Twitter and the Cross Section of Stock Returns

Programming for Finance; Assignment Report

Alec Bernasconi
Jan Gobeli
Riccardo Cosenza
Ivana Pallister
2020

Abstract

"The aim of this paper is to analyze stock market performance and its potential correlation with Twitter data. The performance of the stocks were compared to all tweets over the same time period which referenced the stock. The analysis finds a potential correlation between tweet frequency and market perfomance."

Università della Svizzera italiana

TABLE OF CONTENTS

1. Introduction

The purpose of this report is to give insight into the process of our Twitter analysis. The analysis has been conducted using the R programming language, any R files will be enclosed in a separate appendix.

The goal of this assignment was to:

1. Obtain twitter data for 80% companies in the S&P500.
2. Produce daily time series of the number of tweets regarding each company.
3. Produce time series for all tweets and analysize for any cyclicality.
4. Compare the “market share” of tweets to the index weight, analyze whether companies are “under” or “overtweeted”.
5. Assess whether the tweets are “bullish” or “bearish”, incorporate this analysis into the subsequent strategy.

2. Research Methodology

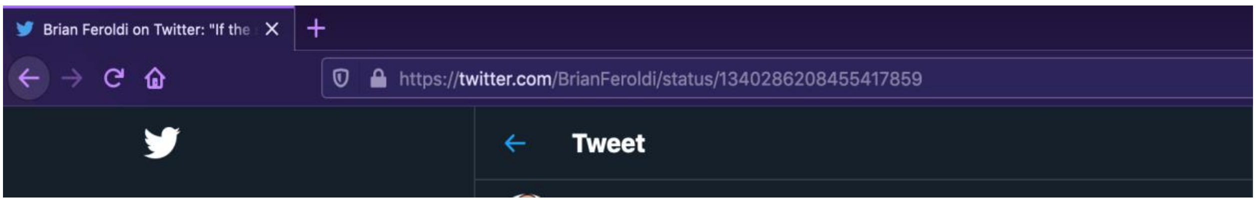
2.1. Data Collection

To complete this project there is a need for a large data set of tweets. We choose to analyze 80% of the firms in the S&P500 based on market capitalization. We referenced these companies based on their cashtags, e.g. $AAPL.

First, the amount of tweets that we could obtain officially from Twitter, as the API limited us to seven days of tweets. As seven days would not be diverse enough for a reliable data set we chose an alternative way to collect the data. The aim was to recover all the tweets for the designated companies for the past year, 2020.

We know that every tweet exists on a webpage and thus has URL. At the end of every Twitter URL there is an ID number for that specific tweet, see figure 1 (Feroldi, 2020). This means that when we perform a search query for all the cashtags we can save the ID numbers at the end of the URL so we can find them later. This process is called webscraping.

Figure 1: Example of tweet $AAPL



Reference: (Feroldi, 2020)

We then call the API with the string of ID's. Using this process we can collect twitter data as far back as 2007. However, in 2020 we collected approximately 20 million tweets which we concluded was sufficient for our analysis. Besides the content of the tweet, we were also able to collect other characteristics, such as, screen name, if the user is "verified", followers, follows, retweets and favorites.. These factors are indicated in figure 2, on the following page.

Figure 2: Recorded Parameters

Nr	Parameter	Explained
1	Screen name	The name of the user that made the tweet.
2	Verification	Whether the user is "verified" by Twitter.
3	Followers	How many followers the user has.
4	Following	How many acoounts the user himself follows.
5	Retweet	If the tweet itself is a retweet.
6	Amount Retweeted	How many times the tweet was retweeted.
7	Amount Favorited	How many times the tweet was favorited by other users.

Source: Own Representation

2.2. Data Cleaning

The raw data set obtained as described in section 2.1 was not reliable for our analysis. The Twitter API did not distinguish between a cashtag, e.g. $AAP and a mention e.g. @AAP. This presented a problem as one indicates a companies stock data while the mention could be about an entirely different firm. We thus filtered out all of these mentions. For the company Ebay, this lead to 22,670 relevant tweets out of the original 4,801,001.

With the cleaned data we ordered our date based on time, thus creating a time series. We did this both daily and quarterly. This allows us to analyze the time series for any cyclicality.

2.3. Sentiment Analysis

Our data set essentially consists of messages between users about the companies relevant to us. Thus far in our analysis we have had access to the count and statistics about the data set. However, we have not looked at what the messages actually mean.

We can establish the sentiment of the tweet by giving each a sentiment score, the tweets can be sorted into three groups; positive, negative and neutral. These sentiment scores are given by SentimentR (Rinker, 2019).

Every tweet is broken into sentences, punctuation is removed from each sentence and the tweet is broken down into an ordered pack of words. The package considers punctuation such as commas and questionmarks as separate words. It then cross-references these word-packs with its existing libraries and sentiment rules. The result is an indication of the tweets sentiment (Rinker, 2019).

2.4. Portfolios

The data and time series can be used to create a variety of portfolios which can be compared based on performance within eachother and the S&P500 performance.

The first portfolio is based on the amount of tweets about each company. The weights of the companies in the portfolio is based on the amount of tweets abou

Based on the amount of tweets per company we are also able to create an index weighted by the amount of tweets per company. Thus, this created a portfolio based on “market share”, which is reindexed quarterly.

The subsquent indexies consists of two equally weighted separate portfolios. The aim was to create a portfolio that is based on the “over- and undertweeting” of companies. Every quarter we assess the market share of tweets and compare the amount of tweets to the original market share in the S&P500. If the weights of the twitter index are higher than the weights of the S&P500 weights the company is deamed overweighted. Whereas if the company’s twitter index is lower than its S&P500 index it is undertweeted. Thus, from this analysis we are left with two pools of companies, one pool undertweeted and one pool overtweeted. We then index these two pools by equal weights.

The third portfolio tests whether an overtweeted company performs better compared to an undertweeted company. In portfolio three, there is a long position on overtweeted companies and a short position on undertweeted companies. The weights in each pool are based on the twitter index. The sum of the two pools does not go to 1, the remaining portion of the company is invested into a risk free asset.

Finally, we created a similar portfolio to the over- and undertweeted, however this time based on the sentiment of the tweets. Every tweet receives a sentiment score as described in section 2.3. This score is averaged over every quarter and the positive and negative averages are devided into two separate pools. Subsequently, each pool is weighted equally to form one overall portfolio.

Figure 3: Overview Portfolio

Nr.	**Portfolio**	**Summary**
1	S&P500	Base case, S&P500 index based on market share.
2	Twitter Index	Companies indexed by their share of overall tweets compared to the sum.
3	Over/undertweeted index	Companies indexed equally in two pools, overtweeted and undertweeted. Over- and undertweeted is defined by the share of overall tweets in excess to their S&P500 marketshare.
4	Short/long Index	Companies indexed in two pools based on their twitter market share, overtweeted and undertweeted. The portfolio takes a short position on undertweeted and long position on overtweeted.
5	Sentiment Index	Companies indexed equally in two pools, negative and positive overall tweet sentiment.

Source: Own Representation

3. Results

3.1. Base case

The portfolios described in chapter two, are compared to S&P500 data. Since our analysis is based on only 80% of the S&P500, our base case is not identical. We thus created a custom S&P500 index that accurately represents the analyzed companies.

The performance of this altered index is summarized in figure 4 below. The market experienced a return of 9.93%, an annual volatility of 35.83% and a sharpe ratio of 0.2664.

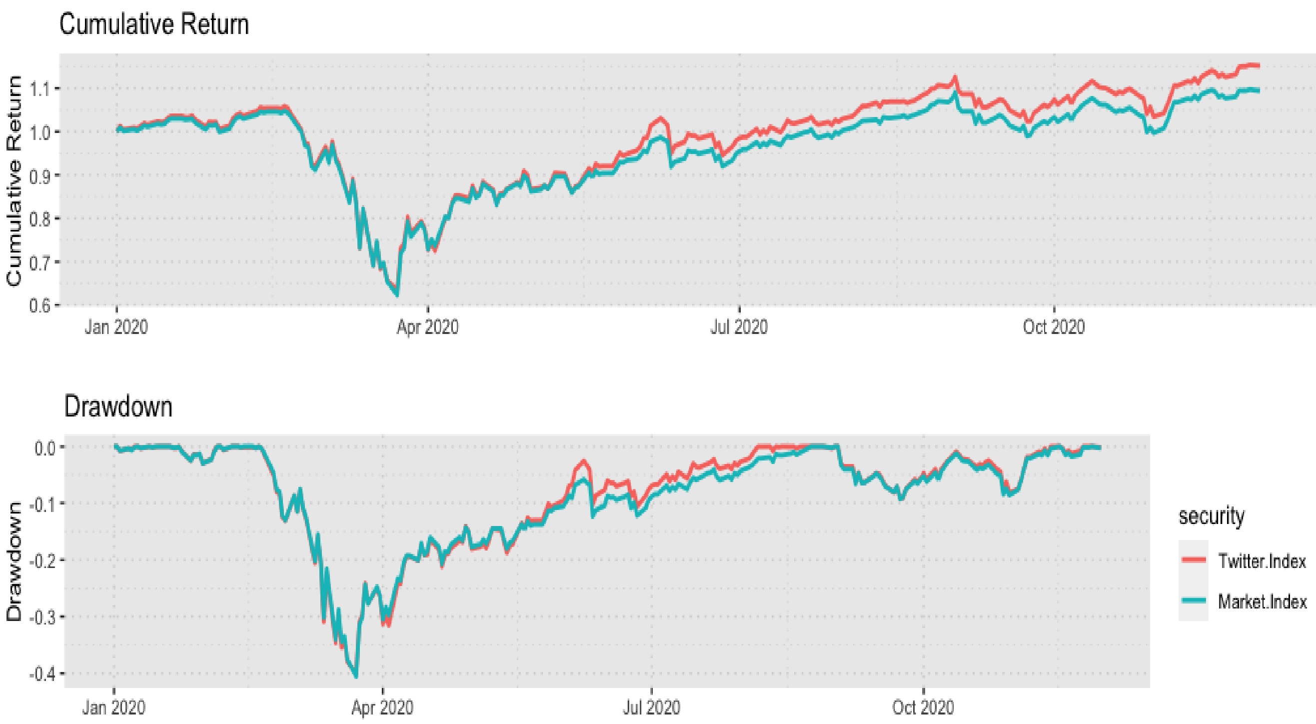
Figure 4: Altered S&P500 index

Source: Own Representation

3.2. Twitter Index

The results of the twitter index suggest that there might be a positive advantage to using Twitter data. The Twitter index beats the market by 6.07%, with an annualized return of 16.00%. However, the volatility is higher as well at 37.05%. The two indices are compared in figure 5 below. Overall, the Twitter index beats the market sharpe ratio by twofold at 0.4215.

Figure 5: Twitter Index compared to S&P500



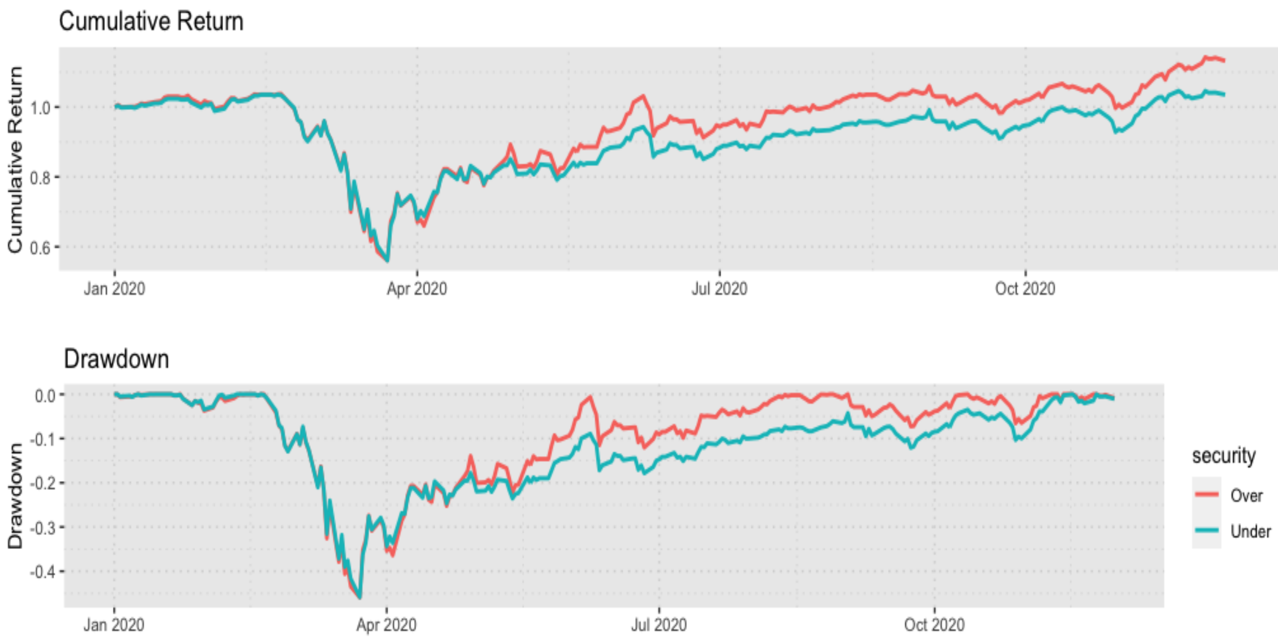
Source: Own Representation

3.3. Over/undertweeted Index

The results in section two are supported by the findings of the over- and undertweeted indices. The overtweeted index has an annualized return of 13.93%, a volatility of 39.36% and a sharpe of 0.3443. Again, there appears to be a positive correlation between Twitter performance and returns.

This effect is emplified by the performance of the undertweeted index. With a return of 3.73% this index was the worst performing portfolio during the analysis. Its annualized volatility is 36.69% and has a sharpe of merely 0.0912. The two indices are summarized in figure 6 on the following page.

Figure 6: Summary Over- and undertweeted Indices

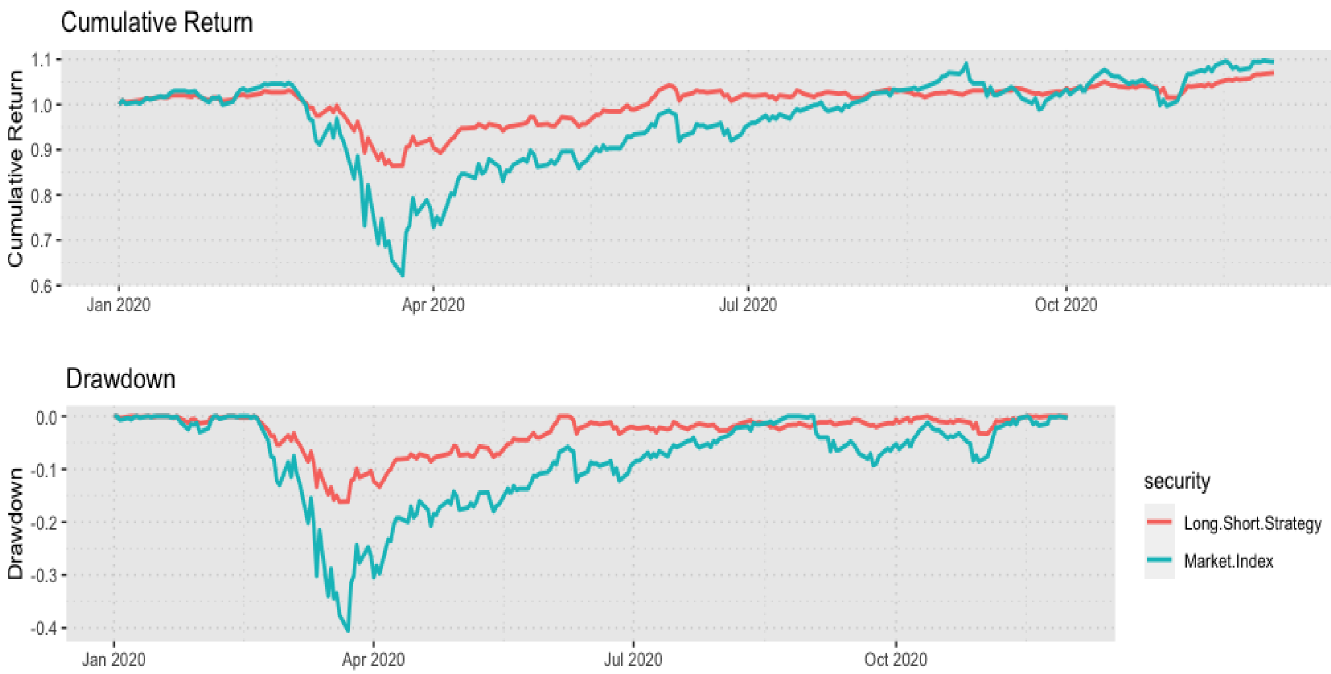


Source: Own Representation

3.4. Short/long Index

The long/short portfolio further explores whether there is a positive correlation between tweets and market performance. In summary, the portfolio takes a long position on the overtweeted stocks and a short position on the undertweeted stocks analyzed in section 3.3. The portfolio has a return of 7.38%, a lower return compared to the base rate. However, it has a significantly reduced volatility at 14.88%. This reduction results in the highest sharpe ratio of the set, namely, 0.4701. The strategy also shows a much lower drawdown during a market decline as shown in figure 7.

Figure 7: Summary Short/long Index



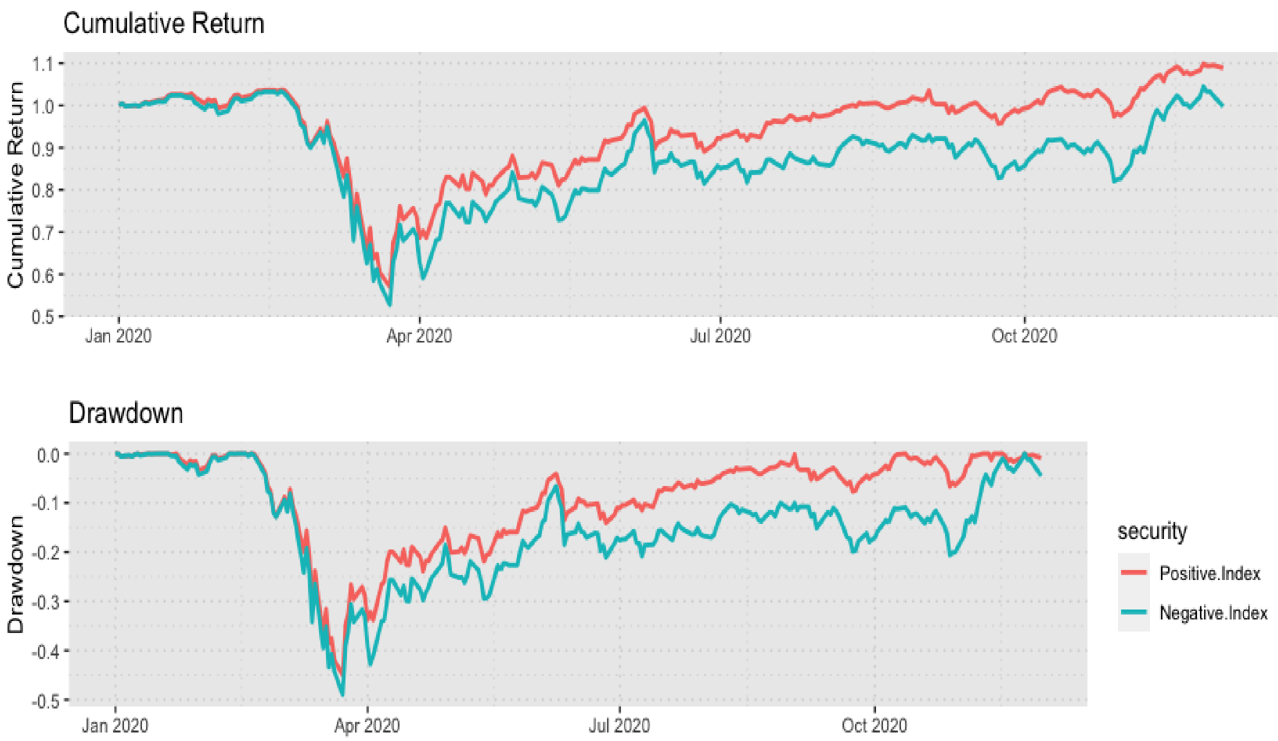
Source: Own Representation

3.5. Sentiment Index

The sentiment index is less reliable compared to the above mentioned portfolios. The positive sentiment pool has a quarterly return of 9.37% which is lower than the market level. The volatility is 37.49% and it has a sharpe of 0.2396. For sentiment, there is an argument that quarterly weights are too wide spread. If we test for positive sentiment on a weekly basis the portfolio slightly increases to a return of 10.71% , volatility of 39.69% and a 0.2602 sharpe ratio.

Furthermore, the negative sentiment pool is the lowest returning portfolio. It has an expected return of -0.23%, standard deviation of 43.00% and a sharpe of -0.0143. Altered to weekly weights it returns of 0.15%, with 40.93% volatility and a sharpe of -0.0056.

Figure 8: Summary Sentiment Analysis



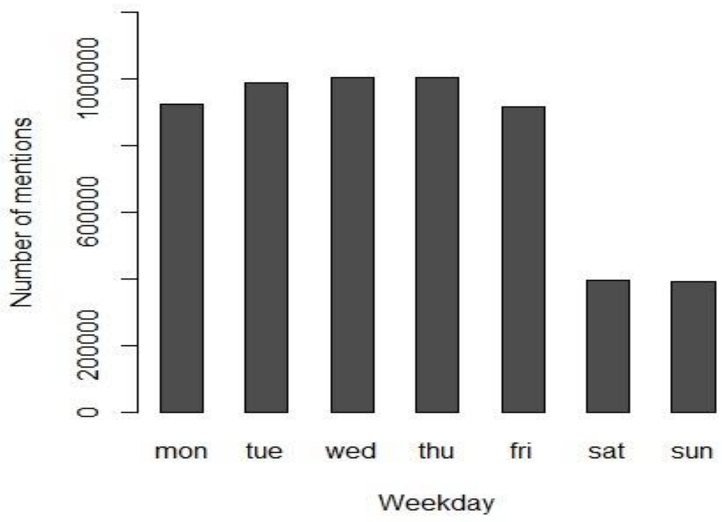
Source: Own Representation

These two pools are not the best performing in the analysis. There is an obvious limitation at play, that the sentiment analysis for tweets is complicated and our method may not have been sophisticated enough to capture the full effect. However, it could also be possible that there are more relevant indicators available.

3.6. Cyclicality

When analyzing the frequency of tweets we find that there are almost no tweets outside of working hours. Thus, there is very little impact of weekend tweets on Monday's performance. Distribution of tweets per weekday are shown in figure 9 below.

Figure 9: Distribution of tweets per weekday



Source: Own Representation

This could be due to a large amount of tweets that only mention a few cashtags with little added value. These tweets may be used to drive up the traffic for certain cashtags during trading hours. As well as bet on algorithms that trade on the amount of cashtags.

3.7. Conclusion

In conclusion, there may be evidence to suggest that Twitter data can serve as an indicator for stock performance. Based on the Twitter Index we may speculate that there could be a correlation between the amount of attention paid to a company on Twitter and its stock performance. In figure 10 on the following page, the results are summarized.

Figure 10: Summary Results

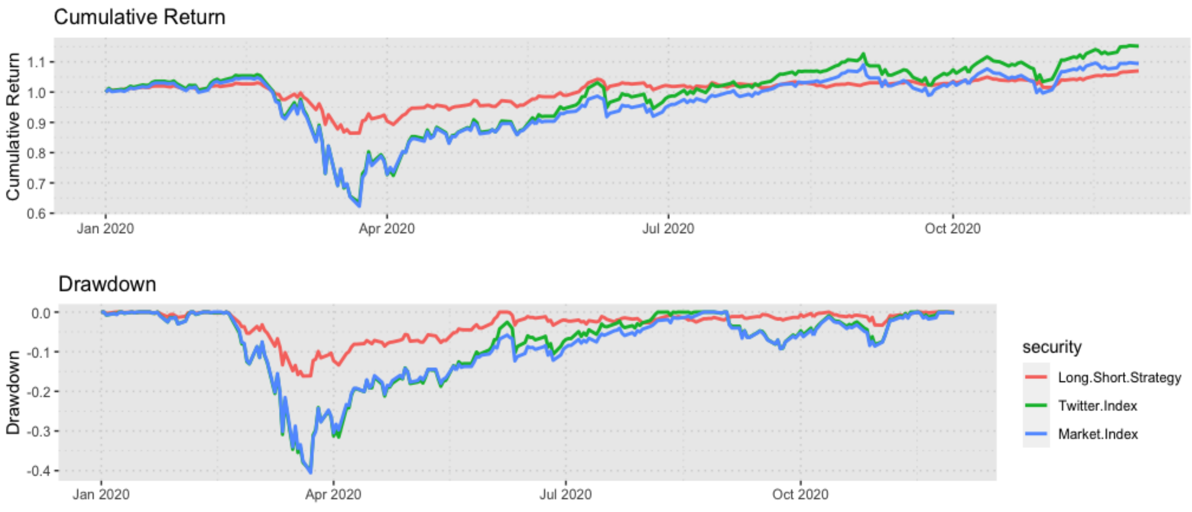
Index	Annualized Return	Annualized Volatility	Annualized Sharpe Ratio
S&P500	9.93%	35.83%	0.2664
Twitter	16.00%	37.05%	0.4215
Over	13.93%	39.36%	0.3443
Under	3.73%	36.69%	0.0912
Long/Short	7.38%	14.88%	0.4701
Positive	9.37%	37.49%	0.2396
Negative	-0.23%	43.00%	-0.0143

Source: Own Representation

Sentiment analysis was the worst performing indicator, this is most likely due to its complexity. Despite its performance as a trading strategy, it still supports the hypothesis, as negative sentiment performed significantly worse compared to any other indicator.

Figure 11 below, shows a graphical representation of the top performing portfolios.

Figure 11: Portfolio Summary



Source: Own Representation

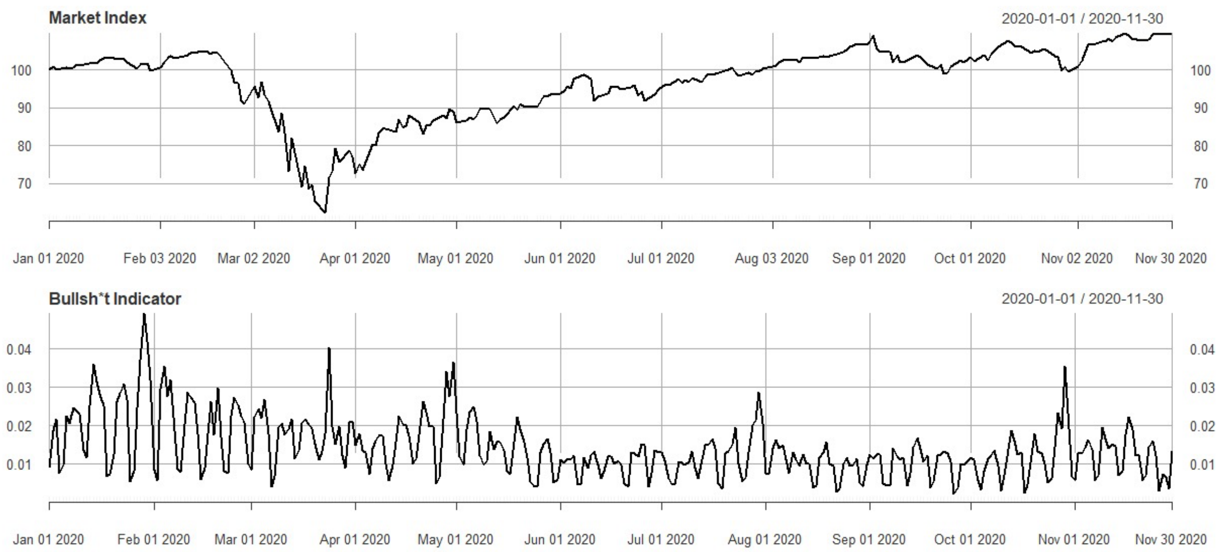
4. Discussion

4.1. Bullsh*t Indicator

Based on the analysis there is evidence to suggest that there may be a positive relationship between Twitter and stock returns. However, based on the dataset the team was fairly skeptical and thus conducted further statistical tests to determine if Twitter weights indeed have the power to generate significant excess returns.

The data set provided contains more information than just the tweet itself and the date. This additional data can also be analyzed to gain insight into what might impact stock prices. For example, a "Bullshit"-indicator was created to determine the number of tweets made by people not verified by Twitter. In other words, to decide how many tweets per day are posted by users who are important according to Twitter. This idea can be extended with indicators such as the number of followers a person has, the number of likes or retweets a person receives.

Figure 11: Bullsh*t Indicator



Source: Own Representation

4.2. Speculation

Based on the analysis there may be cause to further analyze Twitter data. Based on the found other interesting points such as a potential correlation between a decline in “important” accounts and a general decline in the market value of the stock.

Furthermore, the research above is a mere starting point for statistical analysis of Twitter data. There is much more to gain in mentions of companies by tag (@apple) instead of just cashtag ($AAPL). As well as comparing more data, from more locations to more companies.

5. Limitations

5.1. Geographic and Cyclic Limitations

The analysis was conducted based on data between January 1, 2020 and November 30, 2020. In addition, the data set only consideres companies based in the United States. This limites the results of the analysis, as 2020 in the U.S. is not representative of multi-year data, nor can it be relied upon to accurately predict in other geographical locations.

5.2. Stock data

Stock data is based on the S&P500 on December 12, 2020. Any alterations to the index succeeding this date are not taken into consideration.

5.3. Sentiment Limitations

The conducted sentiment analysis has its limitations as the libraries were not adjusted for twitter sentiment analysis. Finding the sentiment of a tweet is complex and a model will not accurately predict it every time.

6. Bibliography

Dinco, J. (2020, September 29). Downloading historical tweets using tweet_ids via snscrape and tweepy. Melbourne, Australia.

Feroldi, B. (2020, December 19). *Tweet ID 1340286208455417859*. Retrieved from Twitter: https://twitter.com/BrianFeroldi/status/1340286208455417859

Refinitiv. (2020, December 13). Datastream. London, United Kingdom.

Rinker, T. (2019). *Calculate Text Polarity Sentiment*. R Documentation.

Twitter. (2020, December1 13). Twitter API v2: Early Access. San Francisco , California, United States.