ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

MSc Economics & Business

Specialization Financial Economics

The Effect of Market Sentiment on Cryptocurrency Prices.

Author: R.B. Hurjui

Student number: 466956rh

Thesis supervisor: Dr. Jan Lemmen

Second reader: Dr. Ronald Huisman

Date: 20th April 2022

Preface and Acknowledgements

I would like to thank my supervisor, Dr. Jan Lemmen, for his assistance and feedback provided throughout the thesis writing process. The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

I. Executive Summary

This paper studies the effects of social media sentiment on the daily returns of cryptocurrencies, and the effect of social media attention on the daily volatility of cryptocurrencies.

In this century social media usage has grown enormously, becoming one of the main sources from which people obtain their news and express their opinions, therefore what people are exposed to online can greatly influence their perceptions and decision making. When it comes to cryptocurrencies, it is often argued that one of the main drivers of price changes is speculation, which is most often spread through social media platforms. Consequently, it is interesting to study whether social media sentiment can influence by itself the prices of some of the biggest cryptocurrencies, as that could change our perception about the world of crypto. It is also worth looking at whether increased market attention measured using social media attention can be a driver of volatility.

Daily social market sentiment is obtained for the year 2021 from tweets with the specific cryptocurrency hashtags and Reddit posts from the respective subreddits of the cryptocurrencies, with Twitter and Reddit sentiment acting as proxies for social market sentiment. The cryptocurrencies which will be studied are Bitcoin, Ethereum, and Cardano ADA. The total number of daily tweets with the specific cryptocurrency hashtags and the daily number of posts on each cryptocurrency subreddit will proxy for social media attention. Daily price values along with the lowest and highest daily price are obtained for each cryptocurrency, along with the daily values of the control variables will be used in this study.

The results indicate that lagged Twitter sentiment is a significant driver of the daily return of Bitcoin only, while lagged Reddit sentiment is a significant driver of the daily return of all three cryptocurrencies studied in this paper. Lagged Twitter attention is found to have a significant effect on the daily volatility of Bitcoin, while lagged Reddit attention is found to have a significant effect on the daily volatility of Bitcoin and Ethereum.

II. Table of Contents

Contents

1. Introduction

By October 2021, the global cryptocurrency market cap had reached 2.63 trillion US Dollars, a 50-fold increase when compared with the May 2017 market cap of 50 billion US Dollars. Although it is still currently far from reaching the same market cap of the NYSE for example (26.2 trillion US Dollars), one factor that differentiates cryptocurrencies from a stock within a company is how their value is determined. A common stock, or a common share of company is a unit of equity ownership in a company, which is backed by the company itself and has its price determined by the performance of the company and the markets' belief for the prospects of the company. For cryptocurrency, the value is determined partially through the cost of obtaining the cryptocurrency (mining in the case of Bitcoin), the regulatory developments regarding the use of cryptocurrencies, for example when China announced the ban of financial institutions from providing crypto transaction related services on May 18th, 2021, and the price of Bitcoin dropped 13% in one day. Another factor that may influence cryptocurrency prices is market sentiment and speculation. Kristoufek (2015) found that despite the belief that Bitcoin is a purely speculative asset, over the long term standard fundamental factors such as usage in trade, money supply and price level also determine the long-term price of Bitcoin. The stock market price fluctuations can also be determined in part by market sentiment, which can sometimes not be rational as shown by multiple studies which concluded that the efficient market hypothesis does not always hold (Malkiel, 2005). Therefore, it is interesting to look at how market sentiment influences the price of cryptocurrencies, as it is often argued that they have no intrinsic value in the way that company shares do. Whether the cryptocurrency market is efficient also came into question in a paper by Caporale and Plastun (2019), in their study about day of the week effects in the cryptocurrency market. The paper found that the Monday effect was present for Bitcoin, and that a trading strategy utilising Bitcoin's higher than usual returns on Monday would have been profitable throughout the sample period. Although this research does not bring assurance that the cryptocurrency market is inefficient, this is an indicator of its potential inefficiency, therefore the market being influenced solely by speculation is not out of the question.

In recent times there has been an increase in the prevalence of social media websites such as Twitter and Reddit, where people can share their opinions on any topic including cryptocurrencies, and this information is spread very quickly among other users. As it is often stated that cryptocurrency price fluctuations are mostly speculation based, it could be argued that market sentiment has led to a very high volatility amongst cryptocurrencies, with sharp increases in price and crashes becoming regular. For this reason, it is very important to study how market sentiment, which will be measured using Twitter and Reddit, directly influences the price of cryptocurrencies. One recent example where social media attention and sentiment have come in the spotlight is in early 2021, when members of the subreddit r/wallstreetbets riled up to short squeeze the stocks of GME and AMC. It has been shown that a strong relationship between the number of comments posted per hour and GME prices existed, while tones of sadness, anger, and surprise tones in posts were shown to exhibit a significant impact on the 1-minute returns of GME stock (Long et al., 2021). Social media overreaction and mass agreement such as in the case of GME, and AMC stocks in January-February 2021 could lead to more bubbles and crashes of cryptocurrency prices, should a clear link between the two be demonstrated. In cryptocurrencies, trading strategies against overreactions where also shown to not be profitable, while trading strategies following the direction of the overreaction appeared to be profitable (Caporale and Plastun, 2019). Another item that is worth studying is the effect of investor attention on the volatility of the cryptocurrency markets, as according to the attention bias investors are more likely to purchase stocks, or in this case cryptocurrencies, that catch their attention. Therefore, a cryptocurrency being in the spotlight should result in it being traded more often than its usual amount, resulting in increased volatility.

It can be observed that the future of Bitcoin, the largest cryptocurrency, is quite a controversial topic. On the one hand, Bitcoin maximalists as they are called believe that Bitcoin is the only digital asset that will be needed in the future, and due to its de-centralised nature, inflation hedging abilities, and limited supply, the price will rise significantly even from its current price of approximately 37 000 euros (April 2022). On the other hand, other investors believe that due to its lack of intrinsic value that can be found in other assets, combined with other factors such as cryptocurrencies not being accepted in many markets (China’s ban on crypto mining and trading), Bitcoin and other cryptocurrencies are in a bubble that will eventually burst (Cheun et al., 2015). Studying the intrinsic value of cryptocurrencies, García-Monleón et al. (2021) found for example that coins issued in ICOs can find value similar to stocks in IPOs, through the exchange rate that is established in its issuance agreement along with its conversion rights for certain market goods and services. They also state that cryptocurrencies whose only purpose is the movement of currency through the nodes of the network can be valued according to the utility this generates. If a cryptocurrency can also transfer other kinds of information, namely multiple layer currencies, these will generate additional utilities. Therefore, they will receive the additional value according to the additional utilities they generate, according to García-Monleón et al. (2021).

Varma, Kurisinkel and Radhakrishnan (2017) have pointed out that the prominence of social media content has become as important as any formal document, due to the instantaneity and relevance of the discussions that often appear on or due to social media. While there have been many studies researching the effects of investor attention and market sentiment on stock prices, such research on cryptocurrencies is currently quite limited due to the relative novelty of cryptocurrencies. Cryptocurrencies have also been characterised as being very volatile, as shown by Chan et al. (2017) who have found that multiple cryptocurrencies, of which also Bitcoin and Ethereum, show a very high amount of volatility relative to their inter-daily prices. The extremely high volatility and lack of fundamentals relative to stocks for cryptocurrencies means that the prices can be easily influenced by market sentiment. A good example of that would be when in January 2021 Elon Musk posted a series of tweets in favour of the cryptocurrency Dogecoin, such as "Dogecoin is the people's crypto" or just simply ‘’Doge’’. The price of Dogecoin shot up by 50% in the days following these tweets. Besides Twitter, one of the major social media sites that can also capture market sentiment quite well is that of Reddit. Reddit works through subreddits, which are user-created boards that focus on specific topics. The subreddits that would be used in this study would be those related to Cryptocurrencies, such as r/Bitcoin (3.6 million users), r/Ethereum (1.2 million users), and r/Cardano (677 thousand users). Subreddits can oftentimes also function as ‘’echo-chambers’’ where the most popular opinions are repeated and rewarded by the users through upvotes, awards and agreement, while opinions which are unpopular can get disregarded and hidden away through being massively downvoted, sometimes leading the original poster to delete the original post to no longer receive negative replies and downvotes, or even unpleasant private messages. For example, when a post receives more downvotes it will not show up on the ‘’top’’ page of the subreddit unlike massively upvoted posts, and comments that similarly receive mostly downvotes will linger towards the bottom of the discussion, oftentimes being hidden or removed by moderators who work voluntarily on moderating subreddits. This can further exacerbate investor sentiment in one direction on the subreddits, therefore making them good places to perform an analysis on the effects of investor sentiment on short-term prices of cryptocurrencies.

Given the evidence that market sentiment may have a strong effect on the daily returns of cryptocurrencies, the first research question of this paper is:

‘’*Does market sentiment measured through Twitter and Reddit sentiment influence the daily returns of Crypto?*’’.

This research question will expand on previous studies of the relation between Twitter sentiment and the returns of crypto by Abraham et al. (2018), Kraaijeveld and de Smedt (2020), by also adding Reddit sentiment as a proxy for overall market sentiment.

Therefore, the first and second hypotheses of this paper are:

H1: Positive investor sentiment on Twitter has a positive effect on the daily returns of cryptocurrencies.

H2: Positive investor sentiment on Reddit has a positive effect on the daily returns of cryptocurrencies.

Another element separate from investor sentiment is that of investor attention, where cryptocurrencies that are receiving more attention than usual should experience a higher level of volatility, as shown by Bank, Larch, and Peter (2011) regarding the stock market. Bijl, Kringhaug, Molnár, and Sandvik (2016) have found that for the stock market, higher Google search volumes predict high returns for the first one to two weeks, with subsequent price reversal. Barber and Odean (2008) have found similar results where stocks that received higher attention through being in the news have higher volatility, and positive abnormal returns as well. For cryptocurrencies Zhang and Wang (2020) find that media attention measured through Google searches can predict cryptocurrency returns and volatility, as are they are positively correlated.

The paper will also study the effect of investor attention on the daily volatility of cryptocurrencies, making the second research question of this paper be:

''Does market attention measured through the total number of daily Tweets and Reddit posts influence the daily volatility of Crypto?''.

Consequently, the third hypothesis will be:

H3a: Online attention measured through daily tweet volume has a positive effect on daily cryptocurrency volatility.

H3b: Online attention measured through daily Reddit post volume has a positive effect on daily cryptocurrency volatility.

With the increase in the popularity along with useability of cryptocurrencies also potentially increasing soon, cryptocurrency remains a very important topic, which partly due to its novelty has not been studied as extensively as the stock market for example. The speculative and controversial nature of cryptocurrencies in general means that the effect of the market sentiment is expected to greatly influence the price of cryptocurrencies. This study will therefore expand the understanding of the scale and effect of these influences and will help investors who are looking to invest in cryptocurrencies have a better knowledge of what drives the price and volatility of cryptocurrencies. This understanding will help investors determine to what extent they can incorporate social media sentiment into trading models or simply rely on it for indicators of how crypto prices will change.

This study will also analyse the effect that the sentiment and investor attention on Reddit, the 19th-most-visited website in the world and 7th most-visited website in the U.S according to Alexa Internet, has on cryptocurrency prices and volatility, something which has not been studied as extensively as the effect of Twitter sentiment.

2. Theoretical Framework

2.1 Social Media Sentiment and Online Attention

The predictive capabilities of social media attention and sentiment on stock prices has been researched quite extensively over the past few years. Barber and Odean (2008) show that individual investors display attention-driven buying behaviour, however this behaviour does not result in higher average returns. Bollen et al. (2011) have found in their study on Twitter data that sentiment in tweets to be predictive of stock market returns, as well as that tweet volume can be used to predict stock financials.

Although research on market sentiment on cryptocurrencies is a novel topic, there have been previous studies that have looked at this issue, which yielded mixed results. Firstly, when looking at Twitter sentiment Mohapatra et al. (2019) finds that Twitter sentiment along with the price of Bitcoin were not sufficient by themselves to accurately predict the future price of Bitcoin as it did not capture the trend of the price movement, something which was remedied by adding the historical price of Bitcoin. Similarly to the results of Mohapatra (2019), in their study that looked at predicting cryptocurrency prices using tweet volume and sentiment analysis Abraham et al. (2018) show that tweet sentiment was not a reliable measure to determine cryptocurrency prices when they were falling, as tweets regarding cryptocurrencies were generally positive or simply reporting news and therefore neutral. They do show however that the volume in Google searches and tweet volume are positively and significantly correlated with the price of cryptocurrencies. Twitter sentiment was found to be a significant predictor of cryptocurrency prices by Kraaijeveld and de Smedt (2020), who studied the predictive power of Twitter sentiment for the forecasting of cryptocurrency prices. Using bivariate causality testing for tweets between June and August 2018, they determined that Twitter sentiment can be used to predict the price returns of Bitcoin, Bitcoin Cash and Litecoin, while the tweet volume could be used to predict returns for Litecoin and Ripple. Their study also included Cardano, for which they did not find that tweet sentiment or volume could be used to determine price returns. A further interesting finding that they came across is that for all 9 cryptocurrencies studied, the one that had the highest presence of tweets originating from bot accounts was Cardano. Similarly, the usefulness of online sentiment in the prediction of cryptocurrency prices was shown by Colianni et al. (2015). In this paper cryptocurrencies were traded using algorithmic trading, which was based on

Twitter sentiment. By cleaning the data and applying supervised machine learning algorithms, they achieved a prediction accuracy exceeding 90%. Naeem et al. (2021) utilise the FEARS index of Da et al. (2015) and Twitter Happiness sentiment to study their effects on the returns of multiple cryptocurrencies. The paper states that the happiness sentiment index significantly predicts the returns of Bitcoin, Litecoin, Ripple, and Monero.

When looking at Reddit sentiment, Telli and Chen (2021) utilise similarly to this paper the number of Reddit posts, the sentiment of the post titles, along with the number of Wikipedia pageviews for four different cryptocurrencies to analyse the relationship between public attention and the evolution of certain multifractal characteristics. They conclude that information from online platforms can have a significant role in the decision-making process, especially for smaller cryptocurrencies such as Litecoin and Ripple. The effect of Reddit sentiment on the stock price surge of GME in early 2021 was studied by Wang and Luo (2021) and was found that although a strong relationship between sentiment and price movement was not certain, semantic embeddings from the subreddit r/wallstreetbets can be useful in predicting the price of GME. Staying with the r/wallstreetbets, Hu et al. (2021) found that Reddit social media activity encourages retail buying behaviour and deters short selling stocks. The paper studied the subreddit r/wallstreetbets and its effects on the 50 stocks that Robinhood halted trading for on January 28th 2021. Results indicated that high Reddit traffic, positive sentiment, high dispersion, and high connectedness resulted in less future shorting flow. Phillips and Gorse (2017) showed in their study that there is a strong relationship between Reddit usage and cryptocurrency prices, with the trading strategy based on this information outperforming a buy and hold strategy. They argue that due to Reddit's design based on subreddits that focus on specific topics, it becomes a valuable source to analyse market sentiment. Subreddits for cryptocurrencies in general, but also for specific cryptocurrencies exist, making it very easy to differentiate between them.

Investor attention was found to have a positive effect on the volatility of cryptocurrencies by Sabah (2020). By using new business venues that accept cryptocurrencies as a form of payments as a proxy for investor attention, Sabah (2020) finds that the number of crypto-accepting venues is a significant driver crypto volatility, through the usage of Vector Autoregressive Models and Granger Causality tests. El Bahrawy et al. (2019) implemented a trading strategy based on Wikipedia page views, which proved to be more successful in its returns than other baseline strategies, but only for certain periods of time. The relation between the investor attention proxied by number of tweets, retweets, favourites, and the volatility of Bitcoin, Al Guindy (2019) determined using Vector Autoregressive Regressions that investor attention can be used to predict the volatility of cryptocurrencies. Using Google Search Values, the relationship between market attention and cryptocurrency prices was examined by Subramaniam and Chakraborty (2020), who found that increased attention led to higher returns for Bitcoin and Ethereum. The paper also found an inverse relationship, in that higher returns of Bitcoin cause higher investor attention. Finally, an interesting study conducted by Ben-Rephael et al. (2017) looked at the relationship between Bloomberg terminal news searching and the news reading for specific stocks, which is used primarily by institutional investors, and Google search interest in specific stocks, which is used primarily by retail investors. The proxy for institutional attention was found to be different from the retail attention proxy, and the paper concluded that institutional attention responded quicker to major news events, which resulted in more trading.

2.2 Sentiment analysis

Understanding emotions and being able to decipher the sentiments of online messages is a task that has become crucial for marketing and financial predictions. Different methods exist that are used to perform sentiment analysis on written messages, of which the 3 main ones are: hybrid approaches, statistical methods, and knowledge-based techniques.

In knowledge-based techniques, the text is classified into different categories based on the words that occur in the text which indicate its tone, such as ''happy'', ''bad'', or ''alright''. While knowledge-based techniques do provide an advantage in terms of ease of implementation and lower cost, they can fall short when it comes to the recognition of the meaning of sentences that contain more complex structures or sarcasm, such as ''Bitcoin is going to do great in the future, totally not a scam at all''. A knowledge-based technique might struggle to classify this sentence as negative due to the way in which it is written, even though for a human the sarcasm is very easily seen (Cambria et al., 2017).

Statistical methods use machine learning and algorithms by inserting keywords and phrases, which result in the text falling under a specific classification using deep learning. Due to their design, statistical methods must be provided a large input of text for the output to be accurate, making them rather labour-intensive processes.

The third sentiment analysis method, the hybrid approach, utilises knowledge-based techniques along with statistical methods to detect emotions from textual inputs. An example of this would be SenticNet (Cambria et al., 2017).

Market sentiment online is expressed mostly through opinions posted by users, and therefore it is important to define and dissect the contents of an opinion to analyse it fully. Liu (2017) defines an opinion as a ‘’quadruple’’, in which one can find the sentiment target, the opinion about the target, the opinion holder, and the time at which the opinion was posted. If any of these four components are missing, it becomes difficult to obtain accurate analysis. Not knowing when it was posted, what is the sentiment target, and what the opinion about the target is deems it useless for analysis, while not knowing the opinion holder can also be problematic, as explained earlier that influencers such as Elon Musk for example have a much greater effect on the opinions of people than the average person. Liu (2017) also emphasises that it is particularly important to identify the target of the opinion in complex sentences with more targets, such as for example ‘’Bitcoin is going to perform very well, compared to the US Dollar which will soon collapse’’. In this example sentence, the sentiment displayed towards bitcoin is positive, and while negative sentiment is also present, it does not target Bitcoin, therefore it is crucial for the algorithms that will be used to be able to make such distinctions.

Similarly, Mohammad (2017) denotes that one of the problems surrounding sentiment analysis is detecting the sentiment expressed at different levels of text in more complex sentence structures such as long paragraphs, but also detecting sentiment at targets that are not explicitly mentioned but can be easily deducted by a human using context. In the same paper it is noted that when performing sentiment analysis, one must decide whether to label objective information as ‘’neutral’’, or whether the sentiment of the speaker will be classified as ‘’positive’’ or ‘’negative’’ based on the news, so in the case of good news objective statements being classified as positive.

3. Data and Methodology

This paper will research the effect of market sentiment on the daily returns of three cryptocurrencies, Bitcoin, Ethereum, and Cardano coin (ADA). The effect of market attention on the volatility of the above-mentioned cryptocurrencies will also be studied. The period for which daily data was obtained is between the 1st of January 2021 and the 31st of December 2021. This period was chosen as it was a particularly interesting period for the cryptocurrency market due to the rallies experienced by the three cryptocurrencies, with the first being from late 2020 to May 2021 for Cardano and Ethereum, to March 2021 for Bitcoin, along with a second rally from July to September for Cardano, July to November for Bitcoin, and July to December for Ethereum. Subsequent crashes have occurred after both rallies for the cryptocurrencies as well. These factors make the year of 2021 a very volatile and speculative year overall, which could be useful in demonstrating whether market sentiment is at least partially responsible for such wild swings, along with whether higher attention from the market results in higher volatility. The cryptocurrencies that are being studied are currently the top 2 cryptocurrencies in terms of market cap, Bitcoin and Ethereum, along with a third cryptocurrency which ranks 9th in market value terms as of February 2022, Cardano (ADA). The reason the first two cryptocurrencies have been chosen is to get a general understanding of how market influences cryptocurrencies through analysis of the effects on its two largest coins, which make up a combined total of 61% of cryptocurrency market cap as of February 2022. Cardano is a public blockchain with its own internal cryptocurrency, with the name ADA. For simplicity, when referring to Cardano in this paper, the reference will be made to the ADA coin, and not to the overall blockchain. What makes Cardano interesting to study is how it is one of the cryptocurrencies to really make a first breakthrough on a broader scale in 2021, with its market cap being 5.5B U.S. Dollars in January 2021, reaching up to 75B U.S. Dollars in November of the same year. Cardano markets itself as a third-generation cryptocurrency, compared to Bitcoin which would be first generation and Ethereum second generation cryptocurrency, therefore having certain advantages such as no mining being required to obtain it making it more eco-friendly than Bitcoin. According to the Cardano Foundation, Cardano is much more energy efficient due to how its proof of ownership works when compared to Bitcoin. Cardano uses a proof of stake which is based on a virtual resource, while

Bitcoin uses a proof of work which requires the process of ‘’mining’’ using hardware machines. This results in Cardano using approximately 6 gigawatts hours of energy annually compared to Bitcoins 115 terawatts hours of energy annually, making Cardano a more sustainable alternative to Bitcoin for example. It is also interesting to compare whether the effect of market sentiment differs between larger cryptocurrencies and a smaller one, especially since it could be argued that at the beginning of the period that will be analysed Cardano was not yet widely known by the general market.

Twitter daily sentiment and tweet volume were both obtained from the publicly available data science company intotheblock, which provides multiple analytical factors such as fundamental analysis, price predictions, and market analytics for many cryptocurrencies, including Bitcoin, Ethereum, and Cardano. The sentiment of tweets is obtained through Machine-learning powered sentiment analysis, which uses multiple NLP (Neuro-linguistic programming) based models. For each individual date it is then determined what the number of positive, neutral, and negative tweets is. Intotheblock states that their sentiment analysis algorithms are specifically based on a recurrent neural network trained on specific terms in the crypto space.

Reddit daily sentiment was obtained using the Python extension PRAW, which stands for Python Reddit API Wrapper, and is a Python package that allows for simple access to Reddit's API. After the Reddit Client Secret, Secret, and Client ID tokens were obtained, Reddit post titles, post content, submission date, and the score of the post were obtained from the respective subreddits r/bitcoin, r/Ethereum, and r/Cardano from the 1st of January 2021 to the 31st of December 2021. The data obtained is then cleaned in R Studio of posts that are not in English, hyperlinks, filler words, posts that do not contain any relevant information, posts which were posted by bots, and posts which have had their post content deleted or removed. For each of the cryptocurrencies, this reduced the number of posts to be analysed by about one third. The sentiment of the post content was determined using the sentiment extraction tool developed in the NLP group at Stanford, called syuzhet (Mohammad et al., 2013). Emotional valence, which categorised the tone of the text under different sentiments such as happy, sad, angry, etc., was obtained for each sentence in every post’s content. The emotional valence was given a value of -1 for negative sentiment posts, 0 for neutral sentiment posts, and 1 for positive sentiment posts. For each day the number of posts was summed for each sentiment category. For each individual day the percentage of positive, neutral, and negative posts was then determined, for the data to be fully comparable to the one obtained from Twitter.

For both Twitter and Reddit number of tweets/posts, a variable which will be called Overall Sentiment is created, which is obtained through subtracting the percentage of negative tweets or posts from the percentage of positive tweets or posts. For example, if on a given day 25% of posts will be positive and 8% of posts will be negative, the overall sentiment score of that day will be of 0.17. A higher overall sentiment score therefore indicates that there are more positive posts or tweets being shared that day. This process is done as the percentage of positive or negative tweets/posts on their own does not tell us much, as for example polarisation effect could have that a day which sees a relative rise in negative sentiment could also see an even bigger relative rise in positive sentiment, making the perceived rise in negative sentiment not fully reflect the story.

To determine the effect of sentiment on the crypto prices, OLS Regressions will be performed where Twitter/Reddit sentiment will be used as an independent variable, with the daily return of the cryptocurrencies as the dependent variable. As it can be assumed that higher daily returns of cryptocurrencies result in turn in higher sentiment, lagged variables of the market sentiment will be used to avoid simultaneity bias. Granger causality tests will then be used to determine not only the direction of the causality, but whether the independent variable of market sentiment can be used to predict future daily returns. Five control variables were also added to the sentiment analysis regressions, first one being the daily return of the S&P500, the most common indicator used to track the performance of the American Economy. Secondly, the daily return of the West Texas Intermediate (WTI) oil price will be used to proxy the daily price change of oil. Both oil and gold have been traditional investments for investors who wanted to hedge again economic and political turmoil, with Bitcoin and other cryptocurrencies emerging as a recent additional alternative investment. Selmi et al. (2018) explore whether Bitcoin can be a hedge for oil prices and find that both Bitcoin and Gold can be used to hedge against not only oil price movement, but also against the overall economy. Therefore, the price of gold will be added as a third control variable in our regressions, as there is reason to believe it may be significantly correlated with the price of Bitcoin and perhaps other cryptocurrencies as well.

The VIX is often used to proxy how volatile the United States stock market is along with an indicator of current overall market sentiment for the American economy and will be used as the fourth control variable that will be added to the sentiment analysis regressions.

Finally, the fifth and final control variable that will be added is the total sum of tweets/posts of that day, as a proxy for total market attention towards each respective cryptocurrency.

The hypothesis H1: *Positive investor sentiment on Twitter has a positive effect on the daily returns of cryptocurrencies* will be therefore tested through the OLS Regressions:

$$Crypto\ Daily\ Return_t = \alpha + \beta CryptoDailyReturn_{t-x} + \beta OverallSentiment_{t-x} + \beta ReturnGold_{t-x} + \beta ReturnS\&P500_{t-x} + \beta ReturnOilWTI_{t-x} + \beta ReturnVIX_{t-x} + \beta Sumoftweets_{t-x} + \varepsilon$$

The hypothesis H2 *Positive investor sentiment on Reddit has a positive effect on the daily returns of cryptocurrencies* will be tested through the OLS Regressions:

$$Crypto\ Daily\ Return_t = \alpha + \beta CryptoDailyReturn_{t-x} + \beta OverallSentiment_{t-x} + \beta ReturnGold_{t-x} + \beta ReturnS\&P500_{t-x} + \beta ReturnOilWTI_{t-x} + \beta ReturnVIX_{t-x} + \beta Sumofposts_{t-x} + \varepsilon$$

The optimal lags that will be used in the OLS regressions for sentiment analysis were determined individually for each cryptocurrency using the Akaike Information Criterion. The Akaike Information Criterion (AIC) is widely used to determine which statistical models have the best fit and can identify best future values according to Akaike (1974). The AIC will therefore compare models with different lag lengths of the independent variable and its control variables, where the model with the AIC will be used for the OLS Regressions.

To determine whether online attention has a positive effect on the volatility of cryptocurrencies, the number of daily tweets about each currency along with the sum of Reddit posts from the respective subreddit of each cryptocurrency will be used to proxy investor attention. These variables will be used as the independent variables. The volatility of the cryptocurrency will be measured through the intraday spread (highest hourly price of each day-lowest hourly price of each day), and will act as the dependent variable, similarly to the research done by Kaminski (2016). Correspondingly to the sentiment analysis, lagged variables of online attention will be used to avoid simultaneity bias, as it can be argued that higher volatility in cryptocurrency returns can also lead to a higher online attention. Additionally, a Granger causality test will then be used.

$$Intraday\ Spread_t = \ Highest\ hourly\ price_t - Lowest\ hourly\ price_t$$

Three control variables will be used in the online attention analysis. Firstly, the trading volume of the respective cryptocurrency will be used. Girard and Omran (2009) concluded in their study on the relationship between trading volume and stock price volatility on the Cairo and Alexandria stock exchanges that volume tends on average to be positively related to volatility. Bol and Henke (2003) found that in the Polish market volatility persistence tends to disappear when trading volume is included in the conditional variance equation.

The second control variable that will be used is the overall sentiment variable created in the previous regression. It will be interesting to see whether a rise in negative sentiment which can be attributed to excess selling and higher uncertainty may influence the overall volatility of a cryptocurrencies price.

Thirdly, the control variable of the VIX will be used similarly to the sentiment analysis regressions. Bouri et al. (2017) find in their study on the return volatility relationship in the Bitcoin market that there is a negative relation between the US implied volatility index (VIX) and Bitcoin volatility, something which will be explored later in this paper as well.

The hypothesis H3a: *Online attention measured through daily tweet volume has a positive effect on daily cryptocurrency volatility,* will therefore be tested through the OLS Regressions:

$$Crypto\ Daily\ Volatility_t = \ \alpha \ + \beta CryptoVolatility_{t-x} + \beta Sumoftweets_{t-x} + \\ \beta CryptoVolume_{t-x} + \beta OverallSentiment_{t-x} + \beta ReturnVIX_{t-x} + \varepsilon$$

The hypothesis H3b: *Online attention measured through daily Reddit post volume has a positive effect on daily cryptocurrency volatility,* will be tested through the OLS Regressions:

$$Crypto\ Daily\ Volatility_t = \ \alpha \ + \beta CryptoVolatility_{t-x} + \beta Sumofposts_{t-x} + \\ \beta CryptoVolume_{t-x} + \beta OverallSentiment_{t-x} + \beta ReturnVIX_{t-x} + \varepsilon$$

Similarly, to in the sentiment analysis, the optimal lags that will be used in the OLS regressions for the attention analysis were determined individually for each cryptocurrency using the Akaike Information Criterion.

The data for the daily opening and closing price of Bitcoin, Ethereum, and Cardano, along with their respective trading volume, highest daily value, and lowest daily value were obtained from intotheblock, which was also used to obtain Twitter sentiment.

For the remaining control variables, namely Gold, S&P 500 index, Oil WTI, and the VIX, the daily closing price was obtained for the period 1 January 2021 through 31 December 2021 from the website investing.com. Daily return for all three cryptocurrencies and control variables was obtained on through the formula:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Where R_t represents the daily return, with P_t representing the closing price of that day, and P_{t-1} representing the closing price of the previous day.

4. Results

This section of the paper will look at the results of the analyses that have been conducted. Firstly, summary statistics will be looked at to get an insight into the data which has been collected.

Secondly, the correlations between all variables which will briefly analysed to get a first look at the relationship between any two variables, including the independent and dependent variables which will later be used in OLS Regression analysis.

Thirdly, before the beginning of conducting the regressions, the optimal lags for the variables will be selected based on the Akaike Criterion Information, for each regression in part. Augmented Dickey-Fuller tests will also be conducted to ensure that the data does not contain a unit-root and is trend-stationary. Should there be non-stationarity found, that could lead to spurious regressions, then the non-stationarity must be avoided by making such data stationary using first differencing.

Fourthly, after determining the optimal lag-length, and ensuring all the data is stationary, OLS Regressions will be performed. Robust standard errors will be used, as they ensure the standard errors will be unaffected by any potential heteroskedasticity, strengthening the validity of any significant relationships found.

Finally, the Granger causality test will determine whether the lagged independent variables and control variables can be used to predict the future values of the dependent variables.

4.1 Summary Statistics

4.1.1 Cryptocurrency Return

In Table 1. we can observe the summary statistics on the daily return of each cryptocurrency for the duration of the year 2021. A pattern can be observed between the market cap size of the cryptocurrency and its performance for the year 2021, with Bitcoin which is by far the largest cryptocurrency of the three having the lowest mean return, but also the lowest standard deviation, and Cardano which is the smallest of the three having both the highest mean return, and highest standard deviation. We can observe how even in the cryptocurrency market which is generally referred to as very volatile and speculative, different categories seem to appear with smaller market cap cryptocurrencies and newer ones being subject to higher volatility but may also be more rewarding as was the case of Cardano in the year 2021.

Table 1. Daily Return Statistics.

Cryptocurrency	Observations	Mean Return	Standard Deviation	Maximum	Minimum
Bitcoin	365	0.0019	0.0350	0.1073	-0.1661
Ethereum	365	0.0054	0.0440	0.1840	-0.2083
Cardano	365	0.0072	0.0587	0.2742	-0.2564

4.1.2 Twitter sentiment analysis

Figure 1a. shows the daily return of Bitcoin and the Overall Twitter sentiment about Bitcoin plotted across the year. The lowest sentiment value can be observed on the 19th of May 2021, with an overall score of 0.0933, coinciding with the worst daily performance of Bitcoin throughout the year, -16.61%. On the 18th of May 2021 Chinese regulators announced that they were banning banks and payment firms from using cryptocurrencies, and as China was one of the largest crypto markets in the world at that time, the effect was felt. The highest sentiment value is observed on the 17th of December 2021, with a score of 0.3443. We can also observe how the average overall sentiment remains very high throughout the whole period, with even the lowest value still being overwhelmingly positive. This is a trend that will remain for the other currencies as well and can be explained on Twitter partially by the fact that many people who regularly tweet about cryptocurrencies are supporters, who may also have ‘’skin in the game’’, and it is therefore in their best interest to try to influence market sentiment positively. Another partial explanation may as well be the large number of bots which are programmed to encourage investing into these cryptocurrencies, especially under the ‘’mentorship’’ of fake profiles who will charge money from people who fall for these scams. Even though there are steps in place that the algorithms that perform this sentiment analysis to remove messages from bots, some messages may well be indistinguishable from a genuine post.

Figure 1a. Daily Return of Bitcoin and Overall Twitter sentiment for Bitcoin.

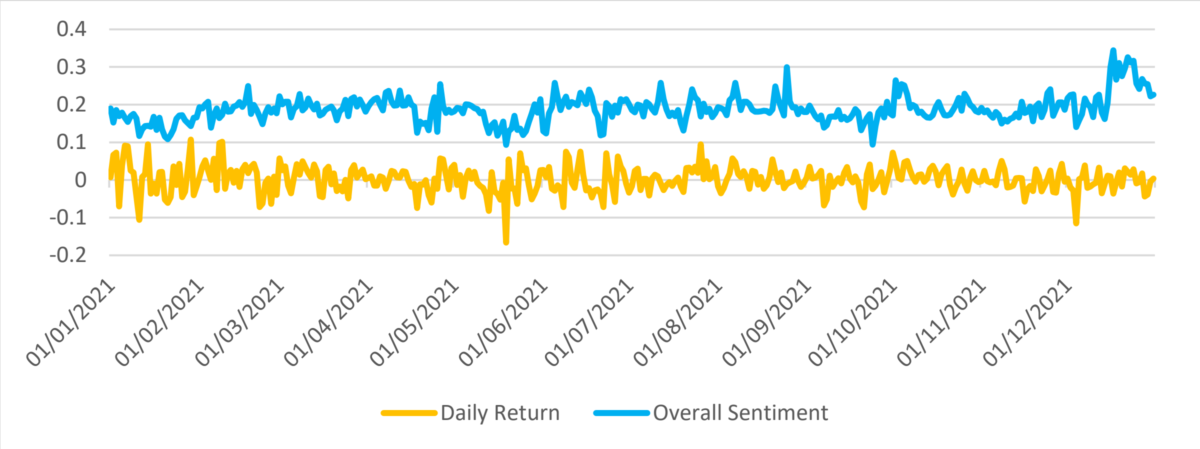


Figure 1b. shows the return of Ethereum and the Overall Twitter sentiment about Ethereum plotted across the year. What we now start to see for the first time is a negative overall sentiment, which occurs only on the 7th of August 2021 and had a value of -0.1130. Daily return on that date was of 6.483%. Highest overall sentiment was felt on the 10th of February 2021, in the middle of the early 2021 crypto rally, with a value of 0.4289.

Figure 1b. Daily Return of Ethereum and Overall Twitter sentiment for Ethereum.

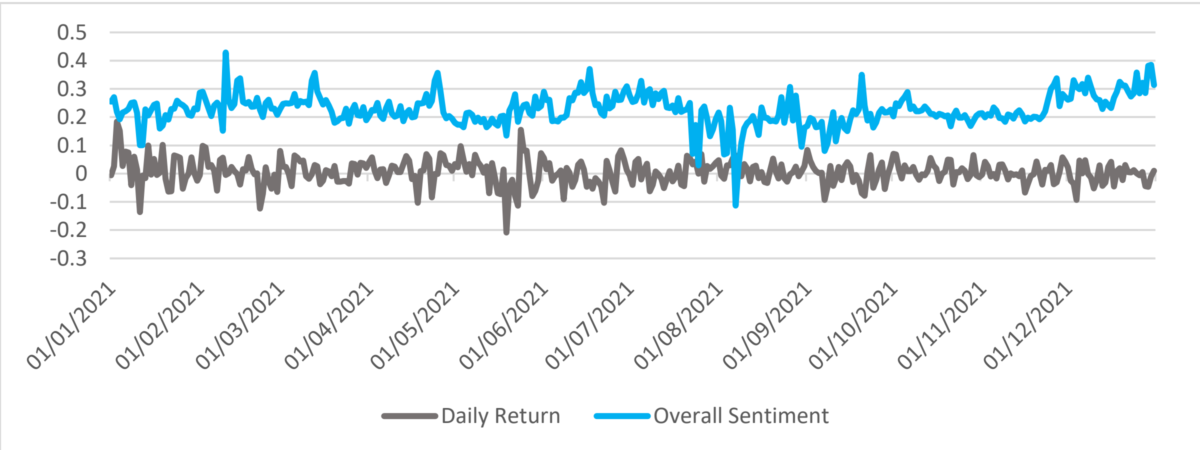
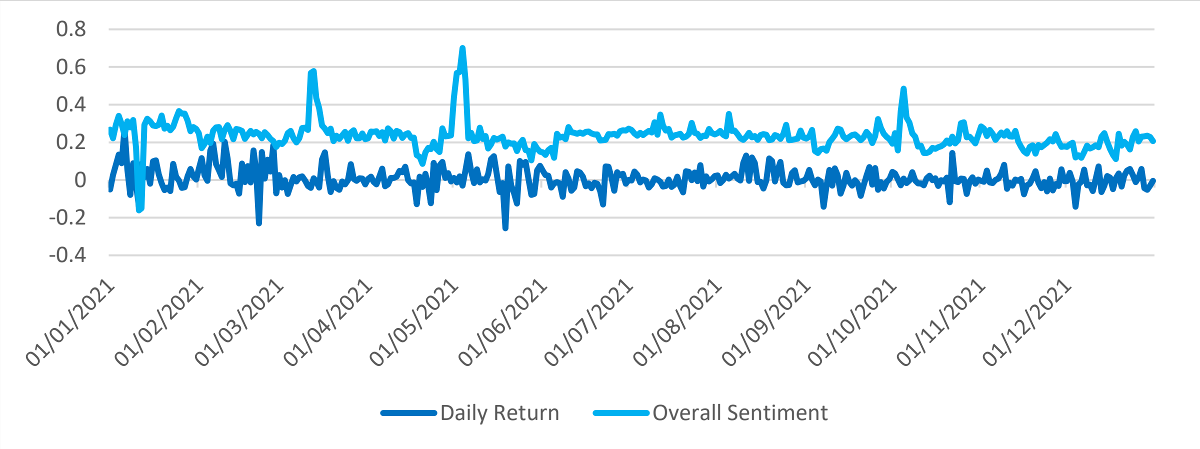


Figure 1c. shows the return of Cardano and the Overall Twitter sentiment about Cardano plotted across the year. Lowest overall sentiment was felt on the 11th of January 2021 with a score of -0.1613, while the highest was seen on the 4th of May 2021 with a score of 0.7001. In early Elon Musk tweeted that Tesla would no longer be accepting Bitcoin as a means of payment, which can also be seen in Figure 1a., with a lower average overall sentiment. Cardano was speculated to be one of the potential alternatives as a cryptocurrency payment

for Tesla, with Elon Musk even stating that Cardano's ADA and Ripple's XRP are two cryptocurrencies with more sustainable environmental costs than Bitcoin.

Figure 1c. Daily Return of Cardano and Overall Twitter sentiment for Cardano.



4.1.3 Reddit sentiment analysis

Figure 2a. shows the daily return of Bitcoin and the Overall Reddit sentiment about Bitcoin plotted across the year. What we will begin to see with the overall Reddit sentiment is that firstly overall sentiment tends to be more positive than for Twitter. This may be explained by how communities of cryptocurrencies tend to be mostly populated by those that are advocates for the cryptocurrencies, and also due to the previously mentioned so called ''echo-chamber'' effect of Reddit, where making posts and comments that fit the general opinion of the subreddit is rewarded through upvotes and awards, therefore propagating the most common opinion even further, while opinions that do not fit the general opinion are likely to be downvoted and receive criticism.

Secondly, we will also see the variation in overall sentiment between days will become larger due to the lower number of posts when compared to the amount of tweets.

For Figure 2a., we see that the lowest overall sentiment is experienced on the 24th of September 2021, with a score of 0.0805. On the 24th of September 2021, 73.7k Bitcoin Options in Open Interest are approaching expiry leading to market uncertainty. Highest overall sentiment is felt on the 28th of January 2021, which occurred during the crypto rally of early 2021, and at the same time as the frenzy around GME stock. Coincidentally, on the 28th of January 2021 Elon Musk tweeted ''#bitcoin'' with no further explanation, an event which might not have affected the price of a cryptocurrency as large as Bitcoin, however it would have been well received by r/bitcoin as the general sentiment towards Elon Musk appears to be very positive on Reddit, especially amongst the investing and cryptocurrency communities.

Figure 2a. Daily Return of Bitcoin and Overall Reddit sentiment for Bitcoin.

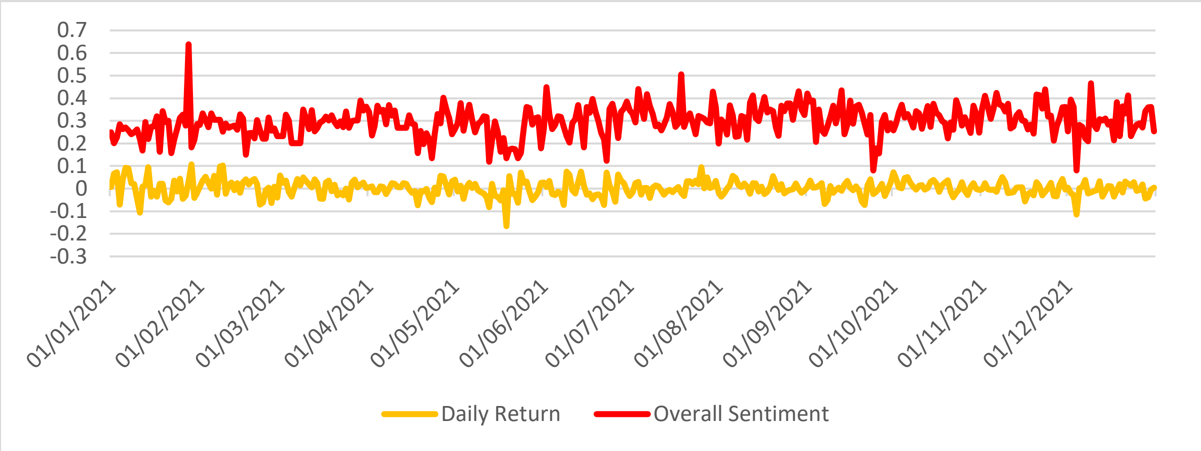


Figure 2b. shows the daily return of Ethereum and the Overall Reddit sentiment about Ethereum plotted across the year. As the subreddit becomes smaller and the average number of daily posts becomes smaller, so does the variance in the daily overall sentiment. Lowest overall sentiment is seen on the 7th of September 2021, with a score of -0.0256, a day which also received an especially low Twitter overall sentiment as seen in Figure 1b., with a score of 0.0799. That day also saw a daily return of -0.0926 in response to the announcement that El Salvador is set to adopt Bitcoin as legal tender, making it the first country to do so.

Figure 2b. Daily Return of Ethereum and Overall Reddit sentiment for Ethereum.

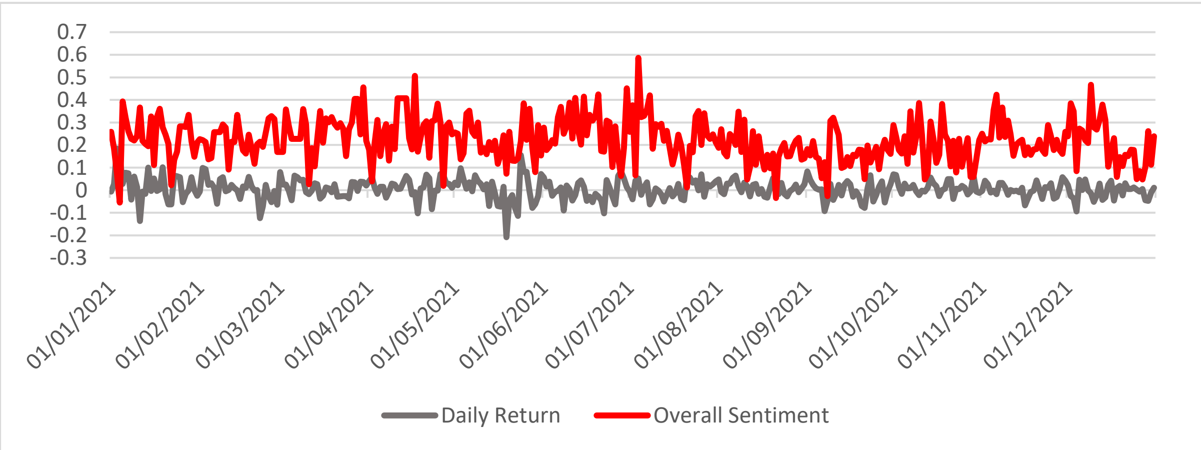
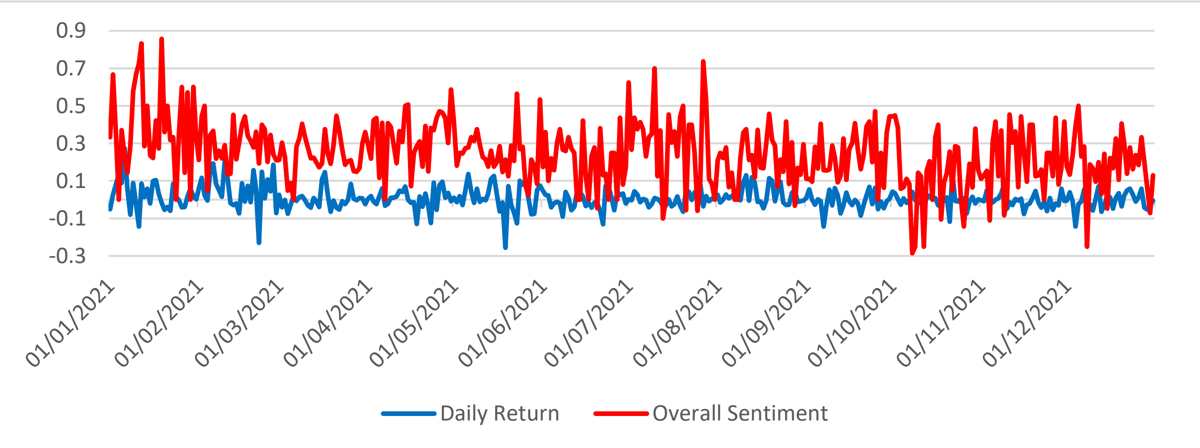


Figure 2c. shows the daily return of Cardano and the Overall Reddit sentiment about Cardano plotted across the year. The subreddit r/Cardano began the year of 2021 with only 94 750 subscribers, ending it with over 677 000 subscribers. When compared to the other two subreddits used in this study, r/bitcoin and r/Ethereum, which began the year with over 1.9 million and over 512 000 subscribers respectively, we really see the explosive growth in popularity of Cardano's ADA. The relatively low number of subscribers also means a relatively low number of daily posts, with some days in early January 2021 having as little as even 5 posts per day, although the average for that period was 14 daily posts. Therefore, there is such a high variance in the daily sentiment, with the lowest overall sentiment being observed on the 8th of October 2021 at -0.2857, and the highest overall sentiment score being seen at the beginning at the crypto rally 19 January 2021 at a whopping 0.8571, a day which followed 3 days with daily returns of 0.0991, 0.1046, and 0.03 respectively.

Figure 2c. Daily Return of Cardano and Overall Reddit sentiment for Cardano.



Tables 2. and 3. show the summary statistics of the daily prices of each cryptocurrency, along with its intraday spread which is obtained by subtracting the lowest hourly price from the highest hourly price from each respective cryptocurrency. What was seen regarding volatility in Table 1. can be seen in Table 3. as well, and that is that Cardano appears to have the highest average intraday spread relative to its average price, with Bitcoin having the lowest.

Table 2. Prices of Cryptocurrency.

Cryptocurrency	Observations	Mean Price	Standard Deviation	Max Price	Min Price
Bitcoin	365	47280.6903	9761.9832	67413.3400	29207.2700
Ethereum	365	2763.7873	1023.2924	4775.5600	733.4250
Cardano	365	1.4928	0.6140	2.9690	0.1722

Table 3. Intraday spread of Cryptocurrency prices.

Cryptocurrency	Observations	Mean Intr. Spread	Standard Deviation	Maximum	Minimum
Bitcoin	365	2910.9959	1613.7186	12864.62	711.95
Ethereum	365	213.2272	144.3317	1485.48	29.41
Cardano	365	0.1388	0.1111	1.00	0.0122

Tables 4a. and 4b. show the total number of tweets and Reddit posts for each cryptocurrency, which will be used throughout this study to measure investor attention. The difference in scale between the two measurements is obvious, with the average number of daily tweets being in the thousands to tens of thousands, while the average amount of Reddit posts is in the tens to hundreds. This difference does not indicate a large disparity in the popularity of the two websites, but rather how the conversations are being held on the respective websites, with the Twitter conversation being done mostly through individuals each tweeting their opinion on the matter, while on Reddit the discussion often takes place in the comments section of the most popular posts.

Table 4a. Total number of tweets for each cryptocurrency.

Cryptocurrency	Observations	Mean Tweets	Standard Deviation	Maximum	Minimum
Bitcoin	365	62696	25805.5776	223483	10882
Ethereum	365	28377	17429.7821	76613	4404
Cardano	365	5496	3759.2450	16284	710

Table 4b. Total number of Reddit posts for each cryptocurrency.

Cryptocurrency	Observations	Mean Posts	Standard Deviation	Maximum	Minimum
Bitcoin	365	178	105.5519	649	51
Ethereum	365	88	39.7819	358	27
Cardano	365	32	21.4301	135	5

4.2 Correlations

Correlations between the dependent variable and the independent variable along with the controls can be used to get an idea about what to expect regarding the relationship between the variables. It can also help us determine relationships between the independent variable and control variables, and amongst the control variables themselves, something which will not be shown in the OLS regressions as it is not a main point of interest of this study but can still provide interesting insight.

4.2.1 Twitter Sentiment Correlations

Table 5a provides the first correlation matrix that we will look at, that of the twitter sentiment variables for Bitcoin. The significance of each correlation is also indicated by the amount of asterixis that follow the correlation coefficient, as indicated under each table that will follow. We note a positive correlation significant at the 1% level between the Return of Bitcoin with Overall Sentiment and VIX, with a negative correlation significant at the 1% level between the Return of Bitcoin with the price of the S&P 500 index and the price of WTI Oil. We also note a positive correlation significant at the 1% level between the Price of Oil and the S&P 500, a positive correlation significant at the 5% level between the price of Gold and the S&P 500, a negative correlation significant at the 10% level between the VIX and the price of Gold, and a negative correlation significant at the 1% level between The sum of tweets and the Overall Sentiment, between the Price of Oil and VIX, and unsurprisingly between the S&P 500 and the VIX.

The strong negative correlation between the Overall Sentiment and the Sum of tweets is interesting as it could indicate that people who tweet consistently about Bitcoin are overwhelmingly positive about it, and when Bitcoin comes up in the news and a wider range of people begin to tweet about Bitcoin the sentiment becomes more diverse.

Table 5a. Pairwise correlation of Bitcoin daily returns, Twitter sentiment and control variables.

	Return BTC	Sentiment	Gold	S&P 500	Oil	VIX	Sum of tweets
Return BTC	1						
Overall Sentiment	0.2164***	1					
Gold	0.0258	-0.0640	1				
S&P 500	-0.2603***	0.0554	0.1359**	1			
Oil	-0.2028***	0.0938**	0.0377	0.3529***	1		
VIX	0.2382***	-0.0305	-0.1199*	-0.8434***	-0.3542***	1	
Sum of tweets	-0.0280	-0.3894***	-0.0349	-0.0323	-0.0419	0.0280	1

**p<0.1,   **p<0.05,   ***p<0.01*

Table 5b provides the correlation matrix between the Twitter sentiment variables for Ethereum. Unlike the correlation for Bitcoin, it can be observed that the return of Ethereum is not correlated significantly with the Overall Twitter sentiment, while the correlations between Return of Ethereum and S&P500, Oil, and VIX are highly significant. The negative significant correlation between the Overall Sentiment and the Sum of tweets remains.

Table 5b. Pairwise correlation of Ethereum daily returns, Twitter sentiment and control variables.

	Return ETH	Sentiment	Gold	S&P 500	Oil	VIX	Sum of tweets
Return ETH	1						
Overall Sentiment	-0.0290	1					
Gold	-0.0597	-0.0233	1				
S&P 500	-0.1904***	-0.0236	0.1359**	1			
Oil	-0.1785***	0.0361	0.0377	0.3529***	1		
VIX	0.1722***	0.0397	-0.1199*	-0.8434***	-0.3542***	1	
Sum of tweets	-0.0211	-0.3752***	0.0275	-0.0901	-0.1081*	0.0959	1

**p<0.1,   **p<0.05,   ***p<0.01*

Table 5c provides the correlation matrix between the Twitter sentiment variables for Cardano. We are seeing similar results as with Ethereum, with one notable change that the Sum of tweets is no longer significantly negatively correlated with the Overall Sentiment, indicating perhaps the less mainstream nature of Cardano and how it is not often discussed by those with a particular interest in cryptocurrency or Cardano itself, unlike for Bitcoin or Ethereum.

Table 5c. Pairwise correlation of Cardano daily returns, Twitter sentiment and control variables.

	Return ADA	Sentiment	Gold	S&P 500	Oil	VIX	Sum of tweets
Return ADA	1						
Overall Sentiment	0.0755	1					
Gold	-0.0610	0.0112	1				
S&P 500	-0.2110***	-0.0109	0.1359**	1			
Oil	-0.1408**	0.0402	0.0377	0.3529***	1		
VIX	0.1898***	0.0426	-0.1199*	-0.8434***	-0.3542***	1	
Sum of tweets	-0.0530	-0.0858	0.0337	-0.0605	-0.0975	0.0517	1

**p<0.1,   **p<0.05,   ***p<0.01*

4.2.2 Reddit Sentiment Correlations

Table 6a provides the correlation matrix between the Reddit sentiment variables for Bitcoin. Unlike the Twitter sentiment for Bitcoin, overall sentiment and the return of Bitcoin appear to not be significantly correlated. However, just as in the case of the Twitter Bitcoin correlations, we note a highly significant inverse correlation between the sum of tweets and the overall sentiment, which may also indicate that the people who consistently post on r/bitcoin are primarily Bitcoin enthusiasts and advocates, in accordance with the findings of Telli and Chen (2021).

Table 6a. Pairwise correlation of Bitcoin daily returns, Reddit sentiment and control variables.

	Return BTC	Sentiment	Gold	S&P 500	Oil	VIX	Sum of posts
Return BTC	1						
Overall Sentiment	0.0834	1					
Gold	0.0258	-0.0367	1				
S&P 500	-0.2602***	-0.0480	0.1359**	1			
Oil	-0.2028***	-0.0302	0.0377	0.3529***	1		
VIX	0.2382***	0.0526	-0.1199*	-0.8434***	-0.3542***	1	
Sum of posts	-0.0488	-0.4475***	0.0033	0.0036	0.0019	-0.0277	1

**p<0.1, **p<0.05, ***p<0.01*

Table 6b provides the correlation matrix between the Reddit sentiment variables for Ethereum. A positive relationship at the 5% level appears for the first time between the sum of posts and the return of Ethereum, which indicates that the subreddit r/Ethereum seems to receive more posts than usual on days with positive return. We also do not see any significant correlation between the return of Ethereum and the Overall Sentiment on r/Ethereum, like the results in Table 5b.

Table 6b. Pairwise correlation of Ethereum daily returns, Reddit sentiment and control variables.

	Return ETH	Sentiment	Gold	S&P 500	Oil	VIX	Sum of posts
Return ETH	1						
Overall Sentiment	-0.0141	1					
Gold	-0.0597	-0.0229	1				
S&P 500	-0.1904***	-0.0387	0.1359**	1			
Oil	-0.1785***	0.0047	0.0377	0.3529***	1		
VIX	0.1722***	0.0457	-0.1199*	-0.8434***	-0.3542***	1	
Sum of posts	0.1378**	-0.0074	0.0604	-0.0425	-0.0773	0.0527	1

**p<0.1, **p<0.05, ***p<0.01*

Table 6c provides the correlation matrix between the Reddit sentiment variables for Cardano. Similar results are seen as for the correlation matrix between the Reddit sentiment variables for Ethereum, with the notable difference that the sum of posts is not correlated only at the 10% level to the return of Cardano.

Table 6c. Pairwise correlation of Cardano daily returns, Reddit sentiment and control variables.

	Return ADA	Sentiment	Gold	S&P 500	Oil	VIX	Sum of posts
Return ADA	1						
Overall Sentiment	0.0369	1					
Gold	-0.0610	0.0630	1				
S&P 500	-0.2110***	0.0607	0.1359**	1			
Oil	-0.1408**	0.0196	0.0377	0.3529***	1		
VIX	0.1898***	-0.0564	-0.1199*	-0.8434***	-0.3542***	1	
Sum of posts	0.1178*	-0.0057	-0.0341	-0.0097	0.0465	-0.0109	1

**p<0.1, **p<0.05, ***p<0.01*

4.2.3 Twitter Attention Correlations

Table 7a. provides the correlation matrix between the Twitter attention variables for Bitcoin. We immediately note that the Volatility of Bitcoin is strongly negatively correlated to the sum of tweets at the 1% level, while the Volatility of Bitcoin is positively correlated to the overall sentiment and the trading volume of Bitcoin, and the 5% and 1% levels respectively. The correlation result between the sum of tweets and the volatility of Bitcoin is a surprising one, as a positive relationship was expected, meaning that more online attention measured through daily tweet volume resulted in higher volatility. However, the opposite appears to be the case when judging solely from this correlation.

Table 7a. Pairwise correlation of Bitcoin volatility, tweet volume and control variables.

	Volatility BTC	Sum Of Tweets	Overall Sentiment	Volume	VIX
Volatility BTC	1				
Sum Of Tweets	-0.1390***	1			
Overall Sentiment	0.1048**	-0.3894***	1		
Volume	0.5180***	-0.2665***	0.1742***	1	
VIX	-0.0675	0.0280	-0.0305	0.0519	1

**p<0.1, **p<0.05, ***p<0.01*

Table 7b. provides the correlation matrix between the Twitter attention variables for Ethereum. Volatility of Ethereum is not significantly correlated with the sum of tweets according to the correlation matrix. Overall sentiment negatively correlated with the volatility of Ethereum at the 5% level, which indicates that days with higher volatility for Ethereum lead to more uncertainty and therefore a lower overall sentiment on Twitter regarding Ethereum.

Table 7b. Pairwise correlation of Ethereum volatility, tweet volume and control variables.

	Volatility ETH	Sum Of Tweets	Overall Sentiment	Volume	VIX
Volatility ETH	1				
Sum Of Tweets	-0.0490	1			
Overall Sentiment	-0.1401***	-0.3752***	1		
Volume	0.1900***	0.3497***	-0.0050	1	
VIX	-0.0221	0.0959	0.0397	0.0519	1

**p<0.1, **p<0.05, ***p<0.01*

Table 7c. provides the correlation matrix between the Twitter attention variables for Cardano. Volatility of Cardano is positively and significantly correlated at the 1% level with the sum of tweets, which indicates that a higher level of online attention may lead to a higher volatility, or vice-versa.

Table 7c. Pairwise correlation of Cardano volatility, tweet volume and control variables.

	Volatility ADA	Sum Of Tweets	Overall Sentiment	Volume	VIX
Volatility ADA	1				
Sum Of Tweets	0.1375***	1			
Overall Sentiment	0.0382	-0.0858	1		
Volume	0.2299***	0.4150***	-0.1161*	1	
VIX	0.0629	0.0517	0.0426	0.0519	1

**p<0.1, **p<0.05, ***p<0.01*

4.2.4 Reddit Attention Correlations

Table 8a. provides the correlation matrix between the Reddit attention variables for Bitcoin. Similarly to what we saw in Table 7a., we see a significant at the 1% level negative correlation between the sum of posts on r/Bitcoin and the volatility of Bitcoin. Overall Sentiment is significantly correlated at the 1% level to the volatility of Bitcoin, a result similar to that of Table 7a.

Table 8a. Pairwise correlation of Bitcoin volatility, Reddit post volume and control variables.

	Volatility BTC	Sum Of Posts	Overall Sentiment	Volume	VIX
Volatility BTC	1				
Sum Of Posts	-0.1554***	1			
Overall Sentiment	0.1980***	-0.4475***	1		
Volume	0.5180***	-0.3174***	0.2045***	1	
VIX	-0.0675	-0.0277	0.0526	0.0519	1

**p<0.1, **p<0.05, ***p<0.01*

Table 8b. provides the correlation matrix between the Reddit attention variables for Ethereum. Interestingly, the only variable which is significantly correlated to the volatility of Ethereum is its Volume, with the other 3 variables showing no significant correlation.

Table 8b. Pairwise correlation of Ethereum volatility, Reddit post volume and control variables.

	Volatility ETH	Sum Of Posts	Overall Sentiment	Volume	VIX
Volatility ETH	1				
Sum Of Posts	0.0802	1			
Overall Sentiment	-0.0377	-0.0074	1		
Volume	0.1900***	0.0159	-0.1059*	1	
VIX	-0.0221	0.0527	0.0457	0.0519	1

**p<0.1, **p<0.05, ***p<0.01*

Table 8c. provides the correlation matrix between the Reddit attention variables for Cardano. The correlations of the Volatility are like those found in Table 7c., with the sum of posts being significantly correlated at the 5% level with the volatility of Cardano.

Table 8c. Pairwise correlation of Cardano volatility, Reddit post volume and control variables.

	Volatility ADA	Sum Of Posts	Overall Sentiment	Volume	VIX
Volatility ADA	1				
Sum Of Posts	0.1136**	1			
Overall Sentiment	0.0769	-0.0057	1		
Volume	0.2299***	-0.1231*	-0.1192*	1	
VIX	0.0629	-0.0109	-0.0564	0.0519	1

**p<0.1, **p<0.05, ***p<0.01*

4.3 Optimal Lag Selection and Stationarity

To perform the Ordinary Least Square Regressions in order to find out what the effect of overall sentiment is on the daily return of the cryptocurrency and of market attention on volatility, it would be tempting to just use the 1-day lagged value of the independent and control variables. However, although information travels faster than ever and posts and tweets can be seen the second they are posted, it might still take a while for information to reach the investors. For example, someone browsing Twitter or Reddit can very well see tweets or posts posted from a few days' prior, and act based upon information that is already a few days old, something which would not be caught in any measure unless there are multiple lagged values. Therefore, in order to determine whether more than 1 day lags are needed, and what the optimal amount of lags is, the models with varying lag lengths will be compared using the Akaike Information Criterion. This Information Criterion will give us scores for each lag, and the lower the score the better the fit of the model.

4.3.1 Lag Selection

Tables 9a and 9b show the optimal lags for the sentiment analysis models, for all three cryptocurrencies for both Twitter and Reddit. The lag length chosen for all 6 models is of 4 days, indicating that there is information up to the 4th lag that increases the fit of the overall model.

Table 9a. Lag selection of Twitter sentiment variables.

Lag	AIC Bitcoin	AIC Ethereum	AIC Cardano
1	-8.2027	-8.0007	-9.2805
2	-8.1787	7.6340	-8.8613
3	-7.6924	-7.0743	-8.9479
4	-10.045*	-8.7646*	-10.6726*

Table 9b. Lag selection of Reddit sentiment variables.

Lag	AIC Bitcoin	AIC Ethereum	AIC Cardano
1	-17.5265	-19.3872	-17.7097
2	-17.2905	-19.0695	-17.5431
3	-17.4872	-19.4758	-17.7706
4	-18.6577*	-20.3026*	-19.466*

Tables 10a and 10b show the optimal lags for the attention analysis models, for all three cryptocurrencies for both Twitter and Reddit. For Bitcoin both for Twitter and Reddit, the optimal lag amount remains of 4 days, while for Ethereum it becomes 2 days for the Twitter analysis and 3 days for the Reddit analysis, and for Cardano 2 days for the Twitter analysis and 4 days for the Reddit analysis. The summarised findings of the optimal lag selection can be found in Tables 11a and 11b.

Table 10a. Lag selection of Twitter Attention variables.

Lag	AIC Bitcoin	AIC Ethereum	AIC Cardano
1	84.9879	75.2219	56.1564
2	85.3386	74.8543*	56.0841*
3	85.1346	75.4817	56.3314
4	84.7351*	75.0665	56.7798

Table 10b. Lag selection of Reddit Attention variables.

Lag	AIC	AIC	AIC
1	75.6675	64.6118	47.7758
2	76.1943	64.1307	47.3229
3	76.1081	64.1211*	47.3639
4	75.5369*	64.4663	47.0169*

Table 11a. Overall Lag Selection Sentiment Analysis.

Cryptocurrency-Website	Lags Selected
Bitcoin-Twitter	4
Ethereum-Twitter	4
Cardano-Twitter	4
Bitcoin-Reddit	4
Ethereum-Reddit	4
Cardano-Reddit	4

Table 11b. Overall Lag Selection Attention Analysis.

Cryptocurrency-Website	Lags Selected
Bitcoin-Twitter	4
Ethereum-Twitter	2
Cardano-Twitter	2
Bitcoin-Reddit	4
Ethereum-Reddit	3
Cardano-Reddit	4

4.3.2 Stationarity

Using a trend augmented Dickey-Fuller test it was tested whether a unit-root was present in the time series of each variable. Data that is non-stationary would not be suitable in usage for forecasting and would have to be transformed into stationary data to avoid spurious results. Under the null hypothesis on the Dickey-Fuller test, a unit-root is present in the data. A negative result below the 5% significance threshold means we reject the null hypothesis, and the data is indeed stationary and can be used for proceeding analysis. The trend augmented Dickey-Fuller test for each variable revealed that the null hypothesis was rejected at the 5% level for every variable, with the null hypothesis not being rejected at the 1% level only for the sum of tweets for Ethereum. These results are satisfactory and mean that all variables can be used in the further analysis as they are, with no transformation needed.

4.4 OLS Regressions

The next step in the analysis is to use Ordinary Least Squares regressions to estimate the coefficients of linear regression equations which will describe the relationship between the dependent variables, namely the daily return of the Cryptocurrency and the daily volatility of the cryptocurrency, and the independent and control variables which were selected.

4.4.1 Sentiment Analysis OLS Regressions

Table 12a. provides us with the result of the OLS regression conducted for Twitter and Reddit Bitcoin sentiment. It is clear from both regressions that the 1-day lagged return has a significant effect at the 1% level for the daily return of Bitcoin. For Twitter sentiment, we interestingly see that the 1-day lagged Overall Sentiment is negatively significant at the 5% level, while the 2-day lagged Overall Sentiment is positively significant at the 5% level. This is a surprising result as perhaps it was expected that the 1-day lagged overall sentiment would have a positive effect on the return of Bitcoin, however the situation is the opposite. The 2-day lagged overall sentiment is a positive predictor of Bitcoin returns, perhaps indicating that investors do not act immediately on information they are exposed to or even share, but rather may wait a few days. The same can be seen in the case of Reddit overall sentiment, where the 3-day lagged variable is positively significant at the 1% level.

Table 12a. OLS Regressions results Sentiment Analysis on Bitcoin Returns.

Variables	Coefficient Twitter	Robust Standard Error Twitter	Coefficient Reddit	Robust Standard Error Reddit
L1.Daily Return	0.3878***	0.1098	0.4265***	0.1283
L2.Daily Return	-0.0762	0.0890	-0.0093	0.0935
L3.Daily Return	-0.0227	0.0933	-0.0252	0.0876
L4.Daily Return	0.1454	0.0940	0.0888	0.0794
L1.Overall Sentiment	-0.2906**	0.1399	-0.0507	0.0554
L2.Overall Sentiment	0.3613**	0.1425	-0.1054	0.0666
L3.Overall Sentiment	-0.1945	0.1871	0.1730***	0.0614
L4.Overall Sentiment	0.0879	0.1583	0.0477	0.0530
L1.Return Gold	-0.0559	0.4256	-0.0292	0.3926
L2.Return Gold	0.3355	0.4671	0.0379	0.3929
L3.Return Gold	-0.0812	0.4089	-0.0358	0.3946
L4.Return Gold	-0.5769	0.4112	-0.8885***	0.4331
L1.Return S&P500	0.5466	1.0061	-0.1890	0.9732
L2.Return S&P500	0.1242	0.7826	0.3778	0.6945
L3.Return S&P500	0.4552	0.8386	-0.6267	0.8166
L4.Return SP500	0.0968	0.6536	0.4427	0.7078
L1.Return Oil	-0.2251	0.1745	-0.2393	0.1867
L2.Return Oil	-0.0729	0.1787	0.0596	0.1634
L3.Return Oil	0.1342	0.1656	0.2214	0.1625
L4.Return Oil	-0.2585	0.1985	-0.1059	0.1778
L1.Return VIX	-0.0578	0.0960	-0.0788	0.0816
L2.Return VIX	0.1142**	0.0477	0.1377***	0.0508
L3.Return VIX	-0.0483	0.0546	-0.1271	0.0631
L4.Return VIX	-0.0146	0.0797	-0.0065	0.0843
L1.Sum of tweets/posts	0.0159	0.0139	0.0114	0.0128
L2.Sum of tweets/posts	-0.0222	0.0139	-0.0223	0.0155
L3.Sum of tweets/posts	0.0074	0.0164	0.0066	0.0119
L4.Sum of tweets/posts	-0.0031	0.0179	-0.0023	0.0079
Constant	0.0299	0.1280	0.0179	0.0899
R-squared	0.4647		0.3327	

**p<0.1, **p<0.05, ***p<0.01*

Table 12b. provides us with the result of the OLS regression conducted for Twitter and Reddit Ethereum sentiment. In the case of the Twitter variables, we can see according to the OLS regression that none of the variables can be used to predict the price of Ethereum, something which was hinted towards by the pairwise correlations. In the case of the Reddit variables, the 2 day lagged variable of overall sentiment is negatively significant at the 5% level, with no other variable achieving any sort of significance besides.

Table 12b. OLS Regressions results Sentiment Analysis on Ethereum Returns.

Variables	Coefficient Twitter	Robust Standard Error Twitter	Coefficient Reddit	Robust Standard Error Reddit
L1.Daily Return	0.0485	0.1300	0.0970	0.1311
L2.Daily Return	-0.0889	0.1324	0.0305	0.1342
L3.Daily Return	-0.0132	0.1244	-0.0093	0.1299
L4.Daily Return	0.0129	0.1124	-0.0493	0.1133
L1.Overall Sentiment	-0.1622	0.1153	-0.0498	0.0623
L2.Overall Sentiment	-0.0860	0.1114	-0.0992**	0.0426
L3.Overall Sentiment	-0.0185	0.0935	0.0214	0.0470
L4.Overall Sentiment	-0.1441	0.0899	-0.0104	0.0498
L1.Return Gold	-0.0703	0.4864	-0.1108	0.5646
L2.Return Gold	0.3495	0.4140	0.6764	0.4435
L3.Return Gold	-0.8328	0.5873	-0.6240	0.6050
L4.Return Gold	-0.5251	0.6467	-0.4254	0.5270
L1.Return S&P500	1.0136	1.1991	0.9504	1.1425
L2.Return S&P500	-0.3058	1.0435	-0.5675	1.1367
L3.Return S&P500	1.7037	1.0817	1.3935	1.0870
L4.Return SP500	1.5221	0.9655	1.0795	1.0724
L1.Return Oil	-0.1384	0.1927	-0.0041	0.2577
L2.Return Oil	0.1811	0.2356	0.1628	0.2399
L3.Return Oil	0.0746	0.1670	0.1553	0.2196
L4.Return Oil	0.1641	0.2625	0.0603	0.2406
L1.Return VIX	-0.0043	0.1029	0.0162	0.0848
L2.Return VIX	0.0315	0.0673	0.0617	0.0731
L3.Return VIX	0.0944	0.0709	0.1003	0.0871
L4.Return VIX	0.1445	0.0949	0.0616	0.1189
L1.Sum of tweets/posts	-0.0068	0.0151	-0.0042	0.0221
L2.Sum of tweets/posts	0.0077	0.0168	0.0143	0.0209
L3.Sum of tweets/posts	-0.0023	0.0235	0.0016	0.0218
L4.Sum of tweets/posts	-0.0088	0.0197	0.0131	0.0192
Constant	0.1919***	0.0761	-0.0812	0.0964
R-squared	0.3322		0.3185	

**p<0.1, **p<0.05, ***p<0.01*

Table 12c. shows the result of the OLS regression conducted for Twitter and Reddit Cardano sentiment. Although overall sentiment is not significant at any lag for Twitter variables, and only significant at the 10% level for the 1st and 2nd lag for Reddit variables, what can be observed that is interesting is the significance of the coefficients for the 3-day lagged return of Gold. The 4-day lagged Sum of posts variable is also significant at the 1% level for the Reddit analysis.

Table 12c. OLS Regressions results Sentiment Analysis on Cardano Returns.

Variables	Coefficient Twitter	Robust Standard Error Twitter	Coefficient Reddit	Robust Standard Error Reddit
L1.Daily Return	0.2409	0.1806	0.2279	0.1621
L2.Daily Return	-0.0930	0.1013	-0.2018*	0.1100
L3.Daily Return	0.0508	0.0872	0.0121	0.0864
L4.Daily Return	0.0435	0.1354	0.0259	0.1128
L1.Overall Sentiment	-0.0471	0.1375	0.0870*	0.0459
L2.Overall Sentiment	0.1323	0.1393	0.0730*	0.0420
L3.Overall Sentiment	0.0203	0.1195	0.0382	0.0330
L4.Overall Sentiment	-0.0606	0.1195	-0.0382	0.0459
L1.Return Gold	-0.6271	0.7604	-0.4597	0.6650
L2.Return Gold	-0.2760	0.9278	-0.4713	0.8913
L3.Return Gold	-2.0442**	0.9110	-3.1844***	0.8326
L4.Return Gold	-2.3572*	1.2174	-1.682*	0.9581
L1.Return S&P500	0.1611	1.3198	-1.8612	1.4691
L2.Return S&P500	-0.1293	1.3202	-0.9487	1.2544
L3.Return S&P500	3.1445*	1.5975	3.2862**	1.5066
L4.Return SP500	1.0338	1.4273	0.4616	1.2529
L1.Return Oil	-0.4250	0.3615	-0.3344	0.4037
L2.Return Oil	0.2911	0.2509	0.0611	0.2708
L3.Return Oil	0.2026	0.3255	0.2034	0.3184
L4.Return Oil	0.2623	0.3603	-0.1502	0.3412
L1.Return VIX	-0.1565	0.1659	-0.3399	0.1674
L2.Return VIX	0.0508	0.1223	-0.0405	0.1129
L3.Return VIX	0.1468	0.1020	0.0460	0.1132
L4.Return VIX	0.0376	0.1483	-0.1658	0.1534
L1.Sum of tweets/posts	0.0048	0.0160	0.0286*	0.0154
L2.Sum of tweets/posts	-0.0150	0.0264	0.0046	0.0138
L3.Sum of tweets/posts	-0.0231	0.0285	0.0074	0.0117
L4.Sum of tweets/posts	0.0296	0.0238	-0.0435***	0.0141
Constant	0.0286	0.0851	-0.0096	0.0394
R-squared	0.3648		0.4741	

**p<0.1, **p<0.05, ***p<0.01*

4.4.2 Attention Analysis OLS Regressions

Table 13a. provides us with the result of the OLS regression conducted for Twitter and Reddit Bitcoin attention. The 2-day lagged volatility coefficient is significant for both regressions, at the 5% level for the Twitter analysis and at the 1% level for the Reddit analysis. The 3-day lagged coefficient for the sum of tweets is negatively significant at the 5% level, however for the Reddit posts there is no coefficient significant at any lag. Interestingly, the 1-day lagged value of the overall Twitter sentiment is positively significant at the 5% level.

Table 13a. OLS Regressions results Attention Analysis on Bitcoin daily tweets/posts.

Variables	Coefficient Twitter	Robust Standard Error Twitter	Coefficient Reddit	Robust Standard Error Reddit
L1.Volatility	0.1691	0.1100	0.0982	0.1011
L2.Volatility	0.2626**	0.1269	0.3110***	0.1182
L3.Volatility	0.1434	0.1292	0.1002	0.1226
L4.Volatility	-0.0122	0.1491	-0.0364	0.1339
L1.Sum of tweets/posts	0.1040	0.2140	-0.1165	0.3150
L2.Sum of tweets/posts	0.2419	0.2378	-0.0438	0.2335
L3.Sum of tweets/posts	-0.6643**	0.3098	0.0279	0.1520
L4.Sum of tweets/posts	0.3041	0.1943	-0.0345	0.1480
L1.Volume	-0.1338	0.1970	0.0368	0.2146
L2.Volume	0.3035	0.3229	0.6419**	0.3037
L3.Volume	-0.4796	0.4403	-0.9781***	0.3682
L4.Volume	0.4607	0.3301	0.6456	0.3904
L1.Overall Sentiment	3.555**	1.6412	1.1815	0.8612
L2.Overall Sentiment	-2.3371	1.9035	-0.5849	0.9237
L3.Overall Sentiment	-0.7912	2.2888	1.0013	0.8662
L4.Overall Sentiment	0.9483	1.7655	-0.0461	1.0667
L1.Return VIX	-0.8020	0.6667	-0.5717	0.6081
L2.Return VIX	-0.0948	0.5003	-0.4631	0.5837
L3.Return VIX	-0.7041	0.6784	-0.6661	0.6224
L4.Return VIX	-1.4567**	0.6521	-1.0778	0.7372
Constant	-0.2041	4.2346	-3.7994	6.4504
R-squared	0.3468		0.3327	

**p<0.1, **p<0.05, ***p<0.01*

Table 13b. provides us with the result of the OLS regression conducted for Twitter and Reddit Ethereum volatility. The 1 and 2 day lagged coefficients of Volatility are both positively significant at the 1% level for both the Twitter and Reddit Sentiment analysis. Sum of posts is negatively significant for the 3-day lag, at the 5% level.

Table 13b. OLS Regressions results Attention Analysis on Ethereum daily tweets/posts.

Variables	Coefficient Twitter	Robust Standard Error Twitter	Coefficient Reddit	Robust Standard Error Reddit
L1.Volatility	0.3504***	0.0706	0.2713***	0.0991
L2.Volatility	0.2906***	0.0799	0.3291***	0.0917
L3.Volatility	-	-	0.1511	0.0958
L1.Sum of tweets/posts	-0.1661	0.1131	0.0941	0.1534
L2.Sum of tweets/posts	0.0554	0.1151	-0.0202	0.1595
L3.Sum of posts	-	-	-0.2235**	0.1309
L1.Volume	-0.1034	0.2057	0.0473	0.2640
L2.Volume	0.2062	0.2006	-0.1489	0.3017
L3.Volume	-	-	0.4125**	0.2116
L1.Overall Sentiment	-1.0772	0.812	-0.0853	0.4960
L2.Overall Sentiment	-0.4241	0.8834	-0.2206	0.4351
L3.Overall Sentiment	-	-	0.7454*	0.4505
L1.Return VIX	-0.0322	0.488	0.0499	0.5200
L2.Return VIX	-0.3707	0.4124	-0.0937	0.3867
L3.Return VIX	-	-	-0.1085	0.4866
Constant	0.8061	2.6159	-5.6020*	3.1966
R-squared	0.3603		0.4254	

**p<0.1, **p<0.05, ***p<0.01*

Table 13c. provides us with the result of the OLS regression conducted for Twitter and Reddit Cardano volatility. It can be noted that the lagged coefficient of the sum of Reddit posts for the 3rd day is significant at the 5% level, however none of the other lagged coefficient of the sum of posts or sum of tweets are significant at any level.

Table 13c. OLS Regressions results Attention Analysis on Cardano daily tweets/posts.

Variables	Coefficient Twitter	Robust Standard Error Twitter	Coefficient Reddit	Robust Standard Error Reddit
L1.Volatility	0.5851***	0.0757	0.5484***	0.0985
L2.Volatility	0.1438*	0.0752	0.1884	0.1226
L3.Volatility	-	-	0.0531	0.1399
L4.Volatility	-	-	-0.0048	0.0951
L1.Sum of tweets/posts	-0.0456	0.0981	0.0708	0.1077
L2.Sum of tweets/posts	0.0531	0.0951	0.0241	0.1180
L3.Sum of posts	-	-	0.1603	0.0980
L4.Sum of posts	-	-	-0.1550	0.1122
L1.Volume	0.1515	0.124	0.1648	0.2022
L2.Volume	-0.0061	0.1339	0.2277	0.2547
L3.Volume	-	-	-0.3403	0.2161
L4.Volume	-	-	-0.0431	0.1444
L1.Overall Sentiment	0.6091	0.6337	0.0484	0.3556
L2.Overall Sentiment	0.216	0.6177	-0.0806	0.3171
L3.Overall Sentiment	-	-	0.2415	0.3132
L4.Overall Sentiment	-	-	0.0302	0.2729
L1.Return VIX	-0.4187	0.5645	-1.3821**	0.5690
L2.Return VIX	-0.0588	0.3462	-0.8227	0.5190
L3.Return VIX	-	-	-0.0448	0.4178
L4.Return VIX	-	-	-0.9164	0.7659
Constant	-4.0499**	1.5766	-0.9426	2.6704
R-squared	0.5023		0.6764	

**p<0.1, **p<0.05, ***p<0.01*

4.5 Granger Causality Test

The Granger Causality Test will be used to determine whether the time series of the independent variables and control variables can be used to predict the time series of the dependent variables. Chu et al. (2016) utilised Granger Causality tests to determine that there is a linear causality from stock returns to investor sentiment in their study on the Chinese Market. The underlying methodology of the Granger Causality test is that a first time series will Granger-cause a second time series if the values of the first time series contain statistically significant information about the values in the future of the second time series. It is necessary to firstly conduct a Granger Causality test to determine whether investor sentiment is a driver of daily cryptocurrency returns, and whether investor attention is a driver of daily cryptocurrency volatility, as a Granger Causality test indicates the direction of the causation and whether the independent variables could be used to predict the dependent variables.

4.5.1. Granger Causality Test Sentiment Analysis

Table 14a. shows the results of the Granger Causality test for the sentiment analysis regarding Bitcoin on its daily return. For the Twitter analysis, the lagged variables of the Return of Oil and the Sum of Tweets are significant at the 1% level. The lagged variable of the Overall Sentiment is significant at the 5% level, and the lagged variable of the return of gold is significant at the 10% level. The joint time-series is also significant at the 1% level.

For the Reddit analysis, the lagged variables of the Overall Sentiment, the Return of Oil, and the sum of posts are all significant at the 1% level, while the return of gold and the return of the S&P500 are significant at the 5% level. The join time series is also significant at the 1% level.

Table 14a. Granger Causality Test results for Bitcoin sentiment analysis.

Equation	Excluded	Chi^2 Twitter	Chi^2 Reddit
Daily Return	**Overall Sentiment**	9.8231**	25.012***
Daily Return	**Return Gold**	8.1603*	10.098**
Daily Return	**Return S&P500**	2.2592	9.7493**
Daily Return	**Return Oil**	15.758***	26.695***
Daily Return	**Return VIX**	1.2912	5.0382
Daily Return	**Sum of tweets/posts**	15.371***	22.944***
Daily Return	**All**	81.339***	141.07***

*p<0.1, **p<0.05, ***p<0.01

Table 14b. shows the results of the Granger Causality test for the sentiment analysis regarding Ethereum on its daily return. For the Twitter analysis, no lagged variable is significant by itself at the 1% level, however the lagged return of gold along with the lagged return of the S&P500 are significant at the 5% level, and the lagged return of VIX is significant at the 10% level. Unlike the analysis done for Bitcoin, the lagged sum of tweets does not Granger cause the Daily Return of Ethereum. The time series is significant at the 1% level.

For the Reddit analysis, the lagged variable of the Overall Sentiment is significant at the 1% level, along with the lagged return of the S&P500 and the sum of posts. The lagged return of Gold is significant at the 10% level. The time series is significant at the 1% level.

Table 14b. Granger Causality Test results for Ethereum sentiment analysis.

Equation	Excluded	Chi^2 Twitter	Chi^2 Reddit
Daily Return	**Overall Sentiment**	7.9555*	40.865***
Daily Return	**Return Gold**	12.063**	9.1434*
Daily Return	**Return S&P500**	10.524**	18.818***
Daily Return	**Return Oil**	2.8544	4.8826
Daily Return	**Return VIX**	7.9145*	7.5087
Daily Return	**Sum of tweets/posts**	4.0145	13.418***
Daily Return	**All**	69.33***	142.81***

**p<0.1, **p<0.05, ***p<0.01*

Table 14c. shows the results of the Granger Causality test for the sentiment analysis regarding Cardano on its daily return. For the Twitter analysis, the lagged return of the Overall Sentiment is not significant at any level. However, all the lagged remaining variables except for the return of oil are significant at the 1% level, with the time series also significant at the 1% level.

For the Reddit analysis, the lagged Overall Sentiment, return of the S&P500 and sum of posts are all significant at the 1% level. The lagged return of Gold is significant at the 5% level. The time series, like all the other time series so far, is also significant at the 1% level.

Table 14c. Granger Causality Test results for Cardano sentiment analysis.

Equation	Excluded	Chi^2 Twitter	Chi^2 Reddit
Daily Return	**Overall Sentiment**	7.5416	27.549***
Daily Return	**Return Gold**	35.202***	11.362**
Daily Return	**Return S&P500**	40.778***	22.536***
Daily Return	**Return Oil**	7.0955	4.0651
Daily Return	**Return VIX**	29.62***	4.8761
Daily Return	**Sum of tweets/posts**	13.583***	18.685***
Daily Return	**All**	118.57***	148.14***

**p<0.1, **p<0.05, ***p<0.01*

4.5.2. Granger Causality Test Attention Analysis

Table 15a. shows the results of the Granger Causality test for the attention analysis regarding Bitcoin on its daily volatility. The lagged sum of tweets/posts is significant at the 5% level for both the Twitter and Reddit analysis. For the Twitter analysis, the lagged Overall Sentiment and lagged VIX both Granger cause the volatility of Bitcoin, significant at the 1% level. For the Reddit analysis, lagged volume and Overall Sentiment are significant at the 1% level.

Both time series are significant at the 1% level.

Table 15a. Granger Causality Test results for Bitcoin attention analysis.

Equation	Excluded	Chi^2 Twitter	Chi^2 Reddit
Volatility	**Sum of tweets/posts**	9.8274**	13.393**
Volatility	**Volume**	2.3728	27.461***
Volatility	**Overall Sentiment**	10.446***	21.069***
Volatility	**VIX**	14.667***	1.0193
Volatility	**All**	38.778***	44.002***

*p<0.1, **p<0.05, ***p<0.01

Table 15b. shows the results of the Granger Causality test for the attention analysis regarding Ethereum on its daily volatility. For the Twitter analysis, only the lagged variable of the volume is significant and at the 5% level. The time series is not significant at any level.

For the Reddit analysis, lagged volume is significant at the 1% level, while lagged sum of posts and lagged Overall Sentiment are both significant at the 1% level, along with the time series as a whole.

Table 15b. Granger Causality Test results for Ethereum attention analysis.

Equation	Excluded	Chi^2 Twitter	Chi^2 Reddit
Volatility	**Sum of tweets/posts**	4.6105	9.546**
Volatility	**Volume**	9.8972**	20.813***
Volatility	**Overall Sentiment**	0.7011	12.172**
Volatility	**VIX**	5.6481	6.7279
Volatility	**All**	22.263	43.186***

*p<0.1, **p<0.05, ***p<0.01

Table 15c. shows the results of the Granger Causality test for the attention analysis regarding Cardano on its daily volatility. For the Twitter analysis, although none of their variables on their own Granger cause the Volatility of Cardano, the time series can be used as a predictor due to its significance at the 1% level.

For the Reddit analysis, lagged sum of posts and lagged VIX are significant at the 10% level, while the time series is significant at the 5% level.

Table 15c. Granger Causality Test results for Cardano attention analysis.

Equation	Excluded	Chi^2 Twitter	Chi^2 Reddit
Volatility	**Sum of tweets/posts**	7.7481	9.0816*
Volatility	**Volume**	-	-
Volatility	**Overall Sentiment**	7.059	12.600
Volatility	**VIX**	3.4192	9.4164*
Volatility	**All**	30.947***	23.649**

**p<0.1, **p<0.05, ***p<0.01*

4.6 Overall Results

When looking at the 4-day lagged variables, we can conclude that Twitter Overall Sentiment Granger causes the daily return only of Bitcoin, while Reddit Overall Sentiment Granger causes the daily returns of all three cryptocurrencies used in this study. The study of the effect that Twitter sentiment has on the price of cryptocurrencies or stocks has led to mixed results, with Reboredo and Ugolini (2018) finding for example that Twitter sentiment has no sizeable impact on returns of renewable energy firms' stocks. Glenski et al. (2019) found that using Reddit social signals from the main cryptocurrency subreddits lead to more accurate models in forecasting prices one day ahead for Bitcoin. This study finds that Overall Sentiment measured through the content of posts on the subreddits of the respective cryptocurrencies does Granger cause the daily return of all three studied cryptocurrencies. The difference in the sentiment analysis between the two social media platforms could be explained in part due to the structure of the websites, where individuals on Twitter may obtain their news regarding cryptocurrencies from people that they already follow, and therefore be exposed to consistent sentiment expressed by the people they follow which may be based on their own beliefs and biases. Subreddits on the other hand work as forums where anyone can make a post expressing their own views on the subject, and visitors to these subreddits may be exposed to a wider range of opinions, despite the echo-chamber effect which was earlier discussed that may lead to subreddits having an opposite effect. The sum of tweets or posts was not found to significantly affect the daily return of any cryptocurrency, at any lag.

For the attention analysis, the results show that the market attention for Bitcoin measured through both the sum of tweets and the sum of Reddit posts Granger causes the volatility of Bitcoin. For Ethereum, only the total sum of Reddit posts Granger causes its volatility. The predictability of Bitcoin volatility using the sum of tweets was also found in the study by Al Guindy (2019).

Table 16. Hypothesis Results based on the results of the Granger Causality tests.

	H1	**H2**	**H3a**	**H3b**
Bitcoin	Not rejected	Not rejected	Not rejected	Not rejected
Ethereum	Rejected	Not rejected	Rejected	Not rejected
Cardano	Rejected	Not rejected	Rejected	Rejected

5. Conclusion

5.1 Discussion

The paper studied firstly the effect of social media attention on cryptocurrency daily returns and attempted to answer the question: *''Does market sentiment measured through Twitter and Reddit sentiment influence the daily returns of Crypto?* ''. Based on the results, we can reject the first hypothesis that *Positive investor sentiment on Twitter has a positive effect on the daily returns of cryptocurrencies* for Ethereum and Cardano, while we cannot reject hypothesis 1 for Bitcoin. The second hypothesis that that *Positive investor sentiment on Reddit has a positive effect on the daily returns of cryptocurrencies* is not rejected for any cryptocurrency. The answer to the first research question is therefore *market sentiment measured through Twitter sentiment does not influence the daily returns of Crypto*, while *market sentiment measured through Reddit sentiment does influence the daily returns of Crypto*. Given that for 2 of the 3 cryptocurrencies studied Twitter sentiment was shown to not influence the daily returns, we can assume that Twitter sentiment can not be accurately used to predict the future daily returns of cryptocurrencies. For Reddit, this paper showed that for all 3 cryptocurrencies can be used to predict the future daily returns of cryptocurrencies.

The paper studied secondly what the effect of investor attention on cryptocurrency volatility is and attempted to answer the question: '*Does market attention measured through the total number of daily Tweets and Reddit posts influence the daily volatility of Crypto?''.* The third hypothesis A that *Online attention measured through daily tweet volume has a positive effect on daily cryptocurrency volatility* is rejected for Ethereum and Cardano, while the third hypothesis B that *Online attention measured through daily Reddit post volume has a positive effect on daily cryptocurrency volatility* is rejected only for Cardano. For Bitcoin we see therefore that both Twitter and Reddit market sentiment can be used to predict its future daily returns, while the online attention the cryptocurrency receives can be used to predict its future volatility. With Ethereum the results are more mixed, with only Reddit sentiment being able to predict its return, and market attention only measured through the sum of Reddit posts being able to predict its volatility. For Cardano we cannot use either proxy for market attention to predict its volatility, however Reddit sentiment can be used to predict its future return. The answer to the second research question is therefore harder to give than for the first research question, due to the mixed results. However, as the hypothesis 3A was rejected for 2 out of the 3 cryptocurrencies for Twitter, and hypothesis 3B was rejected for only 1 out of 3 cryptocurrencies for Reddit, we can state that *online attention measured through daily Twitter post volume does not have a positive effect on daily cryptocurrency volatility,* while *online attention measured through daily Reddit post volume does have a positive effect on daily cryptocurrency volatility*.

The difference in the results between the three cryptocurrencies which were studied is interesting and deserves a further study of its own. One of the potential reasons for why Twitter sentiment cannot explain the price returns of Cardano while Reddit sentiment appears to be able to, is as found out by Kraaijeveld and de Smedt (2020) that the cryptocurrency that had the highest presence of tweets originating from bot accounts was Cardano. This could also be a partial reason as to why the volume of tweets regarding Cardano is unable to predict its future volatility, as the bot accounts would consistently post tweets regardless of overall market attention from humans, leading to a relatively more consistent attention.

Another reason for the lack of effect of the market attention on the volatility of Cardano could be due to its size. With a market cap of approximately 25 billion dollars, it is significantly smaller than Ethereum with a market cap of approximately 300 billion dollars, or Bitcoin with its market cap of approximately 750 billion dollars, the highest in the world by a large margin. When compared to these two cryptocurrencies, Cardano pales in size. This could therefore lead to major news for Cardano not breaking through to the wider public as it would in the case of Bitcoin and Ethereum, but the rise in attention remaining mostly restrained among the already existing community of Cardano enthusiasts. The difference in size between Cardano and Bitcoin or Ethereum was one of the reasons as to why this cryptocurrency was chosen for this study, to see the differences in effects of market sentiment and market attention amongst cryptocurrencies of different rankings.

Throughout academia, research on the effect of Twitter sentiment on Bitcoin and stocks has yielded mixed results. Like the findings of this paper, Li et al. (2019) conclude that Twitter provided powerful social signals in the prediction of Bitcoin, something which this paper does not find. Papers such as Abraham et al. (2018) also were not able to find tweet sentiment as a reliable indicator of price changes for Ethereum and Bitcoin. The results of this paper on the effect of Twitter sentiment on the daily return of Ethereum and Cardano also match those of the study conducted on the effect of stock message boards are related to stock markets' performance by Antweiler and Frank (2004). The study concluded that there is useful information present on the stock message boards, but that its economic effect is quite small. Using tweets and the sentiment of the news, Lamon et al. (2017) created a model that was able to correctly predict the days with the largest percent increases and percent decreases in price for Bitcoin and Ethereum over a specified period of 67 days. Adding a further component of market sentiment in the form of news article sentiment will be discussed in the limitation section of this paper and could explain why in contrast to the Lamon et al. (2017) paper, this paper found no significant link between Twitter sentiment and Ethereum price returns.

When looking at the link between market attention and volatility, the previously mentioned paper by Antweiler and Frank (2004) also studied this topic and found that message posting does help to predict volatility both at daily frequencies and within the trading day.

By using Google search value as a proxy for market attention, Smales (2022) found positive association with volatility when studying the 5 largest cryptocurrencies, with Bitcoin and Ethereum both included a result which is shared with this paper when the proxy for market attention is the number of Reddit posts, but not when the proxy is number of tweets.

An interesting relationship was found by Lin (2021) is that past cryptocurrency returns are a significant positive driver of the attention they receive. This could potentially explain why this study found no relationship between the lagged attention values and the daily returns of cryptocurrencies, as the high attention does not cause high returns, but rather high returns cause high attention.

5.2 Limitations and further research

This paper studied the effects of social media sentiment using Tweet sentiment and Reddit sentiment as proxies, and the effects of social media attention on the volatility of cryptocurrencies, using the sum of tweets and Reddit posts as the proxy for attention. Certain improvements could be made to further similar research, constituting therefore the limitations of this paper.

A first limitation would be the lack of inclusion of the effect that news regarding cryptocurrencies can provide. As noted by Lamon et al. (2017), the sentiment of the news can have a strong predictive power over the future returns of cryptocurrencies. For example, when China declared that it would ban financial institutions and payment companies from providing services related to cryptocurrency transactions on May 18th, the price of Bitcoin dropped by 13.5% the next day. It can be speculated that the crypto prices fluctuate mostly based on good/bad news, as do stocks. However, the fluctuations of crypto prices may be even stronger, as unlike for stocks it is argued by some that there are no ''fundamentals'' to refer to, therefore making it difficult to determine whether an overreaction is taking place. Caporale and Plastun (2019) conducted a study on the price overreaction in the crypto market, and concluded that overreactions cannot be exploited for profit, contrary to the theory that market overreaction are followed by a reversal in returns, or to the inertia theory which states that the overreaction will continue in the same direction. Therefore, including a news variable in future studies would grant the models to determine the relation between market sentiment and crypto price returns a stronger explanatory power, although finding all major news articles relating to crypto and determining their overall sentiment could be a daunting task.

A second limitation of this paper would be the proxy of social media sentiment not covering all major social media sites. While both Reddit and Twitter are both social media behemoths, with Reddit having approximately 430 million monthly users, and Twitter 436 million monthly users, these numbers pale in comparison to some of the largest social media platforms in the world, such as Instagram (approx. 2 billion monthly users), and Tiktok (approx. 1 billion monthly users). According to The Economist (2022), videos with the tag #moneytok had received 10.6 billion views on Tiktok. These videos mostly offer financial advice and can be seen as one of the main ways the new generation of retail investors will be introduced to the world of investments, with crypto being a large point of interest. The reason Reddit and Twitter posts were chosen, and Instagram and Tiktok posts were not, is due to the format in which information is presented on these social media platforms. Text based data which is most commonly found on Reddit and Twitter is easier to obtain and analyse than the image-based macros or video-based data that is used on Instagram and Tiktok. Stronger algorithms that can accurately analyse and store data from these platforms would be needed to perform sentiment analysis using Tiktok and Instagram sentiment. This would also obtain a more accurate and complete proxy for social media sentiment.

A third limitation of this paper would be the lack of linguistical diversity of the social media sentiment. Although English can be considered the lingua franca of the internet and may well be representative of overall sentiment, certain markets may be involuntarily left out of the analysis by limiting the language studied to English.

A fourth limitation of this paper is that the effect that individuals have through their social media accounts is not looked at in detail. The most given example in this instance would be the effect of Elon Musk's tweets on the price of Bitcoin and most notably Dogecoin amongst others. Studying the impact that individuals with a large following may have on the price of cryptocurrencies is important not just for the further understanding of how crypto markets work, but also for legal reasons in determining what is fair use of social media as it can lead to the manipulation of crypto prices.

An additional limitation has to do with the sentiment analysis itself. Although most posts on Reddit and most tweets are text based, image macros are also often posted throughout these social media platforms, which might've as well contained investor sentiment. These posts would have been missed in the analysis, as the technique used could not register their content. Additionally, the algorithm which checked the sentiment of the tweets and Reddit posts could have made certain errors, especially in the cataloguing of sarcastic statements, internet lingo, or jokes. Therefore, the measured sentiment may not always be 100% accurate, and could have missed or mislabelled the sentiment of certain statements.

5.3 Final Remarks

This paper studied the impact of social media sentiment on cryptocurrency returns, and of social media attention on cryptocurrency volatility. The research yielded mixed results, showing that social media sentiment does have a positive impact on the returns of Bitcoin, while results for Ethereum and Cardano were mixed, with Twitter sentiment not being able to predict the future returns, while Reddit sentiment could be used to predict future returns of all cryptocurrencies. Social media attention was also shown to be able to predict the volatility of Bitcoin, while it was not able to predict Cardano volatility. Ethereum results were mixed, with Reddit attention being able to predict its volatility while Twitter attention was not.

Overall, this study is important in helping to deepen the understanding of what drives cryptocurrency prices, as it is a relatively novel topic with research on it yielding varying conclusions. Investors can use social media sentiment as one of the predictors of cryptocurrency returns, however as this paper has shown investors must be wary that social media sentiment may not always be a reliable indicator of future cryptocurrency returns.

6. Bibliography

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review*, *1*(3), Article 1.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Al Guindy, M. (2021). Cryptocurrency price volatility and investor attention. *International Review of Economics & Finance*, *76*, 556-570.

Andrei, D., & Hasler, M. (2015). Investor attention and stock market volatility. *The Review of Financial studies*, *28*(1), 33-72.

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

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36(2), 334-357.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio management*, *25*(3), 239-264.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21 (2), 785-818.

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies*, *30*(9), 3009-3047.

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, *45*, 150-156.

Blitz, D, Huisman, R., Swinkels, L., and Van Vliet, P. (2020). Media attention and the volatility effect, *Finance Research Letters*, 36 (10), 101317.

Bohl, M. T., & Henke, H. (2003). Trading volume and stock market volatility: The Polish case. *International Review of Financial Analysis*, *12*(5), 513-525.

Bollen, J., Mao, H, & Zeng, X (2011). Twitter mood predicts the stock market, *Journal of Computational Science*, 2 (1), 1-8.

Bouri, E., Azzi, G., & Dyhrberg, A. H. (2017). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics*, 11(1).

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). Affective computing and sentiment analysis. In *A practical guide to sentiment analysis* (pp. 1-10). Springer, Cham.

Caporale, G. M., & Plastun, A. (2019). Price overreactions in the cryptocurrency market. *Journal of Economic Studies*, August 29th.

Caporale, G. M., & Plastun, A. (2019). The day of the week effect in the cryptocurrency market. *Finance Research Letters*, *31*.

Chan, S., Chu, J., Nadarajah, S., & Osterrieder, J. (2017). A Statistical Analysis of Cryptocurrencies. *Journal of Risk and Financial Management*. 10(12).

Cheung, A., Roca, E., & Su, J. J. (2015). Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. *Applied Economics*, *47*(23), 2348-2358.

Chu, X., Wu, C., & Qiu, J. (2016). A nonlinear Granger causality test between stock returns and investor sentiment for Chinese stock market: a wavelet-based approach. *Applied Economics*, 48(21), 1915-1924.

Colianni, S., Rosales, S., & Signorotti, M. (2015). Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. CS229 Project, 1(5).

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, 28(1), 1– 32.`

ElBahrawy, A., Alessandretti, L., & Baronchelli, A. (2019). Wikipedia and cryptocurrencies: Interplay between collective attention and market performance. *Frontiers in Blockchain*, 2-12.

García-Monleón, F., Danvila-del-Valle, I., & Lara, F. J. (2021). Intrinsic value in crypto currencies. *Technological Forecasting and Social Change*, *162*, 120393.

Girard, E., & Omran, M. (2009). On the relationship between trading volume and stock price volatility in CASE. *International Journal of Managerial Finance*, 5, 110-134.

Glenski, M., Weninger, T., & Volkova, S. (2019). Improved forecasting of cryptocurrency price using social signals. *arXiv preprint arXiv:1907.00558*.

Hu, D., Jones, C. M., Zhang, V., & Zhang, X. (2021). The rise of reddit: How social media affects retail investors and short sellers' roles in price discovery. Available at SSRN 3807655.

Kaminski, J. (2014). Nowcasting the bitcoin market with twitter signals. *arXiv preprint arXiv:1406.7577*.

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 101188.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*, *10*(4), e0123923.

Lamon, C., Nielsen, E., & Redondo, E. (2017). Cryptocurrency price prediction using news and social media sentiment. *SMU Data Science Review*, 1(3), 1-22.

Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, 7, 98.

Lin, Z. Y. (2021). Investor attention and cryptocurrency performance. *Finance Research Letters*, *40*, 101702.

Liu, B. (2017). Many facets of sentiment analysis. In *A practical guide to sentiment analysis* (pp. 11-39). Springer, Cham.

Long, C., Lucey, B. M., & Yarovaya, L. (2021). " I Just Like the Stock" versus" Fear and Loathing on Main Street": The Role of Reddit Sentiment in the GameStop Short Squeeze. *SSRN Electronic Journal*, 31, 1-37.

Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. *Financial Review*, 40(1), 1-9.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

Mohammad, S. M. (2017). *Challenges in sentiment analysis*. In *A practical guide to sentiment analysis* (pp. 61-83). Springer, Cham.

Mohapatra, S., Ahmed, N., & Alencar, P. (2019). Kryptooracle: A real-time cryptocurrency price prediction platform using twitter sentiments. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 5544-5551). IEEE.

Naeem, M. A., Mbarki, I., & Shahzad, S. J. H. (2021). Predictive role of online investor sentiment for cryptocurrency market: evidence from happiness and fears. *International Review of Economics & Finance*, 73, 496-514.

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Phillips, R. C., & Gorse, D. (2017, November). Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In *2017 IEEE symposium series on computational intelligence (SSCI)* (pp. 1-7). IEEE.

Reboredo, J. C., & Ugolini, A. (2018). The impact of twitter sentiment on renewable energy stocks. *Energy Economics*, *76*, 153-169.

Sabah, N. (2020). Cryptocurrency accepting venues, investor attention, and volatility. *Finance Research Letters*, *36*, 101339.

Selmi, R., Mensi, W., Hammoudeh, S., & Bouoiyour, J. (2018). Is Bitcoin a hedge, a safe haven or a diversifier for oil price movements? A comparison with gold. *Energy Economics*, 74(2018), 787-801.

Shen, D., Urquhart, A., & Wang, P. (2019). Does twitter predict Bitcoin? *Economics Letters*, 174(2019), 118-122.

Smales, L. A. (2022). Investor attention in cryptocurrency markets. *International Review of Financial Analysis*, 79, 101972.

Smuts, N. (2019). What drives cryptocurrency prices? an investigation of google trends and telegram sentiment. *ACM SIGMETRICS Performance Evaluation Review*, *46*(3), 131-134.

Stosic, D., Stosic, D., Ludermir, T. B., & Stosic, T. (2018). Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, *507*, 499-509.

Subramaniam, S., & Chakraborty, M. (2020). Investor attention and cryptocurrency returns: Evidence from quantile causality approach. *Journal of Behavioral Finance*, *21*(1), 103-115.

Telli, Ş., & Chen, H. (2021). Multifractal behavior relationship between crypto markets and Wikipedia-Reddit online platforms. *Chaos, Solitons & Fractals*, 152, 111331.

The Economist (2022). Personal Finance is a hit on Tiktok. Finance and Economics section, January 29th, 2022, edition. https://www.economist.com/finance-and-economics/2022/01/29/personal-finance-is-a-hit-on-tiktok.

Varma, V., Kurisinkel, L. J., & Radhakrishnan, P. (2017). Social Media Summarization. In *A Practical Guide to Sentiment Analysis* (pp. 135-153). Springer, Cham.

Wang, C., & Luo, B. (2021). Predicting $ GME Stock Price Movement Using Sentiment from Reddit r/wallstreetbets. In *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing* (pp. 22-30).

Zhang, W., & Wang, P. (2020). Investor attention and the pricing of cryptocurrency market. *Evolutionary and Institutional Economics Review*, *17*(2), 445-468.