Outsider Trading: Trading on Twitter Sentiment

Dissertation

By: Josh Stevens (STVJOS003)

Supervisor: Professor Paul Van Rensburg

Abstract

This study aims to establish if a relationship between the investor sentiment generated from social media posts, such as Tweets, and the return on securities exists. If a relationship exists, one would be able to obtain an informational advantage from public information and outperform the market on a risk-adjusted basis. This would give the "outsider" information processed the predictive power of insider information, hence the title of the paper. The study makes use of Bloomberg's social activity data, which through natural language processing, allows for investor sentiment to be obtained by analysing a combination of Twitter and StockTwits posts. This paper makes use of a three-prong approach, firstly examining if investor sentiment is a predictor of next-day returns. Next, an event study methodology is used to examine the optimal holding period, which can further be expanded to test market efficiency. Lastly, this paper considers the asymmetric risk aversion as outlined by Kahneman and Tversky (1979). Results show that there is little to no correlation between sentiment and next day returns. There is evidence for a multi-day holding period being optimal but statistically insignificant and there is no evidence found for asymmetric risk aversion.

Key words: Sentiment Analysis, Twitter, Event Study.

Contents

1. Introduction

This paper examines the relationship between Twitter sentiment and next day returns, exploring if trading on Twitter sentiment, which is obtained by making use of Bloomberg's Social Activity Data, can produce returns that significantly outperform the benchmark on a risk-adjusted basis. This was examined in three parts.

The first, can Twitter sentiment-driven trades with a 1-day holding period outperform the benchmark on a risk-adjusted basis. Three portfolios were constructed to test for this, with the same construction used to compile three additional portfolios that made use of news sentiment as their basis for construction. From these portfolios, it was seen that there were some significant results, but due to the small effect size, it can reasonably be expected that the introduction of transaction and short selling costs would result in these returns being either insignificant or negative. Part one concludes that Twitter sentiment is not a predictor of next-day returns and acts rather as a trailing indicator.

Part 2 of the results examine optimal holding periods by making use of both an event study and Cohen & Frazzini's (2008) underreaction coefficient (URC). The event study shows that much of the information pertaining to the sentiment scores is priced in before the sentiment is registered on Twitter. This further provides evidence that Twitter sentiment is a trailing indicator. However, the event study and the URC provide evidence for a 5-day holding period. The multi-day holding period outperforms the 1-day holding period and the benchmark but not at a significant level.

Lastly, the asymmetric risk aversion of investors as explained by Kahneman & Tversky's Prospect Theory (1979) is examined. Making use of the trades driven by positive and negative sentiment and comparing the returns. It was seen that the trades driven by negative sentiment earned greater returns, and had greater volatility, this outperformance was not significant and thus it is concluded that there was no difference in reactions to positive or negative Twitter sentiment.

This paper is then broken up into seven sections, section 1 provides context, section 2 conducts a review of literature, section 3 outlines the research question, section 4 shows an overview of data, section 5 presents the results, section 6 concludes and section 7 outlines limitations and extensions to future studies.

1.1. Context

1.1.1. What is Investor Sentiment and how is it Quantified?

It has long been observed that valuations and the market price of an asset differ, often by a substantial margin. This has left investors unsure about how to adequately account for this price discrepancy. This phenomenon has been called noise or sentiment (De long et al. 1990a).

Investor sentiment has been defined in a 1990 paper (Morck et al.) as the beliefs that investors hold about the valuation of an asset that cannot be explained by the facts at hand.

When the concept of sentiment was initially conceptualised, it was done with no theoretical basis and the role of sentiment was left to be implicit in the studies that observed it. This was likely due to the difficulty in distinguishing between the random walk of the market and market mispricing (Baker & Wurgler, 2007).

Sentiment has been thought to be largely unmeasurable as, in the past, there has been no way to measure what the feelings or sentiment of market participants are. In the past proxies for sentiment have been used. These include things such as surveying investors, looking at trading volumes and retail investor trades amongst other things which have been outlined by Baker & Wurgler (2007).

Social media has become a central part of the way in which humanity communicates. This has resulted in a massive amount of data being generated, and data that is now useful as a result of breakthroughs in machine learning. While users cannot necessarily be linked to their trades when considering the aggregated data, the assets being discussed can be identified and the sentiment of the market with regards to those assets can be obtained (Deng et al., 2018).

Sentiment can be categorized as positive, negative or neutral. Positive sentiment has been defined as an expectation that the market will improve in the coming period, this could be interpreted as a buy signal, a negative sentiment shows an expectation that the markets will decline in the coming period, this could be interpreted as a sell signal, and neutral sentiment is the expectation that markets will remain fairly constant, which could be seen as a signal to hold.

1.1.2. What is Sentiment Analysis and how is it Conducted?

Sentiment analysis is a language processing technique that is used to identify if a text expresses a positive, negative or neutral emotion towards a subject (Gupta, 2018). The increased prevalence of microblogging sites has allowed for people to rapidly share their thoughts on views on a subject. This has increased the amount of data available to be analysed. Analysis of text data has evolved from a basic lexicon approach, which used a list of pre-specified words and rules to conduct an analysis, to more complex machine learning techniques (Lee, 2021).

These processes can examine the text for content and context and are able to output the sentiment of the text and what the sentiment is referring to. The lexicon approach is more basic and is outperformed by the machine learning algorithms (Xing et al. 2020), thus this paper will make use of Bloomberg's social activity data that is collected using a machine learning technique known as a support vector machine[1] (SVM) model (Cui, Lam & Verma, 2016). This supervised machine learning algorithm is used to determine the sentiment of each Tweet from a long investor's perspective. Bloomberg then makes use of a proprietary method to aggregate all the Tweets for a specific company in order to output a sentiment score between -1 (extreme negative) and +1 (extreme positive) and this is updated every minute for all public companies that are mentioned in either a Tweet or a news article. Data is pulled from both Twitter and StockTwits to compute these sentiment figures (Bloomberg, 2022).

If the sentiment is able to be ascertained it may allow for them to have the upper hand and make trades before the market does which would, in turn, result in market outperformance. This assumes that the sentiment is an adequate measure of future returns or losses and that trades can be made before the market moves. The need for speed of the trades may suggest a need for an algorithmic-based trading system as humans may be too slow to fully take advantage of this profitable opportunity.

1.1.3. Natural Language Processing (NLP)

NLP is a machine learning process by which large amounts of text can be sorted and examined for content and tone. NLP is an intersection between computational linguistics and machine learning (IBM, 2020).

NLP is an important development for big data as much of the data generated by humans through things such as social media is unstructured and is unusable unless processed. NLP allows the processing of this data and outputs meaningful metrics which can be used to make more informed decisions. The speed at which data is generated in the modern world requires the automation of the process. This speed of data and effective automation allows for more timely data to be available to aid decision making (MonkeyLearn, 2021).

[1] See Appendix 1 for further explanation of SVM models.

1.1.4. Long/ Short Portfolios and Measures of Performance

Long/ short portfolios are an investment style that allows an investor to make active bets on both the up and downside and allows investors to not be limited by the holdings of the benchmark. Long/ Short portfolios also allow for investors to fund their long positions with their short positions, reducing the initial investment needed while not drastically increasing risk as the longs should balance out the shorts resulting in a market-neutral portfolio (Grinold & Kahn, 1999).

The simple existence of positive returns does not signify skill or value-added. Thus, a way must be found to measure the returns of our strategy and be able to attach some level of statistical significance to those returns so that we may rule out luck being the driver of returns and instead attribute the returns to a superior strategy.

1.1.4.1. Sharpe Ratio

The Sharpe Ratio (Sharpe, 1966) is defined as the excess return earned over the amount of risk taken to earn that return.

$$SR_p = \frac{E\left(R_p - R_f\right)}{\sigma_p}$$

There is an approximate significant t-stat at a 95% or greater level, if the difference in Sharpe ratio is greater than two times the square root of two divided by the number of time periods (Grinold & Kahn, 1999).

$$\left(\frac{\bar{r}_p}{\sigma_p} - \frac{\bar{r}_b}{\sigma_b}\right) > 2 * \sqrt{\frac{2}{N}}$$

Where $\bar{r}_p$ is the return of the portfolio less the risk-free rate, σ_p is the standard deviation of the returns of the portfolio, $\bar{r}_b$ is the return on the benchmark less the risk-free rate, and σ_b is the standard deviation of the returns of the benchmark and N is the number of periods examined.

A study done by Dybvig & Ross (1985) has found that a positive alpha value does not imply a positive Share ratio, but a positive Sharpe ratio will result in a positive alpha. Thus, a Sharpe ratio approach will be used, with Jensen's alpha (Jensen, 1968) used as an explanatory tool, as it will allow us to further expand our analysis by making use of the information ratio, as discussed below.

1.1.4.2. Alphas with significant t-stats

In this paper, alpha will be calculated as the amount by which the portfolio outperforms the benchmark, with the S&P 500 being used as the benchmark. This will be calculated as the return of the portfolio over and above the risk-free rate less the beta of the portfolio multiplied by the benchmark return.

$$\alpha_p = R_p - \beta_p R_b$$

Alpha is defined as the average of the realised residual returns when looking backwards. Alpha is a useful metric to use as it accounts for excess returns but also attaches a t-stat to these returns to identify the significance of the result achieved (Grinold & Kahn, 1999).

The t-stat for alpha is approximately equal to the alpha of the portfolio divided by the risk of the portfolio multiplied by the square root of the number of periods examined:

$$t_p = \left(\frac{\alpha_p}{\omega_p}\right) * \sqrt{T}$$

This t-stat will identify if alpha differs significantly from zero.

1.1.4.3. Information ratio

The Information ratio (IR) measures achievement ex-post and implies opportunity ex-ante. The IR of a portfolio (IR_p) is the ratio of annualised residual return (α_p) to annualised residual risk (ω_p). A high information ratio implies that a manager efficiently makes use of information.

$$IR_p = \frac{\alpha_p}{\omega_p}$$

Information ratios are presented as an annualised figure. Thus, when looking at shorter periods the numbers need to be scaled. The alphas will be additive while the residual risk values will be scaled by the square root of the difference in time period, as risks do not add.

When comparing an active portfolio to the benchmark we will always see that the benchmark has no residual risk or residual return.

$$\alpha_{Benchmark}, \omega_{Benchmark} = 0$$

Making use of the IR we can then go on to calculate the information coefficient (IC). The IC is the correlation of each forecast with actual outcomes and is generally used as a measure of skill for an active manager but in this case, it will be used as a proxy for the value of the information obtained from microblogs.

The IC can also be calculated as the correlation between the signal, in this case, the sentiment, and the next day returns.

$$IC = corr(Sentiment, Next\ Day\ Returns)$$

The IR can be calculated by dividing the average *IC,* $\mu(IC)$*,* by the standard deviation of *IC,* $\sigma(IC)$.

$$IR = \frac{\mu(IC)}{\sigma(IC)}$$

1.1.5. What is an Event Study and Why is it Being Used?

An event study is a method of data analysis, that allows for the examination of the market's reaction to an information shock. It allows for the examination of how prices and returns act leading up to an event, on the day of the event and in the days after the event. It can be a useful tool for analysing whether an event or set of events had a marked effect on the returns of a security or a portfolio. And this allows the inference of how similar securities or portfolios will react to a similar event in future (Hayes, 2020).

This paper will make use of an event study as it will allow for the easy identification of the optimal portfolio holding time. While part 1 deals with portfolios that are rebalanced daily, information may not be fully priced in by the end of day 1. Making use of an event study will allow the determination of this.

1.1.6. Why Twitter and US Equities?

Making use of Twitter as our platform to analyse has been done for a number of reasons. Firstly, the data is readily available from Bloomberg so there is a large element of ease of use. However, we see that websites gain more value when they have more users (Ullrich et al., 2008). Twitter is currently the most widely used microblogging site in the world, with over 300 million users and more than 200 million users using the site daily (Statista, 2021).

The decision to make use of US equities lies in the United States of America having the most Twitter users (Statista, 2021), the depth and breadth of the US market, the prolific nature of indices such as the S&P 500 to be globally traded and a large amount of data available.

1.2. Why is Sentiment Analysis Relevant?

The increased use of social media has availed an entirely new data stream to us. We are now able to see people's and communities' thoughts on a topic and analyse this data to easily establish their feelings on a topic. This along with the increased ease of access for retail traders, through platforms such as Robinhood, have left markets subject to short, fast increases in volatility that can be profitable for traders if they can take advantage of these situations.

The internet has allowed for the easy congregation and discussion of topics. This has evolved into groups such as r/wallstreetbets being able to manipulate the share prices of listed companies, pushing share prices to absurd levels. NLP has granted the ability to monitor these groups and establish which assets are being discussed and the sentiment attached to the asset. If an investor is able to know which asset price is about to spike before it does, this will allow them to earn large risk-adjusted returns, and the use of sentiment will also provide a time to exit the trade before the price normalises. If this is possible it will allow for large profits to be made.

Trading on sentiment may also become more pertinent in the future as stat-arbitrage investors search for more ways to outperform the market. This will both result in the sentiment of microblogs carrying more weight, as posited by Tumarkin & Whitelaw (2001), and the profitable opportunities becoming more difficult to find.

2. Literature Review

While sentiment analysis using NLP models is a relatively new advent as it relies heavily on big data, there is a wealth of research done on investor sentiment that predates the big data era. These studies have laid the groundwork for the approaches used and often correctly theorised the actual results that are found in more recent studies.

The research methodologies used in this paper are based on the work of Oh & Sheng (2011) and Cui, Lam & Verma (2016). It must be noted that both of these papers have certain shortcomings. Oh & Sheng found significant results but the study made use of a relatively small number of data points (72 221 microblog postings) over a short time period of three months, however, this may imply a large effect size. Cui, Lam & Verma (2016) found significant risk-adjusted returns can be earned by trading on microblog sentiment, but this is a professional report done for Bloomberg to show off the tradable opportunities that are available to people who sign up for their service. Thus, the results of both studies should not be taken at face value and should be examined further, as this paper will attempt to do. While there is a reliance on previous work done, the research methods are sound, the possibility of bias lies only within the results of the papers.

2.1. Investor Sentiment

The concept of sentiment has existed since Adam Smith's "Wealth of Nations" (1776), where Smith states that the prevailing market price could be above, equal to, or below the actual intrinsic value of a good. Keynes (1937) also theorised sentiment, by making reference to the "animal spirits" that sometimes drive the market away from their true economic realities. Both indicate the ability of investor sentiment to drive prices away from fundamental values.

Since then, sentiment has been mentioned throughout literature, but in the 1980's it began to develop as its own area of research with papers such as the one written by Summers (1986) laying the foundations for our understanding of sentiment today. Summers theorised that speculators were driving the price away from its fundamental value as there was often no rational reason for the prevailing market price. Summers co-authored another paper (De Long et al., 1990a) expanding on this idea further introducing the concept of noise and noise traders. These traders were theorised to add volatility to the market, which would drive arbitrageurs out as they would not take on the additional risk or could not afford to hold their position which would often cause a further deviation of the market price from the fundamental value of the asset. This increased volatility often resulted in the noise traders outperforming those investing using fundamental valuations. Which goes against what the efficient market hypothesis would lead one to believe (Baker & Wurgler, 2006).

2.2. Sentiment Analysis

While the focus of this paper is making use of microblog sentiment to predict returns, it is pertinent to see the effect that sentiment can have on consumer spending in other areas to establish a causal link between sentiment and changes in demand.

Aggarwal et al (2012) have found that humans will express their emotions about a subject by making use of social media. This now allows companies and other individuals to see the public's emotions that are being expressed towards companies and products.

There have been many instances of studies exploiting this large amount of information to predict social and political activities. Studies such as Tumasjan et al. (2011) find a correlation between political party mentions and the outcome of the German federal election results. Tumasjan et al. (2011) found that this information was valuable as Twitter is used extensively for political debate and thus more debate on a subject may make the sentiment attached to that subject more valuable. Rui, Liu & Whinston (2013) found that positive Twitter sentiment is associated with improved movie sales. This study also found that more significant Twitter users, those with more followers, had a greater impact than smaller Twitter accounts.

Now that research showing that Twitter sentiment holds some predictive ability, this paper examines literature to find evidence that this effect exists in the literature for stock returns.

2.3. Microblogs Predicting Stock Returns

Many studies have been conducted to establish if there is any predictive ability within online platforms. Tumarkin & Whitelaw (2001) stated that microblogs would be useful for predicting returns if they contained new information. Their study posited four ways in which microblogs may contain new information. First, it was theorised that the predictive ability of microblogs was due to insiders attempting to influence other market participants or those with private knowledge inadvertently sharing this when communicating their thoughts on the asset. Next, it was suggested that the microblogs themselves contain no new information but rather provide a good proxy for general market sentiment. The paper then considered the possibilities that a large number of investors follow buy or sell recommendations that they see regardless of the fundamentals of the company. Finally, Tumarkin & Whitelaw (2001) suggested that momentum traders may be attempting to front-run the flows of those making use of microblog sentiment to make trades, which would further fluctuate the price swings that microblogs can cause.

Baker & Wurgler (2006, 2007) have also greatly added to the literature on this topic through two papers. They have noted that the effects of sentiment are far more pronounced on stocks that are newer, smaller, more volatile, unprofitable, non-dividend paying or with extreme growth potential. Thus, it is seen that firms that are relatively difficult to value, and arbitrage, are more likely to be affected by sentiment. Thus, while the focus of this study lies in the S&P 500, it would be anticipated

that the methods used would provide more significant results when applied to an index such as the Russel 2000 which has a larger inclusion of smaller cap stocks.

Baker & Wurgler (2006) find that when sentiment is high, speculative and hard to arbitrage stocks have lower future returns than bond-like stocks. This is supported by Antweiler & Frank (2004b) who found that there is a negative relationship between stock returns and investor sentiment. However, they find that while this negative return is statistically significant, it is not economically significant, and the returns are often overshadowed by transactions costs.

Many studies find a correlation between the number of mentions that a stock receives, and the volume traded (Antweiler & Frank, 2004a) which in turn results in greater volatility in the market but this can be difficult to take advantage of. However, an investor may be able to develop a strategy making use of options and a long Vega strategy to take advantage of the additional volatility in the market or specific assets. Antweiler & Frank (2002) observed that more messages are generally associated with a greater difference of opinion. This is seen in the data obtained, as often assets that have a large number of mentions on a given day will have sentiment scores that are close to zero, but high numbers of positive and negative Tweets.

Oh & Sheng (2011) find that the inclusion of microblog sentiment can help better predict returns, while Cui, Lam & Verma (2016) found that trading on microblog sentiment alone can result in risk-adjusted returns that outperform the market at both a daily and intra-daily frequency. Deng et al. (2018) found that trades based on sentiment resulted in returns that were statistically significant at an intra-daily frequency, but this statistical significance was not observed for trades conducted at a daily frequency.

Deng et al. (2018) highlight the phenomenon that occurs in the markets where there is a causality loop. It is sometimes observed that microblog sentiment affects stock returns, while in other cases stock returns can affect microblog sentiment positing that the relationship may be bidirectional.

2.4. Stock Returns predicting Sentiment

De long et al. (1990b) state that investors are influenced by the returns of the market, this positions the argument that microblogs are in fact as a result of stock returns and thus may have no predictive ability themselves but rather serve as a trailing indicator. Thus, following this line of thinking it would be assumed that a positive return in share returns would result in positive sentiment about that stock.

This effect may be further exaggerated due to the wide range from which sentiment is drawn by making use of both Twitter and an investment-specific microblog such as StockTwits[2].

Roseman & Smith (2001) proposed the Appraisal Theory. This theory states that the emotions that people display are a result of their evaluations of events and circumstances. Under this theory, it would be appropriate to state that a stock that had generated significant positive returns would then elicit positive emotions in those who are talking about that stock. However, Appraisal Theory goes on to add that specific emotional responses to a specific event cannot be mapped. This is due to the nature of humans to respond to the same events with different emotions and for different events to cause the same emotions. This makes investor sentiment a particularly difficult thing to assess, as a stock generating positive returns will cause someone who is long on that stock to be positive about the performance while someone who has taken a short position that stock to be negative about the performance. This was further backed up by Mayew and Venkatachalam (2012) in that individuals will express the change in emotions that are experienced as a result of external stimuli.

Lo et al. (2005) found that market participants have strong emotional responses to the returns of the market. Deng & Poole (2010) stated that the reaction of individuals is the result of the cognitive appraisal of events. Further Deng & Poole (2010) linked this external stimulus to the actions of investment professionals who are likely affected by stock returns. Otoo (1999) found a relationship between increases in stock returns and improvements in investor sentiment. Thus, it can be reasonably expected that social media sentiment may front run stock returns.

2.5. Mean Reversion

Keynes (1937) stated that if the price deviates from its fundamental value by a significant amount with no economic reasoning, then it can reasonably be expected that the price will return to its fundamental value. The extreme sentiment scores, being either the driver of stock values away from their fundamental value or an indication that values are significantly different to fundamental values may create a tradable opportunity to long or short, depending on the initial movement, during the price correction. Benjamin Graham (Graham, Dodd & Cottle, 1934) adds to this through his quote "In the short run, the market is a voting machine but in the long run, it is a weighing machine." Where in the short run prices may drift away from fundamental values but in the long run will revert to the actual value of the company.

[2] StockTwits is a social media platform that has been specifically designed to share ideas amongst investors and traders (https://stocktwits.com/)

2.6. Bearish Sentiment having Greater Predictive Ability than Bullish Sentiment

Bearish sentiment having greater predictive ability than bullish sentiment is rooted in Kahneman & Tversky's (1979) paper which presents Prospect Theory. Prospect Theory states that people are risk-loving on the upside but are risk-averse on the downside. This in turn results in an asymmetric risk function[3], and people take more drastic action in response to a negative shock than they would if they received a positive shock.

This can be extended considering the risk-loving nature of individuals on the upside. People may overreact to a positive information shock which will result in large positive sentiment that may not translate into returns due to individuals being overly optimistic.

Thus, it would be expected that negative sentiment, and the prospect of losses, to cause a larger reaction in investors than positive sentiment would cause. This is, however, opposed by De long et al. (1990b) as they believe that many noise traders, or those who may be particularly sensitive to sentiment, will sell when they see a negative shock but also buy when they see a positive shock as they attempt to chase trends.

2.7. Under-reaction Hypothesis

Cohen & Frazzini (2008) conducted a study to identify if investors' attention was limited, meaning that they did not adequately account for information that was obviously available. Their paper stated that two conditions need to be met to test for investor limited attention, thus under reaction.

First, any information that is overlooked by investors must be available to the investing public before the price evolves. In the case of microblogs, these are public platforms, so the first condition is met. Secondly, the information needs to be salient and an investor must be reasonably expected to gather this information. This second assumption is more questionable in the case of sentiment. The first hypothesis (as outlined below in 3.2) will examine if there is salient information in microblogs. But while sentiment can be obtained by anyone with the right tools, it is not a straightforward process and sources that contain an investor sentiment measure are often stuck behind paywalls. For the purposes of this study, this paper will move forward saying both assumptions are met but opposition to this is valid.

The under-reaction hypothesis was theorised by Cohen & Frazzini (2008) as a proxy for the under-reaction of the market. The under-reaction coefficient (URC) is defined as the fraction of total return from period t to period t + n that occurs in period t.

---

[3] See Appendix 2 for figure.

$$URC = \frac{Return_t}{Return_t + Return_n}$$

Under an efficient market, it would be expected to have a URC that is equal to or close to 1. Any statistically significant deviation below 1 is evidence of an under-reaction, while a deviation above 1 is evidence of an overreaction.

If the URC is not equal to one this is evidence against the EMH. However, a URC that is less than or greater than 1 may not be a signal that the EMH does not hold, as this may be due to the second requirement, that investors should reasonably be expected to gather the information, not being met. The EMH makes provision for information that is not obviously available, stating that only information that is obviously available will be priced in (Fama, 1970). Keane (1983) added to this by stating that "prices adjust rapidly and unbiasedly to new relevant price-sensitive information". If this is the case it provides a rare arbitrage opportunity within the bounds of the EMH.

Oh & Sheng (2011) find that microblog sentiment takes a few days to be fully incorporated in the price and found significant market under-reaction to microblog sentiment, which would offer profitable trading opportunities. This also means that an intraday or one-day holding period as used by Cui, Lam & Verma (2016) would not capture all possible sentiment returns.

2.8. Event Study

The publishing of "Characteristics and Procedure of Common Stock Splits" by Dolley in 1933 is said to be the first event study conducted and this has set the groundwork for the event study space in finance. This was further expanded on by Fama et al. (1969) where a comprehensive outline of how to conduct an event study was laid out.

Fama et al. (1969) made use of a market model, whereby they established a linear relationship between a share and the market and modelled the return of each security as a function of the market return. That paper also produced useful figures to visually model the reaction of the market to the specified events.

MacKinlay (1997) outlined methodology to have multiple events and to effectively aggregate data across them. This paper also outlined how to conduct significance testing to allow the identification of statistically significant results through a t-stat and enable the calculation of a P-value to access the level of significance.

2.9. Prior Study methodology

All the studies that did regression-based/ looked for lagged autocorrelations show little evidence, but the goal is to see if sentiment on microblogs can be used to outperform the market on a risk-adjusted basis thus a portfolio construction approach, making use of information ratios, Sharpe Ratios and t-stats will be used as outlined below.

2.9.1. Trading on Twitter Sentiment

A study conducted by Bloomberg (Cui, Lam & Verma, 2016) outlines a methodology for calculating the effect of including sentiment into a pricing model. The study makes use of a long/ short portfolio construction which allows active bets to be made on both the up and downside. The use of long/short models is also supported by Cohen & Frazzini (2008). They used three portfolio constructions for calculating the returns of trades based on daily sentiment data.

1. High-minus-Low (33%) (HML 33%)

In this approach, the top third of stocks when ranked by sentiment is held long while the bottom third of stocks when ranked by sentiment is held short. Stocks in the long and short portfolios are equally weighted.

2. High-minus-Low (5%) (HML 5%)

The approach is the same as above expect only the top and bottom 5% would be held long and short.

3. Proportional Portfolio

Using this approach, all stocks with a sentiment score above the mean are held long and all stocks with a sentiment score below the mean are held short. The distance from the mean determines the level of over/ underweighting of a stock.

2.9.2. Event studies and The Under-reaction hypothesis

While Fama et al. (1969) outlined an approach to conducting an event study, MacKinlay (1997) provides a more comprehensive approach to event studies and allows for multiple events to occur. Thus, this paper will follow MacKinlay's approach but will draw on some of the formulas used by Fama et al. (1969).

Oh & Sheng (2011) made use of the URC which has been explained above and found the clear presence of an under-reaction in the market even after adjusting the market returns. The study found a clear example of a hump-shaped curve as described by Hong & Stein (1999). Oh & Sheng attribute this under-reaction as one of the reasons that microblog sentiment has the predictive ability, as also found in their paper.

Cohen & Frazzini (2008) looked at how price shocks in companies affected the share price of their suppliers. A simple formula was used, where the change in the price of the supplier would be equal to the change in the price of the customer company multiplied by the percentage of sales that the company made up for their supplier. This allowed for a defined change in price that the supplier was expected to experience. And allowed for a more accurate measurement of the URC than in Oh & Sheng (2011).

2.9.3. Bearish sentiment having more predictive power than bullish sentiment

Oh & Sheng (2011) and Deng et al. (2018) have laid out methodologies for examining possible differing effects of bullish and bearish sentiment.

Oh & Sheng (2011) make use of a J48 classifier, "Being a decision tree classifier J48 uses a predictive machine-learning model which calculates the resultant value of a new sample based on various attribute values of the available data" (Alam & Pachauri, 2017) to conduct a weighted F-measure but found a statistically significant difference between bullish and bearish sentiment when using a class-1 F measure. Oh & Sheng found that the class-1 F measure was appropriate due to its focus on correct predictions as bull labels have a high class-0 F measure due to predicting wrong predictions correctly. A class-1 F measure is a show of goodness of fit of a model, showing the accuracy that a machine learning model has, by indicating the number of true positives shown. Whereas a class-0 F measure shows the inaccuracy of the machine learning model by showing the number of false positives obtained (Zuccarelli, 2020).

Deng et al. (2018) make use of a vector autoregression (VAR), in a method similar to that of Lütkepohl (2007). The VAR model has been used in other studies to examine the potential bidirectional returns between variables (Tetlock, 2007). The VAR model also has the added advantage of being able to address the shortcomings of alternative econometric models as the VAR model can account for biases such as endogeneity, autocorrelations, and causality loops (Luo, Zhang, Duan, 2013).

3. Problem Statement

3.1. Research Question

The primary research question: Can trading on Microblog sentiment result in significant outperformance of the S&P 500? This will be explored through three sub-questions:

1. Do Twitter sentiment-driven trades with a 1-day holding period outperform a benchmark on a risk-adjusted basis?
2. Does a multi-day holding period outperform a 1-day holding period?
3. Does bearish sentiment hold greater predictive ability than bullish sentiment?

3.2. Hypothesis

This paper will adopt a similar approach as Oh & Sheng (2011) and will look at the 3-hypothesis approach to observing the predictive ability of microblogs.

1. A portfolio with a sentiment-based style will outperform the market on a risk-adjusted basis. If this is true, the difference between the Sharpe Ratio of the portfolio (SR_p) and the Sharpe Ratio of the benchmark (SR_b) will be statistically different with the Sharpe Ratio of the portfolio being greater. From Dybvig & Ross (1985) it is seen that if there is a positive Sharpe Ratio, then there is a positive alpha.

H_0: $SR_p = SR_b$

H_1: $SR_p > SR_b$, Alpha is positive

2. The under-reaction hypothesis. It takes a few days for the information to be fully incorporated into the price. If this is true it would be expected that the Sharpe Ratio of a multi-day ($SR_{Multi-day}$) holding period would be greater than the Sharpe Ratio of a portfolio that has a one day (SR_{Daily}) holding period.

H_0: $SR_{Multi-day} = SR_{Daily}$

H_1: $SR_{Multi-day} > SR_{Daily}$

3. The predictive power of bearish sentiment is higher than the predictive ability of bullish sentiment. Therefore, it would be expected that the Sharpe Ratio from bearish sentiment driven trades will be greater than the Sharpe Ratio from trading on bullish sentiment.

H_0: $SR_{Bear} = SR_{Bull}$

H_1: $SR_{Bear} > SR_{Bull}$

4. Data

Data has been obtained from Bloomberg Terminal. Bloomberg makes available the sentiment of a stock, from both Twitter and traditional news outlets, the number of times a stock has been mentioned by a Tweet or news story.

This study will consider the past two years of the S&P 500 at a daily frequency, beginning 02/01/2020 and ending 31/12/2021. While this is not a long timeframe and it is a period of significant market turmoil due to Covid-19, a large amount of data available and the existence of a market crash in March of 2020 and the outstanding returns as the world recovers from the global pandemic provide a broad backdrop for this study. It will allow us to examine how a market reacts to sentiment during both a crash and a boom period, this paper will also have the opportunity to examine if sentiment predicted the large crash or large gain.

The previous two years also provide us with almost 18 million Tweets and more than 23 million news articles and thus allowing increased confidence in our results obtained. The study is able to be scaled to any portfolio given the ability to acquire a sentiment measure.

When obtaining sentiment Bloomberg makes use of an SVM model, which allows sentiment to be directly observable (Cui, Lam & Verma, 2016). While there are more complex models such as long short-term memory (LSTM) or Bidirectional Encoder Representations from Transformers (BERT), when conducting sentiment analysis on financial data SVM models perform approximately as well as the more complex aforementioned models (Xing et al., 2020).

5. Results

The results section of this paper is broken into three parts. Part 1 examines the relationship between sentiment, obtained from both Twitter and news articles, and next day returns, making use of Pearson correlation, Sharpe Ratios, and alphas to explore this relationship. Part 2 looks to find an optimal holding period for the portfolio in order to maximise the returns generated. This is done by making use of an event study methodology and then is further unpacked using the URC from Cohen & Frazzini (2008). Part 3 will attempt to find an asymmetric relationship between positive and negative sentiment, as outlined by Kahneman & Tversky (1979). This is done by comparing Sharpe Ratios from trades driven by positive sentiment and trades driven by negative sentiment and examining for significance.

5.1. Trading on Twitter Sentiment

5.1.1. Correlations

Before conducting any portfolio construction to establish the relationship between sentiment and returns, with both Twitter and news sentiment used, the information ratio (IR) of the signal can be obtained by dividing the mean information coefficient (μIC) by the standard deviation of the IC (σIC). The IC can be observed directly as the Pearson correlation between the previous day's sentiment and the current day's return (Grinold & Kahn, 1999).

$$IC = Corr(Prev\ day\ sentiment, Return)$$

The IR of our data is:

$$IR = \frac{\mu(IC)}{\sigma(IC)}$$

There is some dispute in the literature whether returns follow investor sentiment or if sentiment follows investor returns. Thus, a comparison between previous, current, and next day's sentiment and current day's return has been considered.

The significance of the correlations can be calculated using the following formula:

$$t_{\rho(n-2)} = \frac{\rho}{\widehat{\sigma_\rho}}$$

Where ρ is the correlation which significance is being ascertained, n is the number of time periods and $\widehat{\sigma_\rho}$ is equal to:

$$\widehat{\sigma_\rho} = \sqrt{\frac{1-\rho^2}{n-2}}$$

Table 1a: Correlation between Twitter sentiment and 1-day stock returns. This is the average correlation using daily data over the period 2020-01-02 to 2021-12-31.

Twitter	$\boldsymbol{\mu(IC)}$	$\boldsymbol{\sigma(IC)}$	**T-stat**	**P-value**	**IR**
Previous day Sentiment and Returns	-0.002	0.060	-0.045	0.964	-0.034
Current day Sentiment and Returns	-0.004	0.067	-0.090	0.928	-0.061
Next day Sentiment and Returns	0.070	0.067	1.628	0.104	1.047

Table 1b: Correlation between News sentiment and 1-day stock returns. This is the average correlation using daily data over the period 2020-01-02 to 2021-12-31.

News Articles	$\boldsymbol{\mu(IC)}$	$\boldsymbol{\sigma(IC)}$	**T-stat**	**p-value**	**IR**
Previous day Sentiment and Returns	-0.001	0.073	-0.022	0.982	-0.013
Current day Sentiment and Returns	0.001	0.072	0.022	0.982	0.020
Next day Sentiment and Returns	0.082	0.119	1.845	0.066	0.687

Considering both table 1a and table 1b, there seems to be almost no support for investor sentiment, obtained from news articles or Tweets, predicting next day returns. There, is evidence that the relationship between previous day investor sentiment and returns is negative. The information obtained so far would support the arguments raised by De long et al. (1990b), Deng et al. (2018) and Roseman & Smith's (2001) Appraisal Theory that investor sentiment is driven by returns, over a short time horizon of one day thus leaving no profitable trading opportunity.

Observing the significance of the correlations it is seen that statistical significance exists at a 10% level only for returns and next day news sentiment.

However, this correlation test alone is not sufficient to negate the tradable opportunity that investor sentiment may present. The next section will attempt to construct a range of portfolios to capture the possible advantage presented by investor sentiment.

5.1.2. Sharpe Ratios

This paper has examined 3 methods and applied them to both the sentiment obtained from Tweets, covered in 5.1.2.1 and the sentiment obtained from News articles, covered in 5.1.2.2.

When considering a benchmark to use the SPX, the market-cap weighted index was the obvious choice. However, the SPW was also considered. The SPW is an equal-weighted index that is made up of all the constituents of the SPX where each security is held with an equal weight or with a 0.2% holding in each stock. This is rebalanced quarterly (Bloomberg, 2022). The SPW is also considered as it may outperforms the SPX when companies with a lower market capitalisation outperform. Considering both possible benchmarks it is seen that the SPX outperformed the SPW on both a return and a risk-adjusted basis, as we see the following results over the two-year period observed.

Table 2: A comparison of the returns and Sharpe Ratios of possible benchmark portfolios. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Index	SPX	SPW
Total Return	46.29%	40.46%
Annualised Return	23.15%	20.23%
Annualised Return in excess of risk-free rate	22.89%	19.89%
Volatility	26.06%	28.47%
Sharpe Ratio	0.88	0.70

The total return is equal to the value at the end of the period minus 1, as the starting value of the portfolio is 1, or 100%. The annualised return is the total return divided by 2, as the period examined is two years long. The Annualised Return in excess of the risk-free rate is the annualised return less the risk-free rate, where the risk-free rate used in the Sharpe Ratio calculations done throughout this paper is equal to the 1-year T-bill rate of 0.34% (Bloomberg, 2022). The volatility is directly observed as the standard deviation of the daily returns. This figure is then scaled to an annual figure by multiplying it by $\sqrt{253}$ as there are 505 trading days in the period.

It should be noted that this is above the long-term average volatility of the SPX, which normally sits around 15% [4](Bloomberg, 2022). This is due to the turmoil that the Covid-19 pandemic caused on global equities and will have a negative effect on our Sharpe Ratio. This negative effect should, however, be countered by the superior returns earned by the market over the past two years. These

[4] With a long-term average return of 11.88%, we see that the SPX has a Sharpe Ratio of about 0.8, thus making the observed periods Sharpe Ratio slightly above the long-term average.

abnormal figures should not affect our study as the figures used exist within the same period and thus will experience the same increased volatility and returns.

The returns of each portfolio will be calculated using a geometric return formula, as this will provide for correct compounding, the daily return on each stock is calculated as:

$$R_{L,it} = \frac{P_{Close,it}}{P_{Open,it}}, R_{S,it} = \frac{P_{Open,it}}{P_{Close,it}}$$

Where $R_{L,it}$ is the return on stock i that has been held long for day t, $R_{S,it}$ is the return on stock i that has been held short for day t, $P_{Open,it}$ is the opening price for stock i on a given day t and $P_{Close,it}$ is the closing price of stock i for a given day t.

And the daily return of the portfolio is calculated using the following formula (Dyer et al, 2014):

$$R_p = \left[\left(\prod_{t=1}^{T} R_{L,it} \right)^{\frac{1}{n_{Long}}} + \left(\prod_{t=1}^{T} R_{S,it} \right)^{\frac{1}{n_{Short}}} \right] * 0.5$$

Where R_p is the daily return of the portfolio, $R_{L,it}$ is the return of each stock i that has been held long on each day t, $R_{S,it}$ is the return of each stock i that has been held short for each day t, $\frac{1}{n_{Long}}$ accounts for an equal weight applied to each of the stocks held long and $\frac{1}{n_{Short}}$ accounts for an equal weight in each of the stocks in which a short position has been taken. The sum has been divided by 2 as both the long and the short makes up 50% of the total portfolio. The total value of the portfolio (V_p) is then calculated as the product of all daily returns:

$$V_p = \prod_{t=1}^{T} R_p$$

All portfolios are rebalanced daily and thus are rebalanced 505 times over the two years.

A t-stat will be generated for each portfolio by making use of the following formula:

$$t_p = \left(\frac{R_p}{\sigma_P} \right) * \sqrt{T}$$

Where R_p is the return of the portfolio, σ_P is the standard deviation of the portfolio and T represents the time period over which the returns are examined. In this case, the t-stat will be considered at a daily level. This t-stat will then be used to obtain a P-value to show the significance of the returns.

5.1.2.1. Twitter Sentiment

a) HML-33%

The HML-33% portfolio was constructed by ranking all securities by their Twitter sentiment score each day and then taking an equal-weighted long position in the top 33%, and an equal-weighted short position in the bottom 33%.

The initial results ignore both transaction and short selling costs and resulted in positive returns, and a greater risk-adjusted return than the benchmark, but a smaller absolute return.

Table 3a: Exploring returns of HML-33% Twitter portfolio with a 1-day holding period. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Component	Annualised Return	Annualised Volatility	Sharpe Ratio	T-stat	P-value
Long	0.73%	16.90%	0.02	0.061	0.951
Short	8.90%	18.42%	0.46	0.683	0.495
Overall HML-33% Portfolio	6.33%	5.99%	1.60	1.493	0.136

Table 3b: Average number of stocks in which a position is taken in HML-33% Twitter portfolio. Examining the daily average holding for the period 2020-01-02 to 2021-12-31.

Component	Average no. stocks held
Long	155
Short	136
Overall HML-33% Portfolio	291

The total return of the portfolio has exceptionally low volatility, and this allows for a Sharpe Ratio that is above that of the benchmark even though the total return is less than that of the benchmark. There are also a lower number of average stocks in the short side of the portfolio, this may be due to a large amount of the negative sentiment being in relation to stocks that have been removed from the SPX, as the dataset used includes only the latest composition of the index. It is also seen that the returns earned by each component and the overall approach are not significant. The t-stat for the short side of the portfolio is much larger than that of the long side which may provide evidence for hypothesis three and will be further examined in section 5.3 below.

Considering the market neutrality of the portfolio. To determine the market neutrality of the portfolio, the beta of each stock held was multiplied by the weighting of the stock held. This outputted a net beta for each day, the average of each day's beta was then taken to get the net market exposure of the portfolio over the period.

Daily Beta:

$$\beta_{Daily} = \sum_{t \in Long}^{T} \beta_{it}^{Long} * \frac{1}{n^{Long}} + \sum_{t \in Short}^{T} \beta_{it}^{Short} * \frac{1}{n^{Short}}$$

Where β_i^{Long} is the beta on each day t of each share i in the long side of the portfolio and β_i^{Short} beta on each day t of each share i in the short side of the portfolio.

And the overall portfolio Beta, β_{net}, is the average of the daily betas, $\mu\beta_{Daily}$:

$$\beta_{net} = \mu\beta_{Daily} = 0.016$$

Therefore, as expected the net beta of the portfolio is approximately zero, which results in limited market exposure for the HML-33% Twitter portfolio.

b) HML-5%

The HML-5% Twitter portfolio has been constructed similarly to that of the HML-33% Twitter portfolio, but this approach aims to capture the more extreme sentiment scores and overweight them in an attempt to earn greater profits.

As with the HML-33% portfolio, the initial results ignore both transaction and short selling costs. The long side of the portfolio had a negative return, which is in line with the relationship found in the previous section. There were, however, positive returns for both the short-side and total portfolio. But both the total returns and the risk-adjusted returns of the total portfolio were less than that of the benchmark. It is also seen that the returns earned by each component and the overall approach are not significant at any level but again it is observed that the t-stat on the short side was much larger than that of the long side.

Table 4a: Exploring returns of HML-5% Twitter portfolio with a 1-day holding period. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Component	Annualised Return	Annualised Volatility	Sharpe Ratio	T-stat	P-value
Long	-0.40 %	18.44%	-0.23	-0.031	0.975
Short	6.01%	19.05%	0.29	0.446	0.656
Overall HML-5% Portfolio	2.48%	4.95%	0.43	0.708	0.479

Table 4b: Average number of stocks in which a position is taken in HML-5% Twitter portfolio. Examining the daily average holding for the period 2020-01-02 to 2021-12-31.

Component	Average no. stocks held
Long	24
Short	23
Overall HML-5% Portfolio	47

The more concentrated holdings have resulted in higher volatility of the portfolio but have not resulted in a proportional increase in the returns and thus a lower Sharpe ratio is seen.

Examining the net beta for market neutrality, it is seen that the HML-5% Twitter portfolio is too approximately market neutral with a net beta of 0.026.

c) Proportional Portfolio

The construction of the proportional portfolio is more complex, under this approach the goal is to short all shares that have a sentiment score that is below the sentiment mean and long all shares with a sentiment score that is above the mean. The portfolio then also grants greater weighting to those shares with a more extreme sentiment score.

The weighting of each share will be calculated according to the following formula (Cui, Lam & Verma, 2016):

$$w_{it}^{Long} = \frac{SS_{it}^{Long} - \mu_t}{\sum_{i \in Long_t}(SS_{it}^{Long} - \mu_t)}, w_{it}^{Short} = \frac{\mu_t - SS_{it}^{Short}}{\sum_{i \in Short_t}(\mu_t - SS_{it}^{Short})}$$

Where w_{it}^{Long} is the weight of share i within the long portion of the portfolio on day t, w_{it}^{Short} is the weight of share i that falls within the short-side of the portfolio on day t, μ_t is the mean sentiment score on any given day t, SS_{it}^{Long} is the sentiment score of stock i in the long-side of the portfolio on any given day t and SS_{it}^{Short} is the sentiment score of stock i in the short-side of the portfolio on any given day t.

The proportional portfolio underperforms the previous two approaches considered, on both an absolute and risk-adjusted basis. But positive returns are still observed.

The proportional portfolio does not perform well. The long side of the portfolio only just earns a positive return, which is less than the annual risk-free rate of 0.34% which in turn results in a negative Sharpe Ratio for the long portion of the portfolio. The short side fares only marginally better, with the entire portfolio earning a lower return than the benchmark but having a greater Sharpe Ratio due to the lower volatility. Using the proportional portfolio, the overall portfolio returns are seen to be significant at a 10% level. However, the nature of the t-stat calculation can result in a large observation size causing the t-stat to be significant. In this case, there is a small effect size, which would result in the returns being insignificant or even negative if the transaction and short selling costs were introduced into the calculation.

Table 5a: Exploring returns of proportional Twitter portfolio with a 1-day holding period. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Component	Annualised Return	Annualised Volatility	Sharpe Ratio	T-stat	P-value
Long	0.19%	18.39%	-0.01	0.015	0.988
Short	5.28%	17.45%	0.28	0.427	0.670
Overall proportional Portfolio	4.33%	3.63%	1.09	1.685	0.093

Table 5b: Average number of stocks in which a position is taken in proportional Twitter portfolio. Examining the daily average holding for the period 2020-01-02 to 2021-12-31.

Component	Average no. stocks held
Long	72
Short	400
Overall proportional Portfolio	472

There is an odd situation where the formula used to calculate the weights applied to each stock favours underweighting shares, as seen with the large number of short positions taken relative to the number of long positions, as seen in the table 5b above. This may be due to more extreme positive sentiment values existing that increase the mean sentiment score, resulting in fewer shares being above the threshold to have a long position taken in them.

Finally, observing the net beta of the proportional portfolio approach it is observed that the proportional portfolio is too approximately market neutral, with a net beta of 0.04.

5.1.2.2. News Sentiment

In this section of the paper the validity of the signal or the IR, that was calculated previously, will be examined. The purpose of this section of the paper is to explore if the information generated from news sentiment differs from that generated by Twitter sentiment. The process is functionally the same, the difference exists in the source of the sentiment used to construct the portfolios.

a) HML-33%

Making use of the sentiment obtained from news articles, a positive return is observed for this approach, with the return generated from the short side being almost four times greater than that of the long side return. This has resulted in a total portfolio return that is less than that of the benchmark, but a Sharpe ratio that is above that of the benchmark, due to the low volatility of the earnings of the total portfolio. The higher Sharpe ratio relative to the HML-33% Twitter portfolio may show the greater predictive ability of news articles than that of Twitter. It is seen that the returns on the overall HML-33% News portfolio are significant at a 1% level. This provides further evidence that news articles provide more salient information than Twitter does.

Table 6a: Exploring returns of HML-33% News portfolio with a 1-day holding period. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Component	Annualised Return	Annualised Volatility	Sharpe Ratio	T-stat	P-value
Long	2.40%	16.40 %	0.13	0.207	0.836
Short	8.56%	18.21 %	0.45	0.664	0.507
Overall HML-33% Portfolio	7.00%	3.84 %	1.73	2.575	0.010

Table 6b: Average number of stocks in which a position is taken in HML-33% Twitter portfolio. Examining the daily average holding for the period 2020-01-02 to 2021-12-31.

Component	Average no. stocks held
Long	130
Short	94
Overall HML-33% Portfolio	224

There are a smaller average number of stocks where positions are taken relative to the HML-33% Twitter portfolio, this is likely due to an increased number of securities that have no news sentiment score attached to them.

Considering market neutrality, the net beta of the HML-33% News approach is -0.03 and thus it is approximately market neutral.

b) HML-5%

This approach generates positive returns but is once again below that of the benchmark. The total portfolio does have a greater Sharpe ratio than that of the benchmark, again, due to the lower volatility. Again, it is seen that the overall portfolio return is significant, however, this is less significant than in the HML-33% news portfolio as it is only significant at a 5% level.

Table 7a: Exploring returns of HML-5% News portfolio with a 1-day holding period. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Component	Annualised Return	Annualised Volatility	Sharpe Ratio	T-stat	P-value
Long	5.37%	17.60 %	0.29	0.431	0.667
Short	8.43%	19.50 %	0.41	0.611	0.541
Overall HML-5% Portfolio	8.65%	6.22 %	1.34	1.965	0.050

Table 7b: Average number of stocks in which a position is taken in HML-5% Twitter portfolio. Examining the daily average holding for the period 2020-01-02 to 2021-12-31.

Component	Average no. stocks held
Long	22
Short	24
Overall HML-33% Portfolio	46

Considering the market neutrality of the portfolio, a net beta of -0.02 is observed meaning that the portfolio is approximately market neutral.

c) Proportional Portfolio

The same approach to constructing the proportional portfolio has been taken here as done previously. The aim is to take advantage of the more extreme sentiment scores, while still holding a position in the securities with less extreme sentiment scores.

There is a similar situation in this proportional portfolio to the Twitter proportional portfolio, where a majority of securities form part of the short side of the portfolio. This results in the sentiment scores of the long-portfolio being more meaningful and thus breaking the trend of the HML portfolios and outperforming the short-side. Here it is observed that the overall portfolio returns are significant at a 10% level.

Table 8a: Exploring returns of proportional News portfolio with a 1-day holding period. Examining the daily returns for the period 2020-01-02 to 2021-12-31.

Component	Annualised Return	Annualised Volatility	Sharpe Ratio	T-stat	P-value
Long	3.68%	16.70%	0.20	0.311	0.756
Short	4.65%	19.21%	0.22	0.342	0.732
Overall Proportional Portfolio	5.84%	4.49%	1.22	1.838	0.067

Table 8b: Average number of stocks in which a position is taken in proportional News portfolio. Examining the daily average holding for the period 2020-01-02 to 2021-12-31.

Component	Average no. stocks held
Long	431
Short	43
Overall Proportional Portfolio	474

Another interesting phenomenon exists where a large portion of the stocks' sentiment score lies above that of the mean. This may be due to news articles taking extreme negative stances when a negative stance is taken which results in a low mean with only a few values sitting below the mean. This is seen in table 8b, with a larger number of stocks where a long position is taken, while only 43 shares in which a short position is taken.

5.1.3. Testing for Statistical Significance of Returns

Testing for statistical significance of returns can be done in a number of ways, however, this paper will examine for outperformance on a risk-adjusted basis by comparing Sharpe Ratios, making use of the following formula (Grinold & Kahn, 1999):

$$\left(\frac{\bar{r}_p}{\sigma_p} - \frac{\bar{r}_b}{\sigma_b}\right) > 2 * \sqrt{\frac{2}{N}}$$

Where $\bar{r}_p$ is the return of a portfolio less the risk-free rate, $\bar{r}_b$ is the return of the benchmark less the risk-free rate, σ_p is the volatility of the portfolio, σ_b is the volatility of the benchmark and N is equal to the number of periods observed and since all the figures are annualised N will equal 1 in this case.

The formula essentially states that if the difference between two Sharpe Ratios is greater than 2.83 (as we are considering an annualised figure, N = 1) there is a statistically significant outperformance at a 95% level.

Table 9: Significance of returns with a 1-day holding period. Examining the outperformance of the sentiment portfolios constructed for the period 2020-01-02 to 2021-12-31.

Portfolio	$\frac{\bar{r}_p}{\sigma_p} - \frac{\bar{r}_b}{\sigma_b}$	**Significant outperformance**	**Significant Under performance**
HML-33% Twitter	0.72	False	False
HML-5% Twitter	-0.45	False	False
Proportional Twitter	0.21	False	False
HML-33% News	0.85	False	False
HML-5% News	0.46	False	False
Proportional News	0.34	False	False

While five out of six portfolios outperform the benchmark on a risk-adjusted basis, this outperformance is not significant so cannot be pointed to as a success of these portfolios. It should also be noted that these figures are ignoring transaction and short selling costs which, due to the low returns of each approach, will likely result in every portfolio earning a negative return.

The HML-33% News portfolio showed the highest risk-adjusted returns. While this is not a statistically significant figure it is in line with the findings of section 1 of part 1 that showed that news sentiment provided the highest IR.

5.1.4. Examining Alphas of the portfolios

Expanding on both the Twitter and News portfolios making use of the HML-33% approach, as it produced the greatest risk-adjusted returns in both instances.

For each stock for each day, the representative beta figure has been pulled. This beta figure has been calculated by Bloomberg Terminal. This allows the calculation of each stock's alpha on each day, by making use of the following formula (Grinold & Khan, 1999):

$$\alpha_{it} = R_{it} - \beta_{it} R_{bt}$$

Where α_{it} is the alpha on stock i on a given day t, R_{it} is the actual return that stock i earned on day t, β_{it} is the beta of stock i on a given day t and R_{bt} is the return of the benchmark on a given day t.

These individual alphas are then considered in the context of each of the portfolios, which adds all the alphas of all the stocks that the portfolio holds on each day using the following formula:

$$\alpha_d = \sum_{i=1}^{N} \frac{\alpha_{it}}{n_t}$$

Where α_d is the weighted average alpha of each day, and n_t is the number of stocks held during day t. Each day's alpha is then summed and then divided by the number of trading days in order to give a daily average alpha of the portfolio, using the formula:

$$\alpha_p = \left(\sum_{t=1}^{T=505} \alpha_d\right) * \frac{1}{T}$$

Where α_p is the daily average alpha of the portfolio, *t* represents the number of trading days, and this has been divided by 505 as there are 505 trading days observed.

In order to calculate a t-stat to determine the significance of the alphas, the volatility of the portfolio must be calculated. This has been calculated as the tracking error of each portfolio ($R_p - \beta_i R_b$).

$$\omega_p = \sigma(R_p - \beta_i R_b)$$

The t-stat for alpha is approximately equal to the alpha of the portfolio divided by the risk of the portfolio multiplied by the square root of the number of periods examined (Grinold & Kahn, 1999):

$$t_p = \left(\frac{\alpha_p}{\omega_p}\right) * \sqrt{T}$$

Where *T* is to account for time and is equal to 1 in this case as figures are considered at a daily level.

Table 10: Significance of Alphas of 1-day holding period. Examining the average daily alpha for the period 2020-01-02 to 2021-12-31.

Portfolio	Alpha	Tracking Error	T-stat	P-value
HML-33% Twitter	-0.07%	0.24%	-0.31	0.76
HML-33% News	-0.07%	0.24%	-0.29	0.77

This results in negative average daily alphas for both portfolios this result was expected due to each portfolio earning lower returns than the benchmark. The high P-values however show that the returns of the portfolio do not significantly differ from the returns of the benchmark. Dybvig & Ross (1985) stated that a positive Sharpe Ratio implies a positive alpha. While, at face value, this result appears as a contradiction of this, the positive Sharpe Ratios obtained above were as outperformance of the risk-free rate, whereas the negative alphas seen here are as underperformance of the benchmark. If the hurdle rate used for the calculation of Sharpe Ratios was the return of the benchmark, then the Sharpe Ratios would too be negative.

5.1.5. Conclusion to Section 5.1 Trading on Twitter Sentiment

Part 1 of the results of this paper has examined the predictive power of the current day's sentiment, obtained from both Twitter and news sources, on the next day's returns. As initially outlined through the IR at the beginning of this section Twitter's has little to no predictive ability on next day returns, which is counter what the studies of Oh & Sheng (2011) and Cui, Lam & Verma. (2016) found. This lack of correlation was shown through the use of t-stats which showed no significant correlations with the exception of share returns and next day news sentiment which was seen to be significant at a 10% level.

The portfolio constructions used, improved returns, as the IR obtained in section 1 would result in the expectation that all returns would have been negative. The returns of the Twitter portfolios were statistically insignificant with the exception of the proportional Twitter portfolio where the returns earned were significant at a 10% level. The news portfolios produced better results with each portfolio earning significant results when looking at the overall return. The HML-33% news portfolio performed best having results that were significant at a 1% level, while the HML-5% news and proportional news portfolios were significant at a 5% and 10% level respectively.

However, while these results were significant their effect sizes were notably small with the t-stat being inflated by the large number of days in the observation period. This would result in the returns not being significant and likely being negative once transaction and short selling costs are reintroduced.

When considering alphas, it was also observed that even the most successful of the portfolios underperformed the benchmark. However, due to statistical insignificance, it can be stated that their alphas did not differ significantly from 0.

The market neutrality of the portfolios allows for the value and the returns of the portfolio to be removed from that of the market. When graphing the value of the portfolios[5], this is obvious as the portfolios were not affected by the large Covid-19 sell-off seen in March of 2020. Thus, making this approach favourable during extreme market downturns. However, this is not unique to just this approach, any market neutral approach would see the same benefits.

These results are consistent with Deng et al. (2018) showing that significant earnings cannot be earned with daily trading. The results also provide support for Appraisal Theory (Roseman & Smith, 2001), which states that individuals' emotions are in response to external events and stimuli. There is also evidence in line with Otoo (1999) that the returns of the market affect investor sentiment.

[5] See Appendix 3 for Graph of Twitter portfolios' value relative to the benchmark over period.

Part 1 has provided evidence that returns are not predicted by microblog sentiment, but that microblog sentiment is predicted by returns. In section 1, a large correlation between returns and next day investor sentiment is seen. This effect is more pronounced on Twitter than in news articles, this is likely due to the more casual nature and lower barriers to entry that social media platforms provide relative to that of traditional news sources. However, this is when positions are entered into at market open and exited at market close, Deng et al. (2018) and Cui, Lam & Verma (2016) have found evidence of profitable returns at a higher frequency. The high speed with which information is processed may have resulted in daily tradable opportunities no longer existing, but intra-daily trades still being profitable.

The semi-strong form efficient market hypothesis (Fama, 1970) states that only obviously available public information is priced in efficiently. As NLP is still being developed and the data used in this study was obtained from Bloomberg Terminal, which has a high financial barrier to entry, a case could be made that Twitter sentiment is not obviously publicly available and thus this information may take multiple days to be fully priced in. All the portfolios considered in part had a one-day holding period, holding from open until close each day. However, this may not be an optimal holding period as investor sentiment may take multiple days to be fully priced in. Thus, in part 2 this paper will take the HML-33% Twitter portfolio, as it was the best performing portfolio that made use of Twitter sentiment and examines the returns obtained over a multi-day holding period. This will be done by making use of an event study model.

5.2. Optimal Holding Period

As a 1-day holding period does not produce significant results this paper will now examine if a multi-day holding period outperforms a 1-day holding period. This will be done by first making use of an event study, and then looking for evidence that information takes a few days to be priced in using Cohen & Frazzini's (2008) URC.

5.2.1. Event Study

Part two of this paper examines the most successful portfolio from part 1, the HML-33% Twitter portfolio, and makes use of an event study model to discern the optimal holding period. The event study package used 'Eventstudy, Python Package' relies heavily on MacKinlay's paper (1997) detailing event studies.

A market model approach has been used as it compares the returns of the portfolio against that of the market, which is what was done in part 1. The market model assumes a stable linear relationship between the market return, which is the SPX in this case, and the return of the portfolios. The market model's linear relationship with each portfolio follows the following formula (Fama et al., 1969):

$$R_{it} = \alpha_i + \beta_i R_{bt} + \varepsilon_{it}$$

Assuming that on average ε_{it} is zero and $var(\varepsilon_{it}) = \sigma^2_{\varepsilon_i}$. Where R_{it} are the returns on portfolio i for period t and R_{bt} is the return on the benchmark for period t.

This package requires some prior data to fit a model to each of the shares examined. Given the turbulent times that the world has faced over the past two years, a shorter period was used as it better fits the higher volatility environment that global equity markets have been in. 18-months prior data has been used to fit adequate models for the event study, thus the estimation period for the event studies extends from 2018-06-01 until 2019-12-31. With the events event window lasting from 2020-01-02 until 2021-12-31.

In this paper, 2 separate event studies are conducted with 24 events for both the long and short side of the portfolio as it has been theorised that the different emotions attached to the positive and negative sentiment will result in different speeds of the information being priced in. These are then aggregated into the tables seen below. The event is defined as a stock having an extreme sentiment score, in this case falling with the top third or bottom third of shares when ranked by sentiment. The portfolios are then created, and the positions are entered into on the first business day of the month. The positions are then held for 3 weeks or 15 business days and the returns are analysed and observed for the significance of daily returns. Note the returns calculated in part 2 differ from that of part 1, as part 1's return was intra-daily, between the opening and closing price on a given day. Whereas in part 2 returns are calculated as the change from close to close. The returns in part 2 are calculated as:

$$R_{L,it} = \frac{P_{Close,it}}{P_{Close,i(t-1)}} - 1, R_{S,it} = \frac{P_{Close,i(t-1)}}{P_{Close,it}} - 1$$

Where $R_{L,it}$ is the return on stock i held long for period t, $R_{S,it}$ is the return on stock i held short for period t and $Pclose_t$ is the closing price of stock i on day t and $P_{Close,i(t-1)}$ is the closing price of stock i on day t.

The abnormal return (AR_{it}) is the actual return earned by the security (R_{it}) less the expected return if the event did not occur $\left(\hat{\alpha}_i + \hat{\beta}_i R_{bt}\right)$ and is calculated for each stock using the formula (MacKinlay, 1997):

$$AR_{it} = R_{it} - \left(\hat{\alpha}_i + \hat{\beta}_i R_{bt}\right)$$

This paper is doing an event study where there are multiple events that are then accumulated and aggregated, and this results in the AR_{it} figure becoming an AAR_{it} or average abnormal return. This aggregation exists for all calculations as seen below.

$$AAR_{it} = \frac{\sum_{n=1}^{n} AR_{nt}}{N}$$

$\sigma^2(\mathrm{AAR})$ is the variance of AAR, and is calculated as:

$$\sigma^2(AR_{it}) = \sigma_{\varepsilon_i}^2 + \frac{1}{L_1}\left[1 + \frac{(R_{bt} - \hat{\mu}_b)^2}{\hat{\sigma}_b^2}\right]$$

The variance of AAR has two components, the first is the distribution variance $\sigma_{\varepsilon_i}^2$ and the second is the variance that is a result of sampling error in $\hat{\alpha}_i$ and $\hat{\beta}_i$. Where L_1 is the length of the estimation window. Due to the nature of this formula, it is seen that the second part of the equation tends to zero as the event window increases in length. $\hat{\sigma}_b^2$ is the variance of the benchmark and $\hat{\mu}_b$ is the average cumulative returns of the benchmark over the period examined.

The cumulative average abnormal return is calculated using the formula:

$$CAR_{i(-5,+15)} = \sum_{t=-5}^{15} \varepsilon_{it}$$

Cumulative average abnormal returns ($CAAR_i$) for the sample of N portfolios in each time period t:

$$CAAR_{i(-5,+15)} = \frac{\sum_{n=1}^{n} CAR_{nt}}{N}$$

$\sigma^2(\mathrm{CAAR})$ is the variance of CAAR.

$$\sigma^2(CAAR_i) = L_2\sigma^2(AR_{it})$$

Where L_2 is the length of the event window.

Conducting the event study results in the data having a student T-distribution, and this allows the calculation of a Patell T-test (1976) using the following a summarised formula stated in Corrado (2011):

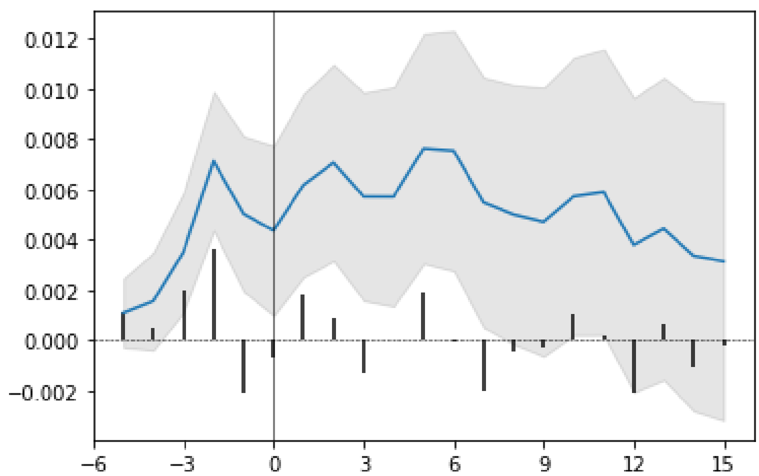
$$t_{CAAR} = \frac{CAAR_i}{\sigma_{CAAR}}$$

And from this T-stat, using a T-table a P-value can be obtained.

Table 11: Long side output of event study. Event study conducted with a -5, +15 window for each month within the two-year period. Day zero is the day immediately following extreme Twitter sentiment. With the significance levels denoted by the asterisks, where *** shows significance at a 1% level, ** shows significance at a 5% level and * shows significance at a 10% level.

	AAR	T-stat	P-value	CAAR	$\sigma(\mathbf{CAAR})$	T-stat	P-value
-5	0.001	0.93	0.353	0.001	0.00108	1.01	0.314
-4	0	0.00	1	0.002	0.00152	1.02	0.307
-3	0.002 *	1.85	0.066	0.004 *	0.00186	1.88	0.06
-2	0.004 ***	3.70	0.0002	0.007 ***	0.00215	3.31	0.001
-1	-0.002 *	-1.85	0.066	0.005 **	0.00241	2.09	0.037
0	-0.001	-0.93	0.353	0.004 *	0.00264	1.65	0.098
1	0.002 *	1.85	0.066	0.006 **	0.00285	2.16	0.03
2	0.001	0.93	0.353	0.007 **	0.00304	2.32	0.021
3	-0.001	-0.93	0.353	0.006 *	0.00323	1.77	0.077
4	0	0.00	1	0.006 *	0.0034	1.68	0.094
5	0.002 *	1.85	0.066	0.008 **	0.00357	2.13	0.033
6	0	0.00	1	0.008 **	0.00373	2.02	0.043
7	-0.002 *	-1.85	0.066	0.005	0.00388	1.41	0.158
8	0	0.00	1	0.005	0.00403	1.24	0.215
9	0	0.00	1	0.005	0.00417	1.13	0.259
10	0.001	0.93	0.353	0.006	0.00431	1.33	0.185
11	0	0.00	1	0.006	0.00444	1.33	0.185
12	-0.002 *	-1.85	0.066	0.004	0.00457	0.83	0.407
13	0.001	0.93	0.353	0.004	0.00469	0.94	0.345
14	-0.001	-0.93	0.353	0.003	0.00481	0.7	0.487
15	0	0.00	1	0.003	0.00493	0.63	0.525

Figure 1: Graphing CAAR of the long side of the portfolio. Where day 0 is the day immediately following the day in which a stock has an extreme sentiment score and the blue line plots the CAAR as seen in table 11, the shaded areas represent the limits of a 90% confidence interval and the grey bars along the x-axis are the AAR's for each day.



Considering both the table of returns and the graph above, it is seen that the positive sentiment is priced in two days before (day -2) the sentiment is registered on Twitter. While significant returns do present after the event date, the CAAR is either below that of day -2 or only slightly above (0.001) which is further evidence that the returns on securities are a predictor of microblog sentiment.

Examining the AARs it is seen that the most significant return occurs on day -2. With the returns after the event date only showing significance and positive returns on days 1 and 5. With these returns only being significant at a 10% level.

Interestingly an increase in returns is seen after the event date, this may be an aftershock effect as those who observe the positive sentiment that was caused by the initial price increase start to buy into the securities, causing the returns to increase from day 0 until day 5. This may be seen as a momentum effect. It is also noteworthy that shares that have a positive sentiment at day 0, have persistent abnormal returns throughout the observation period.

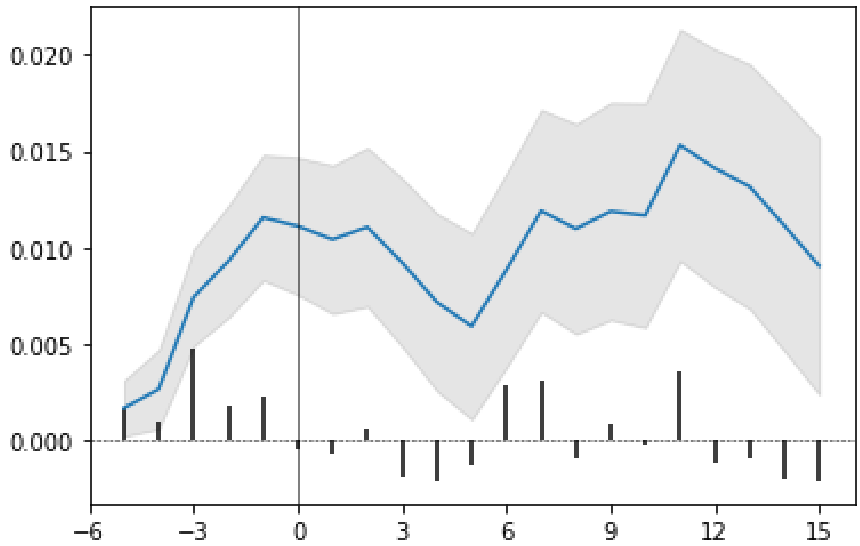
Therefore, it can be stated that an optimal holding period for the long side of the portfolio would be five business days, as it is seen that the CAAR is positive and significant at a 5% level.

Table 12: Short side output of event study. Event study conducted with a -5, +15 window for each month within the two-year period. Day zero is the day immediately following extreme Twitter sentiment. With the significance levels denoted by the asterisks, where *** shows significance at a 1% level, ** shows significance at a 5% level and * shows significance at a 10% level.

	AAR	T-stat	P-value	CAAR	$\boldsymbol{\sigma}(\mathbf{CAAR})$	T-stat	P-value
-5	0.002 *	1.754386	0.080	0.002	0.00114	1.48	0.138
-4	0.001	0.877193	0.381	0.003 *	0.00161	1.66	0.096
-3	0.005 ***	4.385965	0	0.007 ***	0.00197	3.78	0
-2	0.002 *	1.754386	0.080	0.009 ***	0.00227	4.09	0
-1	0.002 *	1.754386	0.080	0.012 ***	0.00254	4.56	0
0	0	0	1	0.011 ***	0.00278	4	0
1	-0.001	-0.87719	0.381	0.01 ***	0.00301	3.47	0.001
2	0.001	0.877193	0.381	0.011 ***	0.00321	3.45	0.001
3	-0.002 *	-1.75439	0.080	0.009 ***	0.00341	2.71	0.007
4	-0.002 *	-1.75439	0.080	0.007 **	0.00359	2	0.046
5	-0.001	-0.87719	0.381	0.006	0.00377	1.57	0.116
6	0.003 ***	2.631579	0.009	0.009 **	0.00394	2.25	0.025
7	0.003 ***	2.631579	0.009	0.012 ***	0.0041	2.91	0.004
8	-0.001	-0.87719	0.381	0.011 **	0.00425	2.58	0.01
9	0.001	0.877193	0.381	0.012 ***	0.0044	2.7	0.007
10	0	0	1	0.012 **	0.00454	2.57	0.01
11	0.004 ***	3.508772	0.005	0.015 ***	0.00468	3.27	0.001
12	-0.001	-0.87719	0.381	0.014 ***	0.00482	2.93	0.003
13	-0.001	-0.87719	0.381	0.013 ***	0.00495	2.66	0.008
14	-0.002 *	-1.75439	0.080	0.011 **	0.00508	2.19	0.028

15	-0.002 *	-1.75439	0.080	0.009 *	0.00521	1.74	0.082

Figure 2: Graphing CAAR of the short side of the portfolio. Where day 0 is the day immediately following the day in which a stock has an extreme sentiment score and the blue line plots the CAAR as seen in table 12, the shaded areas represent the limits of a 90% confidence interval and the grey bars along the x-axis are the AAR's for each day.



Examining the short-side event study an interesting phenomenon is observed, with the most significant return earned the day before the event occurs, or by day -1. This is an indication that the negative sentiment that is observed is a result of poor returns previously earned by the security, as the short side of the portfolio has positive returns when the value of the securities invested in decreases. The short side shows significant CAARs from day -2 through to day 4. However, there is a drop off in the returns earned after day -1, thus no tradable opportunity seems to exist if abnormal returns are considered.

Considering the AARs, the most significant return is seen on day -3, with positive significant returns at a 1% level seen on days 6 and 7 which seem to be sharp mean reversion.

However, it is noteworthy that holding from day 5 until day 11, where CAAR peaks with a 1% significance level. This may be the result of people 'buying the dip' ending after the poor performance between day -5 and day 5 and a continuation of the negative momentum that the security had been experiencing.

5.2.2. Underreaction of the Market

The Underreaction coefficient (URC) was theorised by Cohen and Frazzini (2008) and shows the under or overreaction of the market to an information shock. The formula that their paper outlined divides the return of period 1 over the total return over the periods examined. This paper will modify this, in a similar manner as Oh & Sheng (2011) did, and examine the next day returns over the total returns over 10 days, using the following formula (Cohen & Frazzini, 2008):

$$URC = \frac{1\ Day\ Return}{CR_{day10}}$$

Where CR_{day10} is the Cumulative return earned from day 1 until day 10.

An approach similar to that of the event study above will be used, with the returns of the first day of the month being used as the "Day 1 Return" etc., and this will be conducted for each month and then aggregated across all 24 months observed. This study will make use of a 10-day time period, as both the event study conducted prior and Oh & Sheng (2011) provide evidence for a 10-day period being adequate for a market to fully incorporate information.

Table 13: Exploring the URC over a 10-day period. Where the URC is stated, and each days' return is listed.

	URC	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	CR_{day10}
Long side	0.146	0.003	0.006	0.004	0.002	0.004	0.004	0.001	-0.004	-0.005	0.007	0.023
Short side	0.075	-0.001	-0.004	-0.004	-0.002	-0.005	-0.003	-0.001	0.006	0.009	-0.007	-0.012

There is a large underreaction to the information, this is seen through the URC that is well below 1. If the information was priced in on day 1, and the market completely efficient, the URC would sit closer to 1.

The long side returns experience an underreaction, as seen by the low URC of 0.146. This signals that the information generated from investor sentiment is not quickly priced into securities. Although this low URC may be due to much of the information being priced into the security before the extreme sentiment is seen, as figure 1 provides evidence for.

Note, the short side returns have been calculated in a manner similar to that of the event study, with the daily return calculated as:

$$R_{S,it} = \frac{P_{Close,i(t-1)}}{P_{Close,it}} - 1$$

It is noteworthy that the cumulative average return of the short side of the portfolio over 10 days of all 24 months is negative. This, along with a negative average return on day 1 has allowed for a URC to be calculated within the range defined by Cohen and Frazzini (2008). However, negative returns may provide issues for the measure that they have created, as the figure only accounts for a figure between 0 and 1. This may bring into question the usefulness of this measure in this study.

However, a URC of 0.075 shows an underreaction of the market, this is on the back of the shares within the portfolios examined falling into the bottom 33% of shares when ranked by sentiment. As outlined by the event study prior, much of the information seems to be priced in before the day in which the event occurs. Thus, what is observed here is the reversion to the mean, this consists of the shares within the short side of the portfolio appreciating in value in the wake of the negative information being processed by the market. This price appreciation results in the short side of the portfolio having negative returns over the 10 days examined. Note that the returns from part 1 for the HML-33% Twitter portfolio were positive for a 1-day holding period, this discrepancy exists as the returns of part 1 were calculated from open to close on a given day whereas the returns calculated in part 2 are calculated close to close on consecutive trading days.

It should be noted that following from section 5.1 the requirements for the URC as outlined by Cohen & Frazzini (2008) are not necessarily met, and thus the results from the URC may be questioned. While the investor sentiment is available, due to it being on public forums such as social media the discerning of the sentiment is not a straightforward process. And the sentiment obtained from Twitter has been shown to not be salient for a 1-day holding period. Thus, the URC will not be relied upon for making any statements but will be used to expand on other findings.

5.2.3. Further Analysing Optimal Holding Periods

After examining the graphs and tables from the event study it is seen that the long side of the portfolio has an optimal holding period of 5 days, while the short side of the portfolio has an optimal holding period of 11 days.

The long return is calculated as:

$$\sum_{i=1}^{N} \sum_{t=1}^{T=5} R_{L,it}$$

Where $R_{L,it}$, is the return of each stock i held long on each day t in the holding period, n is the number of stocks held and t is the time period and is equal to 5 as this is the deemed optimal holding period.

The short return is calculated as:

$$\sum_{i=1}^{N} \sum_{t=1}^{T=11} R_{S,it}$$

Where $R_{S,it}$ is the return of each stock i held short on each day t in the holding period, $t = 11$ as this is the deemed optimal holding period.

Each return is then converted to an annualised figure by dividing by 2, as the return is calculated over two years.

The volatility of each approach is directly observable and is calculated as the standard deviation of the returns. This figure is then scaled to make it an annualised figure. The long side returns' standard deviation is multiplied by $\sqrt{\frac{253}{5}}$, as there are 505 days observed and the portfolio holds each position for 5 days. The short side returns' standard deviation is multiplied by $\sqrt{\frac{253}{11}}$, as there are 505 days observed and the portfolio holds each position for 11 days.

After calculating the returns accounting for optimal holding periods, the following returns are obtained.

Table 14: Returns of optimal holding periods of long and short sides of the portfolio. Observing the returns and Sharpe Ratios of optimal holding period returns, for the period 2020-01-02 to 2021-12-31.

	Annualised Return	**Annualised Volatility**	**Sharpe Ratio**
Long: 5-day holding period	27.16%	25.97%	1.05
Short: 11-day holding period	-5.56%	40.47%	-0.07
Extreme negative sentiment shares held long for 5 days	27.77%	29.32%	0.95

Where "Long: 5-day holding period" is the portfolio created by longing shares that have a sentiment score in the top third of shares and then holding this for five days, "Short: 11-day holding period" is the portfolio created by shorting shares that fell within the bottom third of shares when ranked by sentiment and this is held for elven days and "Extreme negative sentiment shares held long for 5 days" is the portfolio where shares that have fallen in the bottom third of shares when ranked by sentiment are held long and held for five days in order to take advantage of the mean-reversion effect that seems to be present.

From table 14, it can be seen that the long side of the portfolio benefits from a multi-day holding period, whereas even the multi-day optimal holding period for the short side of the portfolio produces a negative return. Thus, it can be stated that the optimal holding period for the short side of the portfolio, is open to close on the day immediately after the extreme sentiment, and the optimal holding period for the long side of the portfolio is 5 days.

Before testing for the significance of the long side return Figure 2, seems to display a tradable opportunity, if the stocks that had extreme negative sentiment are held long and held for 5 days. Thus, using the formula used to calculate the returns of the 5-day long portfolio will be reused to obtain the returns of a portfolio that has used the negative sentiment as a buy signal.

Considering all three holdings, the 5-day holding period long portfolio produces the greatest risk-adjusted returns, while the 5-day holding period portfolio that used the negative sentiment as a buy signal produced the highest annual return, but at higher volatility which resulted in a lower Sharpe Ratio.

This return seems to be a mean reversion that the securities experience after a period of negative returns that drove the extreme negative sentiment. This mean reversion lasts for a few days before the negative momentum becomes the main driver of returns.

Constructing a portfolio with a 5-day holding period, where 50% of the portfolio holds a long position in shares that have extreme positive sentiment, in an equal-weighted portfolio, and where 50% of the portfolio holds a long position in shares that have extreme negative sentiment, in an equal-weighted portfolio. This results in the following figures, with the benchmark and 1-day holding period figures from part 1 restated for the reader's convenience.

Table 15: Returns of top performing 1-day and optimal holding period portfolios. Listing figures for a 1-day and multi-day portfolio return alongside the benchmark figures for the period 2020-01-02 to 2021-12-31.

	Annualised return	**Annualised Volatility**	**Sharpe Ratio**
5-Day Holding Period Portfolio	30.13%	27.31%	1.09
1-Day Holding Period Portfolio	6.33%	3.74%	1.60
Benchmark	23.31%	26.01%	0.88

Table 16a: Testing for significance of returns of multi-day portfolio relative to 1-day holding period. Examining for significantly differing performance of a multi-day holding period and a 1-day holding period, for the period 2020-01-02 to 2021-12-31. Where $\bar{r}_M, \sigma_M$ are the return less the risk-free rate and volatility of the multi-day holding period portfolio and $\bar{r}_O, \sigma_O$ are the return less the risk-free rate and volatility of the 1-day holding period portfolio.

Portfolio	$\frac{\bar{r}_M}{\sigma_M} - \frac{\bar{r}_O}{\sigma_O}$	**Significant outperformance**	**Significant Under performance**
5-day holding period portfolio	-0.51	False	False

Table 16b: Testing for significance of returns of multi-day portfolio relative to 1-day holding period. Examining for significantly differing performance of a multi-day holding period and a 1-day holding period making use of a t-stat, for the period 2020-01-02 to 2021-12-31.

Portfolio	**t-stat**	**p-value**
5-day holding period portfolio	0.87	0.385

Table 17a: Testing for significance of risk adjusted returns of multi-day portfolio relative to the benchmark. Examining for significantly differing performance of a multi-day holding period and the benchmark portfolio, for the period 2020-01-02 to 2021-12-31.

Portfolio	$\frac{\bar{r}_p}{\sigma_p} - \frac{\bar{r}_b}{\sigma_b}$	**Significant outperformance**	**Significant Under performance**
5-day holding period portfolio	0.21	False	False

Table 17b: Testing for significance of returns of multi-day portfolio. Examining for significantly differing performance of a multi-day holding period and the benchmark portfolio making use of a t-stat, for the period 2020-01-02 to 2021-12-31.

Portfolio	**T-stat**	**p-value**
5-day holding period portfolio	1.10	0.274

The 1-day holding period outperforms the 5-day holding period on a risk-adjusted basis due to the low annualised volatility figure but earns lower returns. However, this is not a significant outperformance. The multi-day holding period earned greater returns than the 1-day holding period but with a p-value of 0.385, it is seen that this outperformance is not significant.

The 5-day holding period portfolio outperforms the benchmark both on a return basis and on a risk-adjusted basis, however, this outperformance is not significant as seen in tables 17 a and b. There is the added benefit of reduced transaction costs making use of the 5-day holding period and as the portfolio only takes long positions there are no additional short selling costs incurred by the portfolio. However, this portfolio will now have a strong positive relationship to the returns of the market and thus has lost the market neutrality that the long-short portfolios of part 1 provided.

5.2.4. Conclusion to Section 5.2. Optimal Holding Period

Observing the long and short sides of the portfolio separately allows for the observation of how the different emotional triggers (positive and negative) cause those in the market to react. The differing reactions will be further unpacked in part 3, later in this paper.

Examining the event study, where the security having extreme sentiment was the defined event it is seen that the most significant returns are observed before the event date. While significant results are observed after the event date, these returns are both smaller and less significant than those observed before the event date and may only be as a result of the elevated CAAR seen in the days prior to the event. The returns observed after the event date seem to be a reaction to the returns earned before the event date, as is the sentiment that is obtained from Tweets. This phenomenon further adds to the argument that is prevailing in this paper that returns are a predictor of investor sentiment, as stated in the conclusion of Part 1. It should be noted that this paper holds some reservations about the optimal holding period due to the “cherry-picked” nature of the optimal holding periods.

The optimal holding period was established as, being five days when using both positive and negative sentiment as a signal to establish long positions. In contrast to the initial approaches of this study, the best way to extract value out of negative sentiment was to long the shares rather than short them. This trade seems to take advantage of a mean reversion that occurs after a period of poor returns, a period which the negative sentiment seems to signal has ended. This results in the extreme negative sentiment presenting as a buy signal rather than as a sell signal as previously thought. The returns of the multi-day holding period outperform both the 1-day holding period and the benchmark, however, the returns are not significant and are likely as a result of a large beta exposure in an upward market.

The multi-day holding period being optimal for the returns of the portfolio provide evidence against the semi-strong form of market efficiency (Fama, 1970). However, to fully challenge the EMH this return would need to be significant and the signal to trade, the investor sentiment would have to be considered ‘’obviously publicly available” both of which do not hold.

In part 3 this paper will explore if there is a differing response from investors to positive and negative sentiment. The results from the previous two sections, point to a low predictive ability of Twitter sentiment, and as a result, no difference in reaction is expected over a 1-day holding period, however, there is potential for a difference to exist when each side is held for their optimal holding period.

5.3. Bullish vs Bearish Sentiment

In part 3 this paper will examine the positive sentiment (bullish) and negative sentiment (bearish) returns from each of the six approaches and evaluate if there is significant outperformance of either portfolio. Kahneman and Tversky's (1979) asymmetric risk function would result in the expectation that the returns earned from the short side of the portfolios would outperform the returns earned from the long side. This is due to people being risk-loving on the upside, causing individuals to overinflate the possibilities of upside returns and having positive sentiment not translate into returns. And risk-averse on the downside, causing a much harsher reaction to the negative information, or sentiment in this case, which would lead to a more rapid decrease in the value of the security and thus a larger positive return in a short portfolio.

Again, the comparison of returns will be conducted using the Sharpe Ratio comparison formula from part 1 but modified to compare returns earned from the reaction to negative and positive sentiment (Grinold & Kahn, 1999):

$$\left(\frac{r_{-}}{\sigma_{-}} - \frac{r_{+}}{\sigma_{+}}\right) > 2 * \sqrt{\frac{2}{N}}$$

Where r_{-} is the annualised return on the trade made in accordance with the negative sentiment after subtracting the risk-free rate, r_{+} is the annualised return on the trade made in accordance with the positive sentiment after subtracting the risk-free rate, σ_{-} is the volatility of the negative sentiment trade’s returns and σ_{+} is the volatility of the positive sentiment trade’s returns.

As the figures have been annualised, N = 1. Therefore, it can be concluded that if the difference between Sharpe Ratios (or $\left(\frac{\bar{r}_{-}}{\sigma_{-}} - \frac{\bar{r}_{-}}{\sigma_{+}}\right)$) is greater than 2.83 then it can be concluded that there is significant outperformance of the short side of the portfolio at a 95% or greater level, and if the difference between Sharpe Ratios is less than -2.83 then there is significant outperformance of the long side of the portfolio at a 95% or greater level.

The significance of returns can also be seen by making use of a t-stat, which can be calculated using the following formula:

$$t_{p(N-1)} = \frac{\bar{r}_{-} - \bar{r}_{+}}{\sigma_{-}}$$

5.3.1. One-day Holding Period

Table 18a: Examining for significant outperformance of negative sentiment driven trades over positive sentiment driven trades for all portfolios with a 1-day holding period on a risk-adjusted basis. Examining for a significantly different return of the long and short side of the portfolio, for the period 2020-01-02 to 2021-12-31.

	$\left(\frac{\bar{r}_{-}}{\sigma_{-}} - \frac{\bar{r}_{+}}{\sigma_{+}}\right)$	**Significant Outperformance of short side**	**Significant Outperformance of long side**
HML-33% Twitter	0.46	False	False
HML-5% Twitter	0.29	False	False
Proportional Twitter	0.25	False	False
HML-33% News	0.45	False	False
HML-5% News	0.41	False	False
Proportional News	0.22	False	False

Table 18b: Examining for significant outperformance of negative sentiment driven trades over positive sentiment driven trades for all portfolios with a 1-day holding period on a return basis. Examining for a significantly different return of the long and short side of the portfolio making use of a t-stat, for the period 2020-01-02 to 2021-12-31.

	T-stat	**p-value**
HML-33% Twitter	0.44	0.660
HML-5% Twitter	0.34	0.734
Proportional Twitter	0.29	0.772
HML-33% News	0.34	0.734
HML-5% News	0.16	0.873
Proportional News	0.05	0.960

Interestingly with a 1-day holding period, there the short side of the portfolio always outperformed on a risk-adjusted and a returns basis. However, as seen in tables 18 a and b this outperformance is not significant.

5.3.2. Optimal Holding Period

The optimal holding period was established in part 2 of this paper, being five days when using both positive and negative sentiment as a signal to trade. The difference in reaction to different sentiment scores can be examined. The negative sentiment may now be considered as an overreaction to the previous performance of the shares, and thus a tradable opportunity exists if a long position is taken in these shares, where the trader will take advantage of a mean reversion effect. Making use of the same formula as above.

Table 19a: Examining for significant outperformance of negative sentiment driven trades relative to positive sentiment driven trades over optimal holding period on a risk-adjusted basis. Examining for a significantly different return of the long and short side of the portfolio, for the period 2020-01-02 to 2021-12-31.

	$\left(\frac{\bar{r}_-}{\sigma_-} - \frac{\bar{r}_+}{\sigma_+}\right)$	**Significant Outperformance of short side**	**Significant Outperformance of long side**
5-Day Holding Period	-0.1	False	False

Table 19b: Examining for significant outperformance of negative sentiment driven trades relative to positive sentiment driven trades over optimal holding period on a return basis. Examining for a significantly different return of the long and short side of the portfolio making use of a t-stat, for the period 2020-01-02 to 2021-12-31.

	T-stat	**p-value**
5-Day Holding Period	0.04	0.968

Using the optimal holding period, it is seen that the positive sentiment portfolio outperforms that of the negative sentiment portfolio on a risk-adjusted basis. The negative sentiment portfolio earns a higher return, however, this performance is not significant and thus it can be said that they performed in an equal manner.

5.3.3. Conclusion of Section 5.3. Bullish vs Bearish Sentiment

These results provide evidence against an asymmetric risk function as described by Kahneman & Tversky (1979), as there is an equal reaction seen to both positive and negative information shocks. But due to the low predictability of the Twitter sentiment and the likely mean-reversion nature of the results seen in the negative sentiment trade, this paper feels that these results are insufficient to provide any significant opposition to the findings of Kahneman & Tversky (1979).

6. Conclusion

This study has examined the relationship between investor sentiment obtained from Tweets. This sentiment was obtained by making use of Bloomberg's social activity tool. A three-part approach was used, the first explored the existence of a relationship between previous day Twitter sentiment and stock market returns. Multiple portfolios were constructed to identify this relationship, however, there was little to no evidence to support this. Some of the returns earned were significant at varying levels, however, the effect size was small, and the t-stats were inflated by the large number of days being considered. The small effect size would also likely result in insignificant and even negative results once transaction and short selling costs were reintroduced. In fact, part 1 generated evidence that stock market returns influence investor sentiment, which means that investor sentiment is a trailing indicator rather than a leading indicator as this paper initially theorised.

Part 2 made use of an event study to identify the optimal holding period, to maximise the returns generated by the HML-33% portfolio. This portfolio was used as it was the most successful from part 1. Optimal holding periods were established being a 5-day holding period, with the modification that both the positive and negative sentiment would be used as a buy signal. While this portfolio did outperform the benchmark, this outperformance was not significant on a risk-adjusted basis. When conducting the event study, the returns with the greatest significance were seen to occur before the event date, which provides further evidence that the sentiment seen in microblogs is a trailing indicator.

Part 2 also examined the URC, which is a test of market efficiency, which identifies how quickly information is priced in. Large underreactions were found for both positive and negative sentiment when examining a 10-day window, as done by Oh & Sheng (2011). This is a challenge to the semi-strong form of market efficiency (Fama, 1970), however, due to the low predictive ability of the sentiment, and the calculation of sentiment not being straightforward and thus the information not being obviously available, this paper cannot refute the semi-strong form market efficiency as outlined by Fama (1970).

Part 3 explores an asymmetric risk aversion of investors. It would be expected that negative sentiment would result in a stronger signal with greater predictive ability as individuals are risk-loving on the upside but risk-averse on the downside (Kahneman & Tversky, 1979). While the returns deriving from the trades made in conjunction with negative sentiment from part 1 always outperformed those earned from the trades made in conjunction with the positive sentiment on both an absolute and risk-adjusted basis, this outperformance was not significant and thus it cannot be stated with confidence that this is an effect that will always persist. And when considering the optimal holding period, it was found that the trades made in conjunction with positive sentiment outperformed on both an absolute and a risk-adjusted basis.

To conclude, this paper has found evidence that trading on microblog sentiment is not profitable at a daily level, Twitter sentiment-driven trades do not show significant outperformance of the benchmark on a risk-adjusted basis, a multi-day holding period does outperform a 1-day holding period but not significantly and there is no difference in the response of investors to positive or negative microblog sentiment. Thus microblog sentiment is not a leading indicator, but rather is a response to the previous returns of the market.

7. Limitations and Improvements for Future Studies

It should be noted that there are a number of limitations and possible improvements to this study. Making use of a sentiment model from Bloomberg reduces the flexibility of the study as a user must accept their categorisation of the text data. The use of an SVM, while shown to perform well in literature, is possibly inferior to the results that a well-trained BERT model may output. Having a model that one has developed themselves would also allow them to data-mine more extensively and uncover one of the many idiosyncrasies of these online communities that may be a buy or sell signal.

Bloomberg also only offers data on Twitter sentiment. Twitter is the most widely used microblog in the world, and thus the information gathered will be priced in faster reducing the profitable trading opportunities. Extending the study to other platforms such as StockTwits or Subreddits such as r/wallstreetbets or r/cryptocurrency may provide more valuable data.

Conduction a test for normality, would further increase the confidence that one may derive from the T-stats obtained. Further nonparametric test such as a Mann-Whitney U-test (1947) could be used so that no assumption about the underlying data is needed.

Another improvement that could be made would be to rank users and assign them specific weightings on their perceived credibility. The model used in this study assumes that all individuals or Twitter accounts carry the same influential ability which is factually false. A Tweet from an individual such as Elon Musk would have a far larger effect on the price of an asset than a Tweet from an account with a single follower as seen in Rui, Liu, Whinston (2013). Another extension may to be limiting one user one vote, as individuals may Tweet about the same thing multiple times a day which would overweight their voice relative to the crowd.

It should also be noted that observing smaller time frames than this study used may also provide more significant useful results as found by Cui, Lam & Verma. (2016) and Deng et al (2018).

Cryptocurrencies have not been considered by this study due to the lack of availability of sentiment data on these assets. However, Cryptocurrencies are high volatile assets that are almost entirely driven by market sentiment, therefore I believe that a study on Cryptocurrency would provide many interesting results.

8. Reference List

Aggarwal, R., Gopal, R., Sankaranarayanan, R. and Singh, P.V., 2012. Blog, blogger, and the firm: Can negative employee posts lead to positive outcomes?. *Information Systems Research*, *23*(2): 306-322.

Alam, F. and Pachauri, S., 2017. Detection using weka. Advances in computational sciences and technology, 10(6), pp.1731-1743.

Antweiler, W. and Frank, M., 2002. Internet stock message boards and stock returns. *University of British Columbia Working Paper*: 1-7.

Antweiler, W. and Frank, M.Z., 2004a. Does talk matter? Evidence from a broad cross section of stocks. *University of British Columbia Working Paper*.

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

Aylien, N.B., 2016. *Support Vector Machines: A Simple Explanation*. Available: https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

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

Baker, M. and Wurgler, J., 2007. Investor sentiment in the stock market. *Journal of economic perspectives*, *21*(2): 129-152.

Bloomberg Market Specialists, 2017. *How you can get an edge by trading on news sentiment data.* Available: https://www.bloomberg.com/professional/blog/can-get-edge-trading-news-sentiment-data/

Bloomberg, 2022. *Bloomberg Terminal.* Available: https://bba.bloomberg.net/

Cohen, L. and Frazzini, A., 2008. Economic links and predictable returns. *The Journal of Finance*, *63*(4): 1977-2011.

Corrado, C.J., 2011. Event studies: A methodology review. Accounting & Finance, 51(1), pp.207-234.

Cui, X., Lam, D. and Verma, A., 2016. Embedded value in Bloomberg news and social sentiment data. *Bloomberg LP*.

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

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., 1990b. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance*, *45*(2): 379-395.

Deng, L. and Poole, M.S., 2010. Affect in web interfaces: A study of the impacts of web page visual complexity and order. *Mis Quarterly*, pp.711-730.

Deng, S., Huang, Z.J., Sinha, A.P. and Zhao, H., 2018. The interaction between microblog sentiment and stock return: An empirical examination. *MIS quarterly*, *42*(3): 895-918.

Dolley, J.C., 1933. Characteristics and procedure of common stock split-ups. *Harvard business review*, *11*(3), pp.316-326.

Dybvig, P.H. and Ross, S.A., 1985. The analytics of performance measurement using a security market line. *The Journal of finance*, *40*(2): 401-416.

Dyer, J.N., MacKinnon, R.J. and Elder, K.L., 2014. A Survey and Discussion of Competing Mean-Variance Statistics in Portfolio Analysis. *Journal of Financial Education*, pp.22-55.

Fama, E.F., 1970. Session topic: stock market price behavior. *The Journal of Finance*, *25*(2): 383-417.

Fama, E.F., Fisher, L., Jensen, M. and Roll, R., 1969. *The adjustment of stock prices to new information*. International economic review, 10(1).

Graham, B., Dodd, D.L.F. and Cottle, S., 1934. *Security analysis* (Vol. 452). New York: McGraw-Hill.

Grinold, R. and Kahn, R., 1999. Active portfolio management: A Quantitative Approach for Producing Superior Returns and Controlling Risk.

Gupta, S. 2018. *Sentiment Analysis: Concept, Analysis and Applications*. Available: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

Hayes, A. 2020. *Event Study*. Available: https://www.investopedia.com/terms/e/eventstudy.asp

Hong, H. and Stein, J.C., 1999. *A unified theory of underreaction, momentum trading, and overreaction in asset markets*. The Journal of finance, 54(6): 2143-2184.

IBM, 2020. Natural Language Processing (NLP). Available: https://www.ibm.com/cloud/learn/natural-language-processing

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, *23*(2): 389-416.

Kahneman, D, and Tversky, A, 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, no. 2: 263–91.

Keane, S.M., 1983. *Stock market efficiency: theory, evidence and implications*. Philip Allan.

Keynes, J.M., 1937. *The general theory of employment*. The quarterly journal of economics, 51(2), pp.209-223.

Lee, K.C., 2021. *Sentiment Analysis — Comparing 3 Common Approaches: Naive Bayes, LSTM, and VADER.* Available: https://towardsdatascience.com/sentiment-analysis-comparing-3-common-approaches-naive-bayes-lstm-and-vader-ab561f834f89

Lo, A.W., Repin, D.V. and Steenbarger, B.N., 2005. Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, *95*(2), pp.352-359.

Luo, X., Zhang, J. and Duan, W., 2013. Social media and firm equity value. *Information Systems Research*, *24*(1): 146-163.

Lütkepohl, H., 2007. General-to-specific or specific-to-general modelling? An opinion on current econometric terminology. *Journal of Econometrics*, *136*(1): 319-324.

MacKinlay, A.C., 1997. Event studies in economics and finance. *Journal of economic literature*, *35*(1), pp.13-39.

Mann, H.B. and Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, pp.50-60.

Mayew, W.J. and Venkatachalam, M., 2012. The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, *67*(1), pp.1-43.

MonkeyLearn, 2021. *Sentiment Analysis: A Definitive Guide*. Available: https://monkeylearn.com/sentiment-analysis/

Morck, R., Shleifer, A., Vishny, R.W., Shapiro, M. and Poterba, J.M., 1990. The stock market and investment: is the market a sideshow?. *Brookings papers on economic Activity*, *1990*(2): 157-215.

Oh, C. and Sheng, O., 2011. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement.

Otoo, M.W., 1999. Consumer sentiment and the stock market. *Available at SSRN 205028*.

Patell, J.M., 1976. Corporate forecasts of earnings per share and stock price behavior: Empirical test. *Journal of accounting research*, pp.246-276.

Roseman, I.J. and Smith, C.A., 2001 Appraisal Theory. *Appraisal Processes in Emotion: Theory, Methods, research:* 3-19

Rui, H., Liu, Y. and Whinston, A., 2013. Whose and what chatter matters? The effect of tweets on movie sales. *Decision support systems*, *55*(4): 863-870.

Sharpe, W.F., 1966. Mutual fund performance. *The Journal of business*, *39*(1): 119-138.

Smith, A., 1776. *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*: 48

Statista, 2021. *Leading countries based on number of Twitter users as of July 2021*. Available: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Summers, L.H., 1986. Does the stock market rationally reflect fundamental values?. *The Journal of Finance*, *41*(3): 591-601.

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.

Towardsdatascience.com, 2018. *Support Vector Machines (SVM) — An Overview*. Available: https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989

Tumarkin, R. and Whitelaw, R.F., 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal*, *57*(3): 41-51.

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., 2011. Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social science computer review*, *29*(4): 402-418.

Ullrich, C., Borau, K., Luo, H., Tan, X., Shen, L. and Shen, R., 2008, April. Why web 2.0 is good for learning and for research: principles and prototypes. In *Proceedings of the 17th international conference on World Wide Web*: 705-714

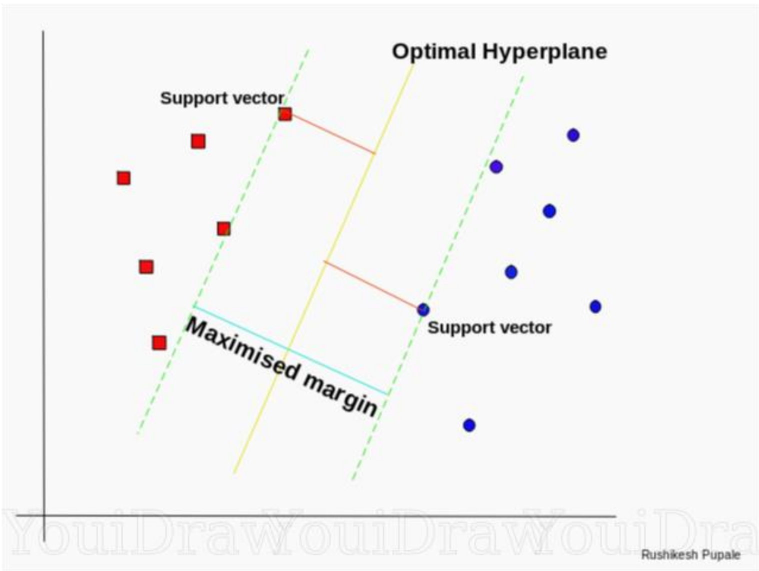
Xing, F., Malandri, L., Zhang, Y. and Cambria, E., 2020, December. Financial sentiment analysis: an investigation into common mistakes and silver bullets. In *Proceedings of the 28th International Conference on Computational Linguistics*: 978-987

Zuccarelli, 2020. *Performance Metrics in Machine Learning — Part 1: Classification.* Available: https://towardsdatascience.com/performance-metrics-in-machine-learning-part-1-classification-6c6b8d8a8c92

9. Appendix

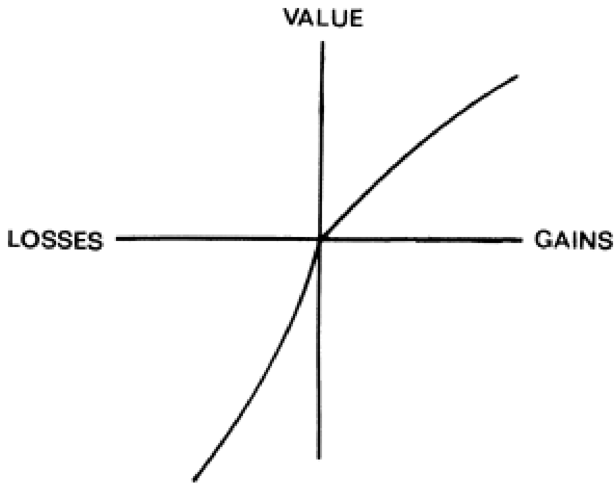
Appendix 1. Further Explanation of Support Vector Machine (SVM) Models

Bloomberg does not disclose the manner in which their SVM conducts sentiment analysis, thus this paper has included an overview of how SVM models work for the interest of the reader. SVM is a machine learning algorithm that is often employed for both classification and regression purposes. An SVM model will distinguish between categories, in the case of investor sentiment there are two categories, positive and negative, and plot all data points on a set of axes. The SVM will then attempt to plot a hyperplane between each category, a hyperplane is just a line that linearly separates the different categories of data, when dealing with two categories the hyperplane is just a line, but as more categories are introduced it will become a plane. The hyperplane is plotted in such a way to maximise the margin, where the margin is the distance between data points and the hyperplane, between each of the two categories considered. (Aylien, 2016). Which will then result in the best classification of positive or negative sentiment, see the visual example below. Bloomberg then conducts a proprietary method to attach a sentiment score between -1 and 1 to the data points that have been categorised as positive or negative.



Source: Towardsdatascience.com

Appendix 2. Kahneman & Tversky (1979) Asymmetric Risk model



Source: Kahneman & Tversky (1979)

Appendix 3. Value of Twitter portfolios over time.