ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Bachelor Thesis Economics & Business

Specialization: Financial Economics

The Impact of Social Media and Investor Sentiment on Financial Markets

Author: Henri Snodgrass

Student Number: 611535

Thesis Supervisor: Sjoerd van den Hauwe

Second Reader: Dr. Ruben de Bliek

Finish Date: June 29th, 2024

Table of Contents

Abstract

This study investigates the effects of information and investor sentiment on stocks trading volumes and stock returns. It focuses specifically on the effects of traditional news and Twitter. The data used is from the Bloomberg Terminal, which contains the number of publications and sentiment for the news and Twitter for the 25 biggest companies in the S&P 500 from 2018-2023. This research quantifies the number of publications, as well as their sentiment, using a regression analysis. This topic was decided due to the drastic changes in information sources for stock information in the past 50 years as well as the introduction of new trading strategies, such as day trading. The findings reveal that both Twitter and traditional news publications have a statistically significant effect on the respective companies trading volumes. Additionally, this paper also finds through the sentiment analysis, that positive and negative sentiments influence daily stock returns, highlighting the role of investor sentiment on market dynamics. The idea of investor sentiment effecting market returns is nothing new; however, before AI it was extremely difficult to go through all news channels and all Tweets in order to collect enough data to confidently identify investor sentiment. These results highlight the constantly changing landscape of financial markets and the importance of ongoing research to identify new market forces.

1. Introduction

Dating back to the inception of the New York Stock Exchange in 1792, investing was a deliberate and somewhat exclusive endeavor. The process of acquiring shares involved intricate procedures, often requiring physical transactions, and limited to those of means and connections. However, with the rise of online trading through platforms like Robinhood, there is an increased access to the stock market like never before. Today, the barriers to entry have been dramatically lowered. With just a smartphone and a modest sum of cash, individuals can quickly become shareholders in esteemed companies such as Apple or Microsoft within a matter of minutes. This accessibility represents a seismic shift, empowering a broader demographic to participate in the wealth-building potential of the market.

This paper aims to explore the effects of social media and traditional news media on the respective companies trading volume and returns. The goal is to see how the modernization of modern-day communication channels has influenced and changed over the span of the stock market's recent life span.

Unlike traditional media, social media has revolutionized communication through passing on news in real time with editorial or delivery delays. Social media companies such as Twitter also have the difference that the content that is posted is created and upload by the users themselves. Many traditional medias such as the British Broadcasting Channel (BBC) in England and Al-Jazeera in Qatar are government owned news channels that are often accused of censorship to favor certain groups or agendas rather than the content being in the hands of the consumers (Safdar, 2023). In most countries, social media has opened a platform for open discussion and freedom of speech for all users. This immediacy and democratization has made Twitter a great tool to find opinions and information for trading activity and general investor sentiment. Tweets are believed by some to reflect public perception and potential market reactions to different events.

Traditional media is still a very important tool in modern day investing, being a much more reliable source of information due to certain checks and balances in most countries (Farhoudinia, Ozturkcan, and Kasap, 2024). Both sources have their advantages and disadvantages and are best used together in combination, to make sure one gets all sides of the story.

This paper quantifies and analyses the effects of Twitter and news sentiment/publications to try and use them to predict trading volume and stock returns. The new data set that is being used in the paper that has highly quantified data for sentiment and publications is what makes this paper unique with respect to past papers. Antweiler and Frank (2004) who researched the effects of messages on stock messaging boards had also quantified data but much smaller sets and much less precise. It also differs from famous papers in the subject such as Barber and Odean (2008) since these papers don't look at direct impact of a publication on returns. This paper contributes quantified research and analysis to the topic, as many papers are mainly qualitative, due human behavior being more easily discussed qualitatively since human behavior is very difficult to quantify.

The definitive research question is: “How do Twitter sentiment and news sentiment correlate with and effect trading volume and stock returns?”. By using AI, the tone and sentiment of tweets and news articles can be interpreted and quantified so the data can be used in a regression analysis. The analysis involves collecting data sentiment and publication data for the two media channels as well as company specific metrics and control variables, to be able to use advanced statistical methods and machine learning to calculate potential causal effects. It aims to provide investors and traders with insights on how non-financial market forces can affect their investments and allow them to make investment decisions with all available information and ability to understand the information.

The main findings of this paper were that both the number of Twitter and news publications have a statistically significant impact on a stock trading volume. These results allow traders to assume that if they see increased activity about a company in the news or on Twitter, that the trading volume for the company will increase accordingly. In addition, it was also found that both Twitter and news sentiment have a statistically significant impact on the company’s stock returns. This allows traders to use the strategy of looking at current news or Twitter sentiment in order to predict what direction stock prices for the companies will move in.

2. Literature Review

The field of financial literature has made many strides and experienced substantial evolution over the past decades. These advancements have been driven by new strategies, technological advancements, a deeper understanding of investor behaviour, and its effect on market dynamics. The following literature review has been organized into three parts, each covering a significant milestone/time in the development of investor strategies and their underlying academic theories.

The first section delves into the early stages of trading strategy development that differed from traditional investment strategies. Academic research in the 1980s marked the shift from traditional value investment strategies to strategies that leveraged market characteristics and certain variable relationships. This research focused mainly on finding specific patterns and variable relationships that proved to be consistent throughout time. This was the spark of algorithm and high frequency trading in the late 80s and 90s, although these came to a short halt after some notable large firm algorithm failures.

The second section focuses on the impact of the invention of the internet on trading and the impact of the increased accessibility to information. By the early 2000s, a large part of trading information was communicated via the internet, allowing real time access for all traders. The internet also facilitated and played a key role in the growth of strategies such as day trading and many over concepts. This period led to a wave of research into the effect of how media coverage and company attention influenced trading volume and returns.

The final section reviews past research that looks into the effect of the introduction of social media on the stock market and trading strategies. Social media and online chats were the first time where there was a platform for recreational investors to discuss their thoughts and strategies with thousands of others across the globe. This social interaction aspect of traders introduced even more variables into the impossible formula of market returns, such as how investor sentiment expressed on social media effects market behaviour, volume and returns. The rise of social medias effect on the market became clear during Covid-19, where increased volatility was seen on the market through events such as the *GameStop Short Squeeze*. By examining these phases one by one, the review highlights the significant steps in the evolution of investor strategies but also emphasizes the continuous adaptation and innovation required to navigate the constantly evolving financial markets.

2.1.1 Evolution of Trading Strategies: From Volume Analysis to Behavioral Insights

In financial literature, investor behaviour and attention has been a key area of study in the past 40 years. Due to investor attention theoretically leading to an increase in trading volume, which causes large price movements and volatility, it is a crucial part of the market to understand. A key steppingstone was Karpoff's

(1987) research which studied the relationship between price and volume. Karpoff's (1987) aim with his research was to disprove past literatures, that had found little or no relations between volume and price. He additionally wanted to research causal effects for increases or decreases in volume to be able to predict the effects on returns through the volume changes, allowing implementation of a strategy in real life.

Initial empirical work done by Granger and Morgenstern (1964) found there was no significant price-volume relation. Karpoff (1987) disproved these findings by finding a correlation between volume and absolute changes of price in equity and futures markets. This breakthrough provided a new lens that markets could be viewed through, turning away from traditional analysis. Upon discovering this, Karpoff (1987) synergized his research with Copeland's (1976) theory of sequential arrival of information. This theory states that traders do not receive information all at the same time but rather all at different points in time; therefore, they all act on it at different points. During the time Karpoff (1987) was conducting his research, these releases of information consisted of mainly earnings reports, economic data, and company announcements. This was mainly due to many information sources such as earnings reports and news being sent via mail, leading to everyone not always getting information on the same day or week in extreme cases.

Though not directly related, this sparked the idea for many researchers of further researching investor attention and trying to proxy investor attention in order to try and predict market returns. Through Karpoff (1987) and Copeland's (1976) research proving that there was a price-volume relationship and volume was affected by investor attention, the importance and interest of research into investor behaviour was sparked.

With groundwork having been laid, many researchers began trying to find ways to "beat" the stock market, so to speak, by using knowledge about volume and investor attention in order to execute better trades and get better prices. Leveraging insights and findings from these papers, some traders began developing more complex strategies to capitalize on market nuances and inefficiencies.

In a new paper, Jain and Joh (1988) also proved that there was a price-volume relationship and further studied volume. Jain and Joh (1988) found that trading volume was highest at market open and market close, allowing traders to take advantages of certain prices through increased volatility. This concept is a key part of today's high-frequency traders and their high-tech algorithms trading systems, which profit from and exploit these rapid fluctuations.

Before then, trading was a concept of investing in value companies and investing in their future growth; however, through papers like this, people began also trading on the market characteristics themselves. Additional research by Bazjik (1995) showed that high trading volume led to overall higher returns. This finding provided further evidence of a complex relationship between trading activity and market performance, motivating more sophisticated and complex trading approaches.

Throughout the late 80s and 90s high frequency trading and high-tech algorithms became increasingly more popular, becoming a leading strategie for many of the top hedge-funds on Wall Street being run by mathematicians, econometricians, and engineers rather than your typical finance or business graduates. One main dilemma with these new trading strategies is that once they are discovered and exploited, they disappear due to market correction. Due to this, still today many researchers and analysts are trying to find the new anomaly and system to beat the market.

As the market is an incomprehensive beast and impacted by millions of factors and variables, there are always issues with making strong assumptions. Researchers found certain causalities and correlations; however, as with all arbitrage opportunities or nuances, relationships often only last in only specific market conditions or time periods. Campbell, Grossman, and Wang (1993) found similar results to past research, that there was serial correlation between volume and returns; however, as volume increases, the predictability of using volume as a predictor of returns decreased rapidly. This meant that certain algorithms, according to past findings, would perform well in regular environment but when trading activity and volume increased, algorithms would continue with the same trading strategies, but they were no longer effective in the extreme environments. Since these trading bots were mere computers basing all actions on the thousands of lines of code they have been fed based on past data, they would not know any better than following the code that in many cases did not account for these market conditions.

One famous example of a hedge going belly up due to a failed algorithm is Long Term Capital Management (LTCM), which managed $120 bn at its peak. The fund was by John Meriweather and had assistance from Nobel Prize winners Myron Scholes and Robert Merton, however; after a failure in their algorithm due to extreme market conditions they lost $4.6 bn in less than four months, leading to a bailout from a number of banks to compensate the investors that lost large sums of money. This event in 1998 gave many investors and funds using market and investor behaviour to build models and algorithms a big wake up call. While this caused investors to take a step back from algorithm trading, academic researchers continued on their quest to finding more variables that could predict market returns.

2.1.2 The Digital Revolution in Trading: The Impact of the Internet on Market Dynamics

The next major innovation in trading was the introduction and use of the internet in the late 20th/early 21st century. This quickly solved the problem of investors receiving information at different points in time. Once the internet matured, the news and earnings reports of companies were available online, allowing all investors to react to news at the same time. This opened up a new line of research, being the effects that this new centralized platform for news has on volume and returns.

Huberman and Regev's (2001) research focused their research on how media coverage on the internet effected stock prices and found that companies with high media coverage had strong short-term price increases, followed by market corrections. Their researched found that while heavily covered companies had higher short-term returns, that companies with less media coverage performed better in the long term. Huberman and Regev's (2001) conclusion was that investor behavior and sentiment led to temporary mis-pricings and needed market correction. Their research went to show that it is crucial to understand stock market dynamics and investors reactions to news.

Chan (2003) studied the effects of news and no-news on stock prices and also looked at price reversals and drifts after headlines. Chan (2003) found that after there were news published about a specific company that their stock price would drift into a certain direction for some time and that if there were no substantial news, that these drifts would correct themselves over time as the market adjusts. This indicates that it takes time for investors to fully digest the information and that the stock continues to drift for some time. This self-correction indicates a reversion to the mean, caused by investors reevaluating the stock after an absence of information to back up the news and it turned out to be a rumor or buzz. Chan (2003) also found that these drifts tended to be stronger when the trading volume was higher. These findings supported many hypotheses that markets were not fully efficient since they did not always incorporate all available information, again proving that there is room for traders and algorithms to find and exploit these patterns.

Another main factor that changed with the access of the internet was not only the rumors and buzz from the news but also the true information posted by companies. Copeland's (1976) theory of sequential arrival of information had now become mainly inapplicable due to instant access to information. By the early 2000's, all companies had moved from mail sent earnings reports to posting them on the internet. Due to this most investors would react to earnings news at the same time.

Baker and Wurgler (2007) studied the effect of investor sentiment around the time of quarterly earnings reports, finding that a company's stock price with reacted stronger positively in the case of a positive announcement. While when there was negative news, or they reported below investor expectation there was a lesser negative price movement. Baker and Wurgler (2007) concluded that strong investor sentiment and optimism can lead to overreaction to good news and an underreaction to bad news and vice versa when there is a negative sentiment; leading to mispricing in the short term around earnings reports.

With news and stock prices being readily available throughout the whole day via the internet for all investors, institutional and recreational, rather than closing prices being found in the newspaper each morning, investors began trading on a more short-term information basis (Barber and Odean, 2008). This began the trend that has taken over most trades in the modern day, which is known as day trading. It led to a change in trader behaviour due to their being an increase in market responsiveness.

Barber and Odean (2008) began researching behaviour and investor sentiment due to short term information, such as basing trades off of one day returns and high volumes. Companies being published in the news or with positive earnings reports were no longer the only thing that could grab a trader's attention on a daily basis when deciding their trading strategy. Barber and Odean (2008) found that investors are net buyers of companies that are defined as attention grabbing stocks. Barber and Odean (2008) defined attention grabbing stocks as those that are recently excessively mentioned in the news, those with extreme one day returns, or with abnormally high trading volume.

They attribute this behaviour to the fact that investors have thousands of stocks to choose from and the average investor does not possibly have the time to go through each one. They need some sort of criteria to select their stocks and method of sorting which stocks to stick out. This leads to investors going through the news and looking for something that catches their eye; with something that has produced a return of 20% in the past week looking much more attractive to some recreational investors than some stable stock that moved only 1%. Barber and Odean (2008) believe short term returns are based mainly on media attention and investor behaviour rather than true company performance.

Barber and Odean (2008) found that while stocks with high media attention tend to have higher short-term returns, that companies with little to no media attention tend to have higher long-term returns. These findings are very similar to those of Huberman and Regev's (2001), that found that companies with high media coverage tend to have high short-term returns that are corrected by the market after some time but companies with little media coverage outperform in the long run. It surprising to see that the same pattern still held seven years after the paper was published, since throughout history it has been seen that once an imperfection or patterned mispricing in the stock market occurs, it is quickly corrected by the market through algorithm traders and other forces in most cases. In order to see if there are still such opportunities in modern-day trading, some of the research on this paper will focus on the effect of news publications on abnormal trading volume.

H1: What effect does the number of total daily news publications have on a company's trading volume

A lot of these new trading strategies are based on short term market inefficiencies rather than long term trading strategies. The research being done in this paper also studies short term effects of investor attention due to news and media attention. The research from the past lacks analysis that links the attention from the news directly to the returns of these companies. The research is rather focused on its effect on volume, which is also of importance, but does not delve into direct effects on returns. This is due to a number of reasons, being that it is difficult to collect data that directly portrays investment sentiment in the sense of buy, sell or hold.

Due to lack of clarity in this area and greater availability of data now, the paper will also look into and analyze the effects of news sentiment on stock returns for specific companies.

H2: What effect does news sentiment have on a company's daily returns

As the internet became more established and widely used, more data and more data became available for researchers to track and know what people were looking up on search engines such as Google. In 2006, Google released a feature known as "Google Trends" which allowed users to track what the months and years most searched things were. Bank, Larch, and Peter (2011) decided to use this tool as a proxy for investor attention, using data of what the top searched public companies or tickers were and regressing these on volume and returns. They found that Google search volume was a strong and effective proxy for investor attention, having high correlations with trading volume.

With relation to returns, Bank, Larch, and Peter (2011) had very similar to results to Barber and Odean's (2008) and Huberman and Regev's (2001) research. Finding that higher search volumes led to short term returns due to short term price pressure from the attention driven trading but also that an increase in Google search volume led to a decrease in illiquidity. Since data availability became greater and more reliable researchers have been able to make more complex and accurate models in order to use these information's channels to predict market characteristics.

2.1.3 The Impact of Social Media on Trading Volume and Stock Returns: A New Proxy for Investor Attention

With the release of the internet came much more than just faster news and more available information to investors. It also created a breeding ground for platforms and messaging boards for people to write each other in real time. Before the internet, the only way of real time long-distance communication was wireless phone calls. The internet allowed millions of people to be able to connect and discuss whatever they like without buffer. One of these things was the creation of stock messaging boards where hundreds or thousands of every day investors would be able to discuss ideas and theories. It opened a channel for people to get opinions and information from a source other than the newspaper or online articles from more trusted sources.

Antweiler and Frank (2004) collected and analyzed data from over 1.5 million messages on different messaging boards across Yahoo! Finance and Raging Bull, examining these messages' ability to predict stock returns, trading volume, and market volatility. Their research was much different than all those off the past because they were analyzing the sentiment and opinions of the investors themselves rather than news reports and company releases. They used the Naive Bayes algorithm to characterize all of the messages into bullish, base, or bearish using linguistics from the messages. Antweiler and Frank (2004) found that on the day following high overall messaging activity about a company/ticker, there was a small negative return. This differs from past papers such as Bank, Larch, and Peter's (2011), Barber and Odean's (2008), and Huberman and Regev's (2001) papers which found positive returns following high investor attention**.** When looking at volume, Antweiler and Frank (2004) found that the more discrepancies in opinion there were, the higher the trading volume that same day, which confirmed one their hypotheses which states that disagreement induces trading. When looking at overall market volatility, it was found that the overall message board activity was positively correlated with market volatility, also while controlling for trading volume (Antweiler and Frank, 2004).

As the internet and social media evolved so did the number of online stock discussions. Platforms such as Twitter, which is a social media for all types of content, was used to discuss stocks but there were also dedicated platforms such as StockTwits, where there are hundreds of forums where there are different types of trading theories and strategies being discussed all day. Drake, Roulstone, and Thornock (2016) extended and updated past literature as there was rapid growth in internet and social media use in the early 21st century, finding that high levels of investor attention can be used to predict short- term returns. They also find that social media attention is a strong proxy for investor attention, as stocks with higher media attention experience higher trading volumes, backing up that attention-driven trading is a key trading dynamic and variable to consider in decision making.

H3: What effect does the number of total daily Twitter publications have on a company's trading volume

Such trading dynamics being found present in the market exposes the further in inefficiencies that attention-driven trading has introduced since the increase in use of the internet and social media. Another event that only boosted the importance of the internet and social media was Covid-19. Forcing everyone to be at home for the better part of the year and communicating almost only online and virtually, definitely fast forwarded the world a few years in terms of tech use.

Lazzini, Lazzini, Balluchi, and Mazza (2022) researched many different relationships between social media and the stock market during the first Covid-19 lockdown since many people were stuck at home, leading many to have greater opinions, emotions, and moods than ever. They found that emotionally charged tweets, especially those portraying a negative sentiment, had significant influences on stock returns as well as market volatility. Lazzini, Lazzini, Balluchi, and Mazza (2022) also found that market overreactions were more extreme than usual and suspected that this was due to heightened emotions during the time. Another result was that high social media activity correlates with abnormal market returns, confirming their suspicions that social media sentiment can predict stock market returns.

These analysis's from past research have led to great strides in the financial research community but many lack the implementation of the data they have collected. Part of the analysis done in this paper will look at ways to directly look at the effect of sentiment data on stock returns. With their being a plethora of literature that shows strong relation between financial markets and medias, there is confidence that the direct connection of trading strategies can be made in order to prove use of real-life strategies.

The trading environment during Covid-19 was unlike anyone had ever seen before and took an even bigger turn when investors took to Reddit to discuss investment ideas. Kim, Lee, and Kauffman (2023) researched the modern trading phenomena known as *GameStop Short Squeeze*, where they analyzed the social media activity and sentiment around the GameStop stock around the time the stock jumped from $15.00 to a high of over $500 in just a matter of weeks. This occasion was one of the first instances of extreme herd behaviour trading, where a bunch of investors come together and execute the same trade in inflate or deflate a stock price. Kim, Lee, and Kauffman (2023) found high correlation between activity on the Sub-Reddit known as "WallStreetBets" and trading volumes and upward price movements of the GameStop stock. This was no surprise due to the Sub-Reddit have a cult like following and the *GameStop Short Squeeze* played out similar to coordinated plan. It was taking advantage of market dynamics using knowledge of derivatives and Greeks in order to get the stock price to shoot upwards and make handsome profits at the cost of those short the stock incurring very heavy losses.

H4: What effect does Twitters sentiment have on a company's daily returns

As seen in the review, trading and market forces have come a long way in the last 40 years with no signs of slowing down with new tech constantly being introduced. The additional contribution of aims to provide investors and traders with insights on how non-financial market forces can affect their investments and allow them to make investment decisions with all available information and ability to understand the information and try to stay one step ahead of the forever evolving market.

3. Data

3.1 Data Extraction

The data used in this paper was extracted from The Bloomberg Terminal, which is the most well-known database in the financial industry. It is mainly used by financial institutions, as it is providing real time and historical data and is able to assist in making investment decisions, though coming at a very large subscription cost. The reliability and accuracy of the data from Bloomberg Terminal are widely accepted and is the most reputable financial asset and database in the industry and has been for some time.

The data for this paper is along the timeline of April 1st, 2018, to December 31st, 2023. The dataset includes the specific financials and metrics of the 25 biggest companies by market cap as of January 1st, 2018. It was decided to use the 25 largest companies due to smaller companies often not being discussed as often online and therefore leading to insufficient data points. The following twelve variables were extracted from the terminal to complete the calculations for the paper.

Table 1 Description of Extracted Data Variables

Variables	Data Source	Description
TwitterPos	*Bloomberg Terminal NLP AI*	Companies/tickers daily positive mentions on Twitter to measure user sentiment
TwitterNeg	*Bloomberg Terminal NLP AI*	Companies/tickers daily negative mentions on Twitter to measure user sentiment
TwitterPub	*Bloomberg Terminal NLP AI*	Companies/tickers daily mentions on Twitter to measure stock attention
NewsPos	*Bloomberg Terminal NLP AI*	Companies/tickers daily positive mentions on the news to measure user sentiment
NewsNeg	*Bloomberg Terminal NLP AI*	Companies/tickers daily negative mentions on the news to measure user sentiment
NewsPub	*Bloomberg Terminal NLP AI*	Companies/tickers daily mentions on the news to measure stock attention
Volume	*Bloomberg Terminal Financial Data*	Companies daily trading volume
Price	*Bloomberg Terminal Financial Data*	Companies daily market closing price

DailyHigh	*Bloomberg Terminal Financial Data*	Companies highest daily trading price
DailyLow	*Bloomberg Terminal Financial Data*	Companies lowest daily trading price
DailyMarketReturn	*Bloomberg Terminal Financial Data*	S&P 500s daily closing price
EPS	*Bloomberg Terminal Financial Data*	Companies quarterly reported Earnings per Share (EPS)
DebtEquity	*Bloomberg Terminal Financial Data*	Companies quarterly reported Debt-to-Equity Ratio

This table includes the variable names and descriptions for the data that was collected for this paper, including sources

Bloomberg was known mainly for providing a central information hub for all financial information that people in the industry could need. Throughout time they have begun including additional data and resources that they have provided themselves, such as additional analysis for the data they provide. Bloomberg has working with this AI for several years improving its ability to detect errors within the code that lead to false counts and errors, speaking for its accuracy and robustness. A main part of the data and research is based on the sentiment analysis, which is done through a new method of AI known as Natural Language Processing (NLP) which allows AI to go through text and read emotion and tone of texts.

Although the Bloomberg Terminal has a reputation of providing strong and accurate data, it must be acknowledged that AI is still new, and it is not perfect. AI often has trouble with dealing with speech such as sarcasm and satire, being unable to understand the true meaning of more complicated human speech and text. This should be considered when basing decisions using AI constructed datasets and when basing decisions based off of AI.

Data in the dataset can also be influenced by external factors such as industry specific news or earnings reports, which often need to be considered when looking at certain variables and are important to keep in the back of one's mind.

Overall, data provided by the Bloomberg terminal is robust and reliable, making it worth examining the relationship between investor sentiment reflected in social media and the news and its impact between stock performance.

3.2 Summary Statistics

Table 2 Descriptive Statistics of Bloomberg Variables

	Obs	Mean	Std. Dev	Min	Max
TwitterPos	34,500	19.87	54.19	0	1961
TwitterNeg	34,500	26.31	83.98	0	3998
TwitterNetEffect	34,500	-6.44	70.81	-3435	1557
TwitterPub	34,500	446.62	970.30	0	46434
NewsPos	34,500	4.87	16.64	0	466
NewsNeg	34,500	9.47	33.21	0	1535
NewsNetEffect	34,500	-4.60	33.50	-1434	420
NewsPub	34,500	469.94	701.04	1	15526
DailyMarketReturn	34,316	.00035	.0132	-.12	.09

This table contains the summary statistics of key variables that were extracted from Bloomberg

For the summary statistics, only variables that did not directly differ in size and relation due to magnitude were included. Looking at variables such as DailyMarketReturn, the results are as expected with the S&P 500 providing a daily slightly positive return. Though when comparing the total number of publications for Twitter and the news, it can be seen that they are almost identical, differing only by roughly 5%. This is an interesting when looking in contrast at the number of positive and negative remarks for Twitter and the news. It can be seen that for Twitter there are roughly 46 daily sentiment comments while for the news there are only about 14. A possible explanation for this is that for the sentiment counts, the AI only picks up on comments or remarks that are strongly opinionated. Due to the news being a platform for the sharing of general information to the public rather than a platform for journalists to share their true opinion, there may be a lack of news articles that have a strongly opinionated syntax. In the US, many news channels are known to either lean to the right or left politically; however, there is a limit on what they can say or what they can tell people to do or not do. On the other hand, social media and Twitter especially are platforms for people to speak their voice on any topic of their choosing and do not have many restrictions on how opinionated their comments or posts are. This is a logical explanation to why more than 3x as many of the Twitter publications are picked up as positive or negative sentiment comments or posts.

An additional interesting observation from the summary statistics is that for both Twitter and the news, there are significantly more negative sentiment posts in comparison to positive ones on average each day. For

Twitter it is around 33% more and for the news there are more than 2x as an many negative remarks on average daily. An explanation for this is that studies have found that people often tend to focus on the bad and that negative events tend to catch people's eyes (Marano, 2003). They find that positive events have to be bigger in order to make the news and a smaller negative event would trump the interest of the viewers. An additional explanation is that the news were originally used to notify people of what was going on in the world and warning them about certain events and dangers, making it bias to including events that people need to be aware of for safety reasons.

There is a similar explanation for the difference in the number of positive and negative remarks on social media. Studies have found that people again rather focus on the negative and make posts on social media about someone that is struggling or negative events going on around the world (Skogen et al., 2023) This could mean that people rather make posts about companies that are doing poorly or facing certain issues rather than a company that maybe just opened a new branch. These reasons are likely the main causes of this landslide of negative sentiment posts or positive ones on Twitter and in the news.

4. Methodology

Given the collected data and variables, there are still additional calculations that need to be done to attain additional variables before the regressions for the hypothesis can be built. The variables extracted from the database can be used to calculate additional variables such as daily returns. Bloomberg data provided the daily closing prices for the 25 largest companies on the S&P 500; however, to find the relationships between investor sentiment and returns, daily returns for each stock must first be calculated. Martinez-Blasco, Serrano, Prior, et al. (2023) use the following formula to calculate a firm's daily returns, which is a commonly accepted method in the financial research community,

Formula 1.1:

$$DailyReturn1 : R_{i,t} = ln \frac{Closing\ Price_t}{Closing\ Price_{t-1}}$$

The next key variable to calculate is the abnormal trading volume using the trading volume data that has been collected from the Bloomberg Terminal. It has been decided to use abnormal trading volume as it provides a number of benefits such as a normalized measure that allows for comparison across time and different sized companies. This is due to the fact that abnormal trading volume is a relative measure and is comparable across different companies in comparison to regular trading volume which is an absolute measure. It is also helpful for the detection of unusual market activities and specific event studies. In order to calculate there are a number of steps, first being calculating an average trading volume using a benchmark time period (Campbell, Lo, Mackinlay, 1997).

Formula 1.2:

$$Period\ Average\ Volume_{i,p} = \frac{1}{N_{i,p}} \sum_{t \in period} Volume_{i,t}$$

This average trading volume is calculated separately for each company. The average trading volumes have also been broken up into four separate periods, to account for events that would drastically change trading volume through the five years, such as a stock split, which could quickly double or triple trading volume due to the split alone. Now with the *Period Average Volume,* the abnormal trading volume can simply be calculated,

Formula 1.3:

$$AbnormalVolume_{i,t,p} = Volume_{i,t} - Period\ Average\ Volume_{i,p}$$

This equation provides the daily abnormal trading volume for each company on each day, relative to the current date and period it has been categorized in (Campbell, Lo, Mackinlay, 1997).

Additionally, to test the relationship between the net sentiment from Twitter users with the daily returns, the net sentiment must be calculated. The Bloomberg data only provides the positive and negative sentiments, so to solve for the net sentiment the following calculation is performed,

Formula 1.4:

$$TwitterNetSentiment_{i,t} = Twitter\ Positive_{i,t} - Twitter\ Negative_{i,t}$$

As the positive and negative remarks are also measured and studied in this paper for the news, a similar calculation is performed,

Formula 1.5:

$$NewsNetSentiment_{i,t} = News\ Positive_{i,t} - News\ Negative_{i,t}$$

An additional variable that must be calculated is the companies daily trading spread, which will be used as a control variable. From Bloomberg, the companies daily high and low trading prices were extracted. The following simple calculation is performed,

Formula 1.6:

$$Spread_{i,t} = Daily\ High_{i,t} - Daily\ Low_{i,t}$$

This calculates spread helps represent the amount of volatility a stock price witnessed during trading hours, since the greater the spread the greater price discrepancies there were during the day.

4.1 Effect of News Publications on Volume

After all needed variables have been calculated, it is now possible to test the effect of Twitter and news attention on a company's trading volume. Two separate regressions will be run in order to test the significance of both the variables separately. Then one will also be able to compare if one platform plays a larger or more significant role in a stock's attention and its trading volume. These regressions are based off of those used by Engelberg and Parsons (2009) when looking for the causal impact of local news coverage on local trading volume. They used control variables such as *Firm Attributes* to account for changes in company structure and *Earnings Surprises* to account for earnings shocks during company earnings report times.

Formula 2.1:

$$AbnormalVolume = \alpha + NewsPub * \beta_1 + DailyMarketReturn * \beta_2 + DailyReturnt1 * + Spread * \beta_4 + EPS * \beta_5 + \varepsilon$$

In the regression above there is the dependent variable *AbnormalVolume* and the main independent variable *NewsPub* who's true causal effect are trying to found. In order to find the best linear unbiased estimator four control variables have been added. This regression will allow us to interpret how one additional news publication will marginally increase trading volume. If there is a resulted coefficient of 500, it can be assumed that for each additional news publication the *AbnormalVolume* marginally increased by 500.

The first control variable is *DailyMarketReturn*, which aims to remove any outside effect th.at general market conditions have on a firm's volume. For example, when S&P 500 has high returns on specific days then daily trading volume increases in those days (Engelberg and Parsons, 2009). Including the *DailyMarketReturn* makes sure to isolate this effect into the variable itself and remove the bias from the *NewsPub* coefficient in the regression.

The second control variable is the corresponding stocks previous one day return, known as *DailyReturn1*. Martinez-Blasco, Serrano, Prior, et al., (2023) find that there is a strong correlation between a stock's price and its one day lagged share price. In certain cases, after a stock has an abnormal one day return it often gets more attention after traders realize the high returns. This then leads to the company gaining more short-term traction potentially leading to more short-term trading traffic with traders trying to get in on the fun.

The third control variable is added is the stocks daily trading spread. This variable helps control for general market price discrepancies that lead to increase in volume. Volatility goes hand in hand with trading volume and when traders see large ranges in stock prices, they often also see opportunity.

The final control variable that is included is *EPS*, which is used to control for sudden spikes in trading volume during earnings season. The variable is dormant throughout financial quarters due to it only being reported in quarterly reports but is useful to and key to controlling bias trading volume bias when earnings are being released. Engelberg and Parsons (2009) use this method through their variable known as *Earnings Suprise* where they use changes in earnings per share to help control for high volume during earnings reports.

To test the econometric strength of the regression, a number of tests can be run in order to check for certain data characteristics such as homoscedasticity, autocorrelation, multicollinearity, etc. A White test was run and there were signs of homoscedasticity found, which is not surprising due to the market being highly cyclical and have more dormant and then later more active periods of trading. An additional test was the Wooldridge test which showed signs of autocorrelation, which is not a shock. This was expected due to publications and sentiment numbers also being cyclical through their being more buzz around stocks at different points in time (Lee, 1985). Additionally, a Hausman test was run to help decide on whether to use fixed or random effects, to help deal with time-invariant characteristics of the companies due to the complexity of the panel data. The Hausman test resulted in a p-value of 0.00, allowing for the rejection of Hausman H0 and leading to the use of fixed effects for this regression.

4.2 Effect of Twitter Publications on Volume

In order to compare the effects of total news mentions and total Twitter mentions, the exact regression will be run using $TwitterPub$ as the main regressor instead of $NewsPub$. All other variables in the regression will be the exact same, including the control variables which are kept in for the same reasons as before.

Formula 2.2:

$$AbnormalVolume = \alpha + TwitterPub * \beta_1 + DailyMarketReturn * \beta_2 + DailyReturnt1 * \beta_3 + Spread * \beta_4 + EPS * \beta_5 + \varepsilon$$

These two regressions will provide results that will hopefully provide econometric clarity on whether or not traditional and social media have an impact on trading volume and bring attentions to certain stocks through discussion and mentions. The interpretation of the regression is very similar to Formula 2.1, where the coefficient will give us the marginal effect of a *TwitterPub* on a companies *AbnormalVolume*. The coefficients for these two regressions will be comparable, since the number of daily *NewsPub* and *TwitterPub* are almost the same in the summary statistics in Table 2.

The same tests as before were also run on the second regression, yielding similar results. The White and Wooldridge test found signs of auto correlation and homoscedasticity, being well explained by the nature of financial markets and the market participants. It also lightly supports the research of this paper, in that once there is movement in a certain sector, there tends to be movement in that direction, leading to auto correlation and homoscedasticity. The Hausman test was also run for this regression; however, finding a p-value of .3135, which indicates that a random effect regression is appropriate in order to deal with time-invariant variables.

4.3 Effect of Twitter Sentiment on Returns

The next set of regressions will analyze the impact that the beliefs and sentiments of this “attention” has on the direction that these stocks move in. The calculated variable *TwitterNetSentiment* represents net sentiment of the Twitter users, and this regression aims to find the effect it has on the corresponding stocks daily returns. Zhao et al., (2024) performed a similar regression analysis in their paper, looking at the effect of investor attention on a stocks daily return, controlling for variables such as Return-on-Equity (ROE) and market capitalization.

Formula 3.1:

$$DailyReturn = \alpha + TwitterNetSentiment * \beta_1 + DailyMarketReturn * \beta_2 + DebtEquity * \beta_4 + EPS * \beta_5 + \varepsilon$$

This regression uses similar control variables such as *DailyMarketReturn,* due to market returns having a large impact on companies returns, which is a relationship that has long been studied which is known as the concept of Beta. The coefficients of this regression will allow for interpretation of the marginal effect of each additional positive sentiment comment on Twitter. If significant, it will allow traders to multiple the *TwitterNetSentiment* by the resulting coefficient, which should give them a strong unbiased estimate of the days stock returns.

Two additional control variables used in this regression are *DebtEquity* and *EPS,* to help control for volatility in price changes during earnings reports season. Debt-to-Equity ratio which helps to account for large trades or attention due to a change in capital structure (Engelberg and Parsons, 2009). When a company reports its earnings, they sometimes decide to take on capital restructuring for any number of reasons, such as needing to raise debt due to issues or take on more equity due to strong performance. These could lead to an increase in *Volume* and the inclusion of the variable *DebtEquity* helps remove the bias out of the target variables *TwitterPub*. This variable is dormant throughout the majority of the quarter due to capital structure changes only being announced on a quarterly basis, however; it is used to control for the chaotic trading environment around a company's earnings report.

EPS is one the key indicators of a company's demise or strong performance when earnings are released each quarter. When there is a surge in a company's *EPS* there is usually an abnormally large trading volume and including the variable helps remove the bias of outside financial performance factors out the *NewsPub* variable (Engelberg and Parsons, 2009). Similar to *DebtEquity*, *EPS* is mainly used to control for earnings surprises as Engelberg and Parsons (2009) use in their methodology.

After running the White and Wooldridge test, autocorrelation and homoscedasticity were found present; however, expected as with the other regressions, due to the nature of dealing with financial market data and this being a common occurrence (Lee, 1985). A Hausman test was run, finding a p-value of .3165, suggesting use of a regression using random effects, to deal with the complexity of panel data and time-invariant variables.

4.4 Effect of News Sentiment on Returns

Many past papers have focused on news outlets and outdated media sites such as messaging boards but the new Bloomberg AI measures Twitter sentiment allowing to measure for sentiments from a modern-day mainstream media. There is little research in looking at the direct impact of social media sentiment on the stocks returns. Due to their being many factors that impact a company's return, this regression seeks to find the causal impact of news sentiment without bias from outside events,

Formula 3.2:

$$DailyReturn = \alpha + NewsNetSentiment * \beta_1 + DailyMarketReturn * \beta_2 + DebtEquity * \beta_4 + EPS * \beta_5 + \varepsilon$$

This regression has very similar characteristics to Formula 3.1 and will also be compared. The interpretations are very similar, with the resulting coefficient allowing traders to multiply *NewsNetSentiment* to get an estimated stock return for that given day.

Looking the same potential negative regression characteristics using the White and Wooldridge tests, autocorrelation and homoskedasticity were found. Similar to the other three regressions, this is a common when dealing with market data in econometrics. When the Hausman test is run, a p-value of .0807 indicating that random effects should be used, similar to the previous regression.

5. Results/Discussion

5.1 Regression #1: News Publications Effect on Abnormal Trading Volume

Table 3 Formula 2.1: Dependent Variable is AbnormalVolume

Variable	Coefficient	Robust Std. Error	Z-score	P-value	95% Conf. Int.
NewsPub	6237***	926	6.73	0.000	(4316,8159)
DailyMarketReturn	-4319988	4904655	-0.88	0.388	(-1.45e+07, 5851644)
DailyReturn1	-1.68e+07***	5594944	-2.99	0.007	(-2.84e+07, -5151823)
Spread	248	190	1.30	0.206	(-146.438, 642.433)
EPS	-151600	106895	-1.42	0.170	(-373285, 70086)
Constant	-2769909***	530398	-5.22	0.000	(-3869888, -1669930)

This regression has an overall R-Squared of 0.0164 and N=34314. The table includes the regression results from formula 1.1, which is a regression with *AbnormalVolume* as the dependent variables and Twitter*Pub* as the main independent variable. The rest of the variables in the regression are included as control variables. The significance stars in this table shows whether the variable is statistically significant at * p<0.10, ** p<0.05, *** p<0.01 significance level. The data represented is derived from the non-experimental sample.

Having built up an understanding of the topic itself and relevant past literature, now it can be seen how these relationships and effects have evolved throughout time. Looking at Table 3, it can be seen there is high significance for the variables of interest, *NewsPub*. *NewsPub* influence on trading volume is not only evident but also robust at a significance level of 1% level. Also given that the *NewsPub* coefficient is positive, the null hypothesis of that the number of news publications about a company have no effect on abnormal trading volume can be rejected at a 1% level. This strong robust relationship underscores and proves the idea that news attention has a significant impact on a company's abnormal trading volume. Additional measures such as the usage of control variables also provide additional confidence in this regression, due to the issue of endogeneity often being untestable.

Having controlled for company specific metrics such as EPS, accounting and removing bias for changes in financial performance. Additionally, the trailing one day is also included to control for the effect of investor attention due to specific returns, as past high one day returns lead to additional attention as high returns often catch people's eye (Jung and Kang, 2021). Daily market return was also included to control for general high market activity, during certain extraneous events such as natural disasters or election times, there is generally higher than average market activity and this helps control for these effects.

These results can be explained by a number of reasons, the main one being that many investors, specifically institutional investors use traditional media as one of their main sources of information or inspiration for their investments. Traditional media continues to play a significant role in the communication of information for all investors, even with many new platforms having been introduced in the last 20 years. This theory is also supported by Barber and Odean (2008), which found that investors need certain tools to narrow down their search of stocks from thousands to maybe a hundred companies one can realistically look at due to cognitive limitations. It is also human nature that if someone reads the paper and sees a company mentioned in multiple headlines or sub articles, that they now be aware and curious about the company. In the majority cases, a public company is usually in the news due to either negative or positive news, meaning that investors would like to either sell their shares due to poor news or buy shares due to positive news. Overall, these results and theories align with the body of growing literature on the topic of the effect traditional media on investor behavior.

5.2 Regression #2: Twitter Publications Effect on Abnormal Trading Volume

Table 4 Formula 2.2: Dependent Variable is AbnormalVolume

Variable	Coefficient	Robust Std. Error	Z-score	P-value	95% Conf. Int.
TwitterPub	3296***	980	3.36	0.001	(1375,5217)
DailyMarketReturn	-4320959	4397904	-0.98	0.326	(-1.29e+07, 4298775)
DailyReturn1	-1.58e+07***	5101704	-3.10	0.002	(-2.58e+07, -5834572)
Spread	-.962	105.043	-0.01	0.993	(-207,205)
EPS	-12467	159902	-0.08	0.938	(-325870, 300936)
Constant	-1447120**	732552	-1.98	0.048	(-2882895, -11345)

This regression has an overall R-Squared of .0423 and N= 34,314. This table includes the regression results from formula 1.1, which is a regression with *AbnormalVolume* as the dependent variables and *TwitterPub* as the main independent variable. The rest of the variables in the regression are included as control variables. The significance stars in this table shows whether the variable is statistically significant at * $p<0.10$, ** $p<0.05$, *** $p<0.01$ significance level. The data represented is derived from the non-experimental sample.

Looking at the results of the comparative regression, including *TwitterPub* instead of *NewsPub* to study the effect of social media on volume, it can be seen that results are similar in many areas. Observing the results in Table 4, it can be seen that the second regression yielded a similar result, with *TwitterPub* being positive and statistically significant at the 1% level. This strong significant effect shows that the importance and rise of social media is not one to ignore. These results align with the growing body of literature around the subject of social media and its impact on investor and media attention. Similar to *NewsPub*, it proves to be a predictive tool that can be used in the real world with viable and realistic implications to assist investment decisions for institutional and recreational investors. In order to ensure robustness, the same control variables were added as the previous regression, covering different firm specific and market factors that could lead to omitted variable bias and endogeneity.

These results were also expected, as social media has become such a large part of many people's lives and has become their main form of communication. Communities in all sorts of subjects and areas have grown to have had such a large impact in their area, such as video games and sports, that stocks and investing were not expected to be an exception to this tidal wave. An additional factor to consider is that investing amongst young adults has become more intriguing and easily accessible through brokerage trading apps such as Robinhood. With a growing youth population investing and the majority of social media users being young adults, it was not a stretch to theorize that social media would become a major variable in the investing world and financial markets. The combination of the increase in social media usage and stock trading has sprouted a unique and new environment where traditional market analytical tools can be supplemented by these digital trends. The younger group truly showed their strength in numbers and importance through the *GameStop Short Squeeze.* Short sellers of the GameStop stock paper losses totaled to $1.4 bn across the whole phenome and caused many hedge funds and institutional investors to have irrecoverable losses (Lipschultz, 2024). The financial impact of this was quite profound, especially after similar short squeezes followed on companies such as AMC Theatres and has led to reevaluation of risk management strategies among institutional investors.

Comparing the first two regressions, they both yielded very similar results and provided strong economic evidence in light of the papers hypotheses, that Twitter and news publications can be used to predict abnormal trading volumes and to better understand what areas investors and traders have their eyes on right now. For instance, the *GameStop Short Squeeze* started as chatter on a sub-reddit and turned into one of the largest trading phenoms of all time. Due to the measurement of the publications for the two platforms not being the same, it is not possible to compare the values of the coefficients themselves in order to see which may have a stronger effect.

5.3 Regression #3: News Sentiment Effect on Daily Stock Returns

Table 5 Formula 3.1: Dependent Variable is DailyReturn

Variable	Coefficient	Robust Std. Error	Z-score	P-value	95% Conf. Int.
NewsNetEffect	.00002***	4.67e-06	4.88	0.000	(.00001, .00003)
DailyMarketReturn	-.008	.006	-1.39	0.164	(-.019, .00)
DailyReturn1	.328***	.064	5.10	0.000	(.202, .454)
EPS	-.00005	.00003	-1.51	0.132	(-.0001, .00001)
Constant	.0007***	.0001	6.93	0.000	(.0005, .0009)

This regression has an overall R-Squared of .3290 and N= 34,314. This table includes the regression results from formula 1.1, which is a regression with *DailyReturn* as the dependent variables and *NewsNetEffect* as the main independent variable. The rest of the variables in the regression are included as control variables. The significance stars in this table shows whether the variable is statistically significant at * $p<0.10$, ** $p<0.05$, *** $p<0.01$ significance level. The data represented is derived from the non-experimental sample.

Now looking at the next regression, where it was tested if the stronger worded and opinionated news publications have an effect on stock prices. There was more skepticism when it came to this regression due to the technology and AI behind the data collection being more complicated and untested. As discussed in the data, it quite difficult for AI to interpret human emotion through text due to the complexity of human emotion with factors such as sarcasm.

Looking at Table 5, it can be seen that the explanatory variable of interest, *NetNewsEffect*, is positive and statistically significant at the 1% level. Given the strong significance and positiveness of the regression, an academic buildup of the effect of news sentiment can be brought to discussion. These results can be explained by a number of reasons, being that with traditional news being a main and often trusted source of information, that when positive reports and opinions are given that traders will act on them. As past literature hypothesized, traders often base decisions on what they see in their surrounding environment and this regression provides strong evidence of this. If confident in the work, a trader could include the evidence from this regression to explore new trading strategies that can be implemented in everyday trading.

There were a number of reasons why the chance of getting causal significant results were feared. A major one being that NLP AI is a work in progress that often has trouble understanding human writing and the true emotion behind a post or sentence. In comparison to the average number of news publications, there are a lot less positive or negative articles. Looking back at the summary statistics, there of 470 news publications but only 14 positive/negative tweets. With improved AI that may be able to pick up more sentiment tweets, the power and reliability of this regressions may increase.

5.4 Regression #4: Twitter Sentiment Effect on Daily Stock Returns

Table 6 Formula 3.2: Dependent Variable is DailyReturn

Variable	Coefficient	Robust Std. Error	Z-score	P-value	95% Conf. Int.
TwitterNetEffect	.00001***	2.16e-06	6.42	0.000	(9.61e-06, .00002)
DailyMarketReturn	-.008	.006	-1.41	0.159	(-.019, .003)
DailyReturn1	.327***	.064	5.09	0.000	(.201, .453)
EPS	-.00005*	.00003	-1.72	0.086	(-.0001, 7.65e-06)
Constant	.0007***	.0001	6.32	0.000	(.0005, .0009)

This regression has an overall R-Squared of .3300 and N= 34,314. This table includes the regression results from formula 1.1, which is a regression with *DailyReturn* as the dependent variables and *TwitterNetEffect* as the main independent variable. The rest of the variables in the regression are included as control variables. The significance stars in this table shows whether the variable is statistically significant at * $p<0.10$, ** $p<0.05$, *** $p<0.01$ significance level. The data represented is derived from the non-experimental sample.

Very similar results were found for the comparative regression looking at the effect of *TwitterNetEffect* on the company's daily stock returns. As seen in Table 6, there is a positive and strong statistical significance found, with *TwitterNetEffect*, being significant at a 1% level. Through the inclusion of control variables and statistical tests to account for certain properties, this regression is to have strong econometric support.

These results make a strong statement in the direction of the argument that social media is officially a part of investing and should no longer be ignored. As one of the main sources of many traders' information and an online platform for regular everyday traders to exchange information and opinions, it is a resource of substantial volume.

There were a number of worries about this regression when collecting data and building up the econometric validity, but it exceeded all expectations. One reason was that lack of observations for the sentiment count compared to the total number of publications due to Bloomberg's AI being in development stages and that in time there may be better data to perform better analyses.

6. Conclusion

To conclude, the findings of this paper have provided extensions on past papers researching the effect of information access and media sentiment on the stock market and found that it plays a big role in today's financial markets. From its humble beginnings of being a gateway and centralized terminal for people to become investors in companies, placing their bet that the company would do well financially, to becoming a complex beast of millions of people buying stocks for hundreds of different reasons. This paper goes to show that the market is forever changing and there will be new factors to account for every day. There will be new information available that can change your whole analysis from the past.

Through these many different information portals, there have been a number of different groups of traders that have formed that trade in completely different ways. You now have the traditional institutional investors, the traditional at home recreational traders, pension funds, day traders, and also meme traders. The whole interesting dynamic of this is that all of them are often trading the same stocks and these leads to great amounts of unpredictability for certain groups. Looking back at the *GameStop Short Squeeze*, the traditional institutional traders who were short GameStop could have never expected for their investment to go this belly up, as GameStop was a dying company in most people's eyes. They ended up losing all this money due to the sheer fact that there was an unknown group of meme traders that they did not of. For more traditional old-style traders it has created this looming shadow in the background that they have to consider when making decisions due to the size of this new community.

There were a number of worries about this regression when collecting data and building up the econometric validity, but it exceeded all expectations. One reason was that lack of observations for the sentiment count compared to the total number of publications due to Bloomberg's AI being in development stages and that in time there may be better data to perform better analyses.

6.1 Limitations

With this paper and its econometric analysis, it is important to account for certain limitations. Through using advanced regressions through fixed and random effects along with a multitude of control variables, there can be confidence that these results are meaningful; however, as with all statistics one can never be one hundred percent certain. Additional limitations include for which companies this analysis can be applied to. The data set contained data for the 25 biggest companies on the S&P 500 and this was due to these larger companies being expected to be talked more and have more data, but this means that the results may not be the same when looks at smaller companies. It is important to only apply assumptions to similar data sets with similar characteristics, as these results may be the complete opposite for penny stocks or mid cap companies. With respect to the control variables, one will never be able to fully control for all forces that effect a company's trading volume or stock returns. There is always a possibility of omitted variable bias and endogeneity, which can often be present and can lead to a Type 1 error, with a false rejection.

6.2 Extensions

This being said, potential extensions to this paper could be to look at these effects for smaller cap companies and comparing the difference in effects. Due to many smaller stocks often having larger movements price movements, this could prove to be a very interesting topic to expand on. Additional data could also be collected to look at different social media companies that are known to be used for stock talk, such as Reddit and also some Facebook groups. With new AI and NLP, it allows for vast data scrubbing across a multitude of websites and pages, and the collection of more quantity and also more accurate data about investor sentiment allowing for more extensive and accurate analyzes.

7. Work Cited

Andrei, D., & Hasler, M. (2014, September 9). *Investor attention and stock market volatility*. OUP Academic.

Antweiler, W., & Frank, M. Z. (2005, November 27). *Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards*. Wiley Library.

Bajzik, J. (2021, October 23). *Trading volume and stock returns: A meta-analysis*. Science Direct.

Baker, M., & Wurgler, J. (2007). *Investor Sentiment in the Stock Market*. NYU Stern.

Bank, M., Larch, M., & Peter, G. (2011, July 9). *Google search volume and its influence on liquidity and returns of German stocks - financial markets and portfolio management*. SpringerLink.

Barber, B. M., & Odean, T. (2008, December 24). *All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors*. OUP Academic.

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993, November 1). *Trading volume and serial correlation in stock returns**. OUP Academic.

Chan, W. S. (2003, July 3). *Stock price reaction to news and no-news: Drift and reversal after headlines*. Journal of Financial Economics.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2016). *The Comovement of Investor Attention*. ResearchGate. h

Fang, L., & Peress, J. (2009, September 28). *Media Coverage and the Cross-section of Stock Returns*. Wiley Online Library.

Farhoudinia, B., Ozturkcan, S., & Kasap, N. (2024, May 18). *Emotions unveiled: Detecting covid-19 fake news on social media*. Nature News.

Godfrey, M. D., Granger, C. W. J., & Morgenstern, O. (1964, February). *The Random-Walk Hypothesis of Stock Market Behavior.* Wiley Online Library.

Jain, P. C., & Joh, G.-O. (1988). *The Dependence between Hourly Prices and Trading Volume*. JSTOR.

Jung, W., & Kang, M. (2021). *The Short-Term Mean Reversion of Stock Price and the Change in Trading Volume*. Emerald.

Karpoff, J. M. (2009, April 6). *The relation between price changes and Trading Volume: A survey: Journal of Financial and Quantitative Analysis*. Cambridge Core.

Kim, K., Lee, S.-Y. T., & Kauffman, R. J. (2023, May 23). *Social Informedness and Investor Sentiment in the GameStop Short Squeeze*. SpringerLink.

Lazzini, A., Lazzini, S., Balluchi, F., & Mazza, M. (2022). *Emotions, moods and hyperreality: Social media and the stock market during the first phase of COVID-19 pandemic*. Emerald.

Lee, C.-W. J. (1985). *Stochastic Properties of Cross-Sectional Financial Data*. JSTOR.

Lipschultz, B. (2024). *GameStop Shorts Dealt $1.4 Billion Paper Losses in Meme Revival*. Bloomberg.com.

Marano, H. E. (2003). *Why We Love Bad News*. Psychology Today.

Martinez-Blasco, M., Serrano, V., Prior, F., & Cuadros, J. (2023, April 11). *Analysis of an Event Study Using the Fama–French Five-Factor Model*. SpringerOpen.

Raza, S., Baiqing, S., Kay-Khine, P., & Ali Kemal, M. (2023, August 4). *Uncovering the effect of news signals on Daily Stock Market Performance: An econometric analysis*. MDPI.

Safdar, A. (2023, November 23). *As Israel Pounds Gaza, BBC journalists accuse broadcaster of bias*. Al Jazeera.

Skogen, J., Andersen, A., Finseras, T., Ranganath, P., Brunborg, G., & Hjetland, G. (2023). *Commonly Reported Negative Experiences on Social Media are Associated with Poor Mental Health and Well-Being Among Adolescents*. Frontiers in Public Health.