedo R^2	0.5089	0.4735	0.3929	0.4058	0.4616	0.4714
Observations	8190	8190	8190	8190	8190	8190

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this Figurein brackets represents the estimated standard error.*

*2. *= p<0.1; **=p<0.05; ***=p<0.01*

Table 21: Pooled Quantile Regression Estimates (COVID sample) with τ=∈[0.05,0.2,0.4,0.6,0.8,0.95]

	Dependent Variable					
	Stock Return					
	0.05	**0.2**	**0.4**	**0.6**	**0.8**	**0.95**
Bullishness	0.014	0.017**	0.009**	0.028***	0.069***	0.066***
	(0.025)	(0.008)	(0.005)	(0.005)	(0.006)	(0.024)
Message Volume	-0.00000***	-0.0000***	-0.00000	0.00000	0.00000***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Message Agreement	-0.016	-0.006*	-0.001	0.001	0.005	0.021**
	(0.010)	(0.003)	(0.002)	(0.003)	(0.003)	(0.009)
Market Return	0.824***	0.822***	0.798***	0.784***	0.825***	0.852***
	(0.041)	(0.017)	(0.012)	(0.012)	(0.019)	(0.047)
Constant	-0.022***	-0.010***	-0.003***	0.002***	0.010***	0.023***
	(0.002)	(0.001)	(0.0004)	(0.0004)	(0.0001)	(0.002)
Psuedo R^2	0.5480	0.5133	0.4389	0.4418	0.4947	0.5211
Observations	4297	4297	4297	4297	4297	4297

Note: *1. The value on the first line of each combination of tweet feature and market feature (in line with the variable name on the left) indicates the regression coefficient of each variable, the value below this Figurein brackets represents the estimated standard error.*

*2. *= p<0.1; **=p<0.05; ***=p<0.01*

WORD COUNT: [28942]