The University of Chicago

How social network sentiments and activities impact cryptocurrencies' trading volume and volatility?

By

Xin Su

June 2022

Faculty Advisor: Toulis Panos

Preceptor: Sanja Miklin

How social network sentiments and activities impact cryptocurrencies' trading volume and volatility?

Xin Su*

June 2021

Abstract

This paper studies how the trading volume and volatility of Bitcoin and Ether are impacted by social media such as Twitter and GitHub. Bitcoin and Ether are the two largest cryptocurrencies in terms of market capitalization, and represent over 70% of the total cryptocurrency market in combined value. Twitter and GitHub are two of the most important social network platforms. By utilizing linear regression models on tweets and GitHub activities, I find that tweet sentiments and GitHub activities are important predictors of cryptocurrency's trading volume and volatility. Specifically, as the total number of tweets increase, the cryptocurrency's trading volume and volatility tend to rise. The sentiment in the tweets is also found to be correlated with trading volume and volatility. The positive tweets increase is associated with decrease in the cryptocurrency's trading volume, while the increase in negative tweets leads to larger trading volume in cryptocurrency. When GitHub activities increase, the trading volume and volatility in cryptocurrency decreases. Theses findings suggest that trading volume and volatility in cryptocurrencies have associations with social media platforms.

*MA of Computational Social Science, University of Chicago, xin4@uchicago.edu.

1 Introduction

Since the rise of the blockchain technology, the cryptocurrencies and their markets have become an increasingly popular social phenomenon for investors, researchers, and policy makers. Cryptocurrency is decentralized and allows people from all over the world to trade freely across borders in a fast, digital, and secure way. This thesis aims to find out how social media sentiments are impacting the trading volume and volatility of cryptocurrencies.

The data used in this project include trading volume data, social media data, volatility data, stock market index, and bonds interest rate data. Methods and models including sentiment analysis and regressions were applied. In the regressions, the dependent variables are trading volumes and volatility of the cryptocurrencies, and the key independent variables are sentiments and activities from social media platforms. Our results indicate that the intensity of social media discussion, different sentiments, and activities have an impact on the trading volume and volatility of cryptocurrency, both in the US and globally.

This paper is linked with several branches of literature. The first branch focuses on the development of the cryptocurrency market itself. It has already been shown that predicting changes in Bitcoin and Ether prices with Twitter data and Google Trends may be possible (Abraham et al., 2018). Also, time series analysis and natural language processing have been used to investigate how Twitter sentiments can impact the price returns for the nine largest cryptocurrencies (Kraaijeveld and De Smedt, 2020). Some research has also suggested that social media sentiments can introduce price bubbles in the cryptocurrency market (Chen and Hafner, 2019). The existence of peer influences and network effects in the cryptocurrency market have also been demonstrated under certain simulated trading data and circumstances

(Krafft, Della Penna and Pentland, 2018; Stylianou et al., 2021). These studies have contributed to a better understanding of the cryptocurrency markets and their functions in the economy at large. This thesis adds to the literature by analyzing how sentiments and activities from social media affect the trading volume and volatility of the cryptocurrencies.

Another branch of literature studies how social networks can impact the actions in different financial asset markets. Research in behavioral finance has shown, for example, that participation and returns from the stock market can be driven by emotion levels extracted from social media applications (Bollen and Mao, 2011; Nofer and Hinz, 2015). Online reviews can also have an impact on music and video game sales (Heimbach and Hinz, 2012; Zhu and Zhang, 2010). The Twitter search volume index and sentiments have also been used to predict prices in oil, gold, and market indices (Rao and Srivastava, 2013). The above studies have investigated the importance of sentiments from social networks in the financial markets. This project extends this topic to the cryptocurrency markets and investigates how social network is correlated with the trading volume and volatility in crypto markets.

In this paper, after the introduction, Section 2 reviews relevant background information for the project; Section 3 is the model description part; Section 4 talks about the data and sources; Section 5 shows the main results and the discussions associated with it; Section 6 consists of the robustness checks and limitations; Section 7 is the conclusion section.

2 Background

This section includes descriptions of cryptocurrencies and social media. It explains the reason why we focused on the Bitcoin and Ether markets, as well as the Twitter and GitHub platforms of social media.

Cryptocurrency & Blockchain

A cryptocurrency is a digital currency used as a medium of exchange based on cryptographic technology, which helps to secure transactions, control volume, and verify transactions (Chohan, 2017). The underlying technology is known as blockchain technology, a digital, decentralized and distributed ledger in which transactions are logged and added in chronological order to create permanent and tamper-proof records. Blockchain technology is based on peer-to-peer connectivity and cryptographic security, allowing a decentralized approach with enhanced transparency and trust instead of the centralized and opaque nature of traditional monetary systems (Treiblmaier, 2018). Between the launch of Bitcoin in 2009 until 2019, more than 1600 blockchain related cryptocurrencies have entered circulation (Wilson, 2019). Until 2017, the cryptocurrency market was valued at approximately $300 billion, with nearly 80% of that value in Bitcoin tokens, while the landscape for cryptocurrencies has grown exponentially over the years (Babkin Alexander et al., 2017; Dimitrova et al., 2019). With large volumes, the natural characteristics of this market, and the fact that many cryptocurrencies are used as a medium of exchange for daily payments, the cryptocurrency market inherently has similar characteristics to other financial markets, particularly precious metals (Omane-Adjepong, Alagidede and Akosah, 2019).

Bitcoin

Bitcoin is the world's first decentralized cryptocurrency. It is a type of digital asset that uses public-key cryptography to record, sign and send transactions over the Bitcoin blockchain, all done without the oversight of a central authority. The Bitcoin network (with an uppercase "B") was launched in January 2009 by an anonymous computer programmer or group of programmers under the pseudonym "Satoshi Nakamoto." The network is a peer-to-peer electronic payment system that uses a cryptocurrency called bitcoin (lower case "b") to transfer value over the internet or act as a store of value like gold and silver. Each bitcoin is made up of 100 million satoshis (the smallest units of bitcoin), making individual bitcoin divisible up to eight decimal places. That means anyone can purchase a fraction of a bitcoin with as little as one U.S. dollar. Bitcoin is classified as a currency under CoinDesk's Digital Asset Classification Standard (DACS).

The past few years have been the most exciting time periods for bitcoin. In November 2021, the price of bitcoin reached an all time high of over $68,000 after starting the year 2021 at just under $30,000, and the crypto industry as a whole has grown to a total market cap of more than $2 trillion [1]. As of June 28, 2022, the bitcoin price had decreased to $20,020.05. The price drop happened in the first 6 months of 2022 in bitcoin has left the token's market capitalization at $381,974,563,477.17. So far this year, bitcoin has experienced a dramatic change of -56.61%[2]. Which makes it even more important to understand the behavioral aspects of bitcoin.

[1] Extracted from this report: https://time.com/nextadvisor/investing/cryptocurrency/Bitcoin-price-history/: :text=Bitcoin%20(BTC)%20reached%20an%20all,of%20more%20than%20%242%20trillion.

[2] Extracted from website: https://app.intotheblock.com/

Ether

Another cryptocurrency of interest is Ether. Just like Bitcoin, Ether is also classified as a software platform under CoinDesk's Digital Asset Classification Standard (DACS). With time, people began to realize that one of the underlying innovations of Bitcoin, the blockchain, could be utilized for other purposes. And that is how Ether was born. Ether is a blockchain-based software platform that can be used for sending and receiving value globally with its native cryptocurrency, ether, without any third-party interference. First proposed in 2013 by Russian-Canadian computer programmer Vitalik Buterin, Ether was designed to expand the utility of cryptocurrencies by allowing developers to create their own special applications. Unlike traditional apps, these Ether-based applications, called "decentralized applications", or "dapps", are self-executing thanks to the use of smart contracts. Smart contracts are code-based programs that are stored on the Ether blockchain and automatically carry out certain functions when predetermined conditions are met. That can be anything from sending a transaction when a certain event takes place or loaning funds once collateral is deposited into a designated wallet. The smart contracts form the basis of all dapps built on Ether, as well as all other dapps created across other blockchain platforms[3].

An estimated 106 million people worldwide now use cryptocurrency exchanges, according to 2021 data from the cryptocurrency exchange Crypto.com[4]. The crypotocurrency market has indeed become enormous. In this market, Bitcoin and Ether are certainly the two most important and influential tokens.

[3] Extracted from website: https://app.intotheblock.com/

[4] Extracted from this report: https://explodingtopics.com/blog/blockchain-stats.

Social Networks

The first social network platform studied in this paper is Twitter, which was created in March 2006. Twitter is definitely one of the most influential social media platforms in the US. According to a report from USA today, Twitter had already reached more than 330 million monthly active users by 2019 (Molina, 2017). To deal with the enormous size of these data, we use sentiment analysis, which provides information on whether a tweet is positive, negative or neutral. Our Twitter sentiment indicator uses IntoTheBlock's proprietary[5] classification machine learning algorithm to infer the connotation of messages discussing a particular crypto-asset and its ticker. These are plotted over time to track market sentiment and the level of activity in relation to a crypto-asset in Twitter.

The Twitter sentiment indicator is helpful to gauge market participants' emotions. Sentiment can be a leading indicator at times in June and July 2021, when the market was tabulated, as was the case for Ether. In most occasions, though, sentiment tends to be a reactive indicator. In other words, there tends to be more positive sentiments when prices rise and negative sentiments when prices fall. Overall, Twitter sentiment can be helpful to estimate the market's current stance on a particular crypto-asset. However, due to its reactive nature, it should be used cautiously when trading.[6]

Another social media platform of interest is GitHub, which is a provider of internet hosting for software development and version control using Git. It offers the distributed version control and source code management (SCM) functionality of Git, plus its own features.

[5] https://app.intotheblock.com/

[6] The sentiments are extracted from the intotheblock website: https://app.intotheblock.com/. The sentiments are classified as Positive, Negative, and Neutral

GitHub provides access control and several collaboration features such as bug tracking, feature requests, task management, continuous integration, and wikis for every project and it has been a subsidiary of Microsoft since 2018. According to the company's official website and internet, GitHub has accumulated approximately 83 million users as of June 2022[7].

As mentioned above, GitHub has many features and functionalities. Some GitHub indicators are selected for this article's purpose, which provide insights regarding development activity for a crypto asset based on commits (changes made to the code of the asset ecosystem by developers), stars and issues (interest and engagement shown by the community) and pull requests (changes and network improvements submitted and approved over time).

Other Important Background

As previous research has pointed out, cryptocurrency and stock prices are correlated after accounting for cryptocurrency's volatility. In addition, many of the factors that affect regular stock prices also affect cryptocurrency prices (Gil-Alana, Abakah and Rojo, 2020). Some popular explanations mention that investors and traders treat cryptocurrency the same way they treat stocks, so prices tend to trend in the same manner. There is also research showing the dynamic spillover mechanism between cryptocurrency and certain bonds (Hassan et al., 2022). These facts are going to be considered in Section 3.

Last but not least, the trading volume data for cryptocurrency will be divided into two big categories: the U.S trading volume and the global trading volume. Specifically, as countries around the world grapple with ways to control crypocurrencies, the United States already

[7] https://en.wikipedia.org/wiki/GitHubcite$_{n}ote - techcrunch - 5$

has a number of rules in place, and is likely to introduce more. It is these existing regulations, both at state and national level, that prevent many crypto exchanges from operating in the US. Exchanges have to register as money service businesses (MSBs) and get money transfer licenses (Hughes, 2017). Some international exchanges have made the decision that the cost and paperwork are not worth the effort. Others provide services that aren't compatible with U.S. laws outside the United States. Therefore, due to cryptocurrency regulations, it is common to see cryptocurrency exchange companies divide their exchanges, softwares, and services between users in the US region and the rest of the world. For example, the largest cryptocurrency exchange in the world, Binance, was banned in the United States on regulatory grounds in 2019. In response, Binance and other investors opened Binance.US, a separate exchange registered with the United States Financial Crimes Enforcement Network and designed to comply with all applicable US laws.[8] Under the same logic, the second largest cryptocurrency exchange FTX also consists of two sections: namely FTX and FTX.US.

Because of these facts, even though the data retrieved from the US exchanges and cryptocurrency market are not the largest data sets we could found, they surely are more reliable and tractable than many of the other ones available. Thus, the US trading volume data are used for the main analysis, while the global trading volume data are used to test robustness in Section 6.

[8] From news: https://captainaltcoin.com/binance-vs-binance-us/

3 Models

Our main interest is to investigate how sentiments are impacting the trading volume in cryptocurrency. Therefore, our regression model's independent variable is sentiment and main dependent variable is trading volume. However, as mentioned above, trading volume can also be impacted by other factors that can be correlated with sentiment. So we include such factors as control variables. The following model is our main specification:

$$Trading\ Volume_t = \beta_0 + \beta_1' Sentiment_t + \beta_2 S\&P_t + \beta_3 Bill_t + \beta_4 Bond_t + \epsilon_t \tag{1}$$

where, $Trading\ Volume_t$ is the total trading volume of the chosen cryptocurrency in week t; $Sentiment_t$ is a vector that contains three variables. The first one is the ratio of the positive tweets over total tweets on the chosen media platform in week t, the second one is the ratio of the negative tweets over total tweets on the chosen media platform in week t, and the third variable is the total number of tweets on the chosen social media platform in week t. $S\&P_t$ stands for a hundred times the log of the S&P 500 index in week t; $Bill_t$ represents the 3-month treasury bill secondary market rate on discount basis; $Bond_t$ stands for the market yield on 10-year US treasury securities.

The model includes several control variables to enhance internal validity. As mentioned in Section 2, the stock markets and bond markets are both correlated to the cryptocurrency markets under certain circumstances. This is why the control variables we include are the stock market index and bond interest rates. For bond interest rates, we include short-term and long-term bonds to provide us both short-term and long-term controls in the model.

More details on these control variables can be found in Section 4.

The relationship between volatility and sentiment is also going to be analyzed. For the investigation in volatility, the same model is used and the dependant variable is changed to volatility.

$$Volatility_t = \beta_0 + \beta_1' Sentiment_t + \beta_2 S\&P_t + \beta_3 Bills_t + \beta_4 Bonds_t + \epsilon_t \tag{2}$$

In this specification, the dependant variable $Volatility_t$ stands for the volatility of the chosen cryptocurrency at week t.

For GitHub, the model is as follows:

$$Y_t = \beta_0 + \beta_1 GitHub_t + \beta_2 S\&P_t + \beta_3 Bills_t + \beta_4 Bonds_t + \epsilon_t \tag{3}$$

where Y_t stands for the dependent variable of interest, either the trading volume or volatility; $GitHub_t$ stands for the GitHub activity of interest, either commits, stars or open issues.

4 Data

This section provides more details on the definitions of the data, the shape of the data, and data sources. The raw data were collected in various sizes and frequencies. For consistency, all the data were transformed into weekly format. Then, the longest overlap was found, which is the period from July 15, 2019 to June 13, 2022, and this serves as the main period of interest of this project.

Some additional processing occurred on specific data sets. A salient example is the US trading volume data for Bitcoin. Initially, the trading volume data from different exchanges in the United States were gathered. Then, all the trading volume data from different US exchanges in this chosen period of time were converted into weekly formats, after which they were summed up into one variable. Thus, the US Bitcoin trading volume consists of all the recorded trading volume in Bitcoin from different major US exchanges between July 15, 2019 and June 13, 2022. The following table and pie chart provide a more detailed view of what the US Bitcoin trading volume variable's structure look like.

Table 1: Bitcoin US Trading Volume Sources

Exchange	Volume(BTC)	Market Share
bit-x	1.2M	3%
bitfinex	8.3M	20%
bitstamp	6.4M	15%
coinbase	17.2M	41%
gemini	2.1M	5%
kraken	5.8M	14%
others	0.63M	2%

Notes: The others included all the exchanges that contributes less than 1% of the total volume

The global Bitcoin and Ether trading volume are much larger compared to the US trading volume. The global data sets are extracted from Yahoo finance, a platform that provides financial news, data and commentary media property and is part of the Yahoo network.

For control variables, the stock market index chosen is the S&P 500 index, the stock market index tracking the stock performance of 500 large companies listed on stock exchanges in the United States. It is one of the most commonly followed equity indices. As of December 31,

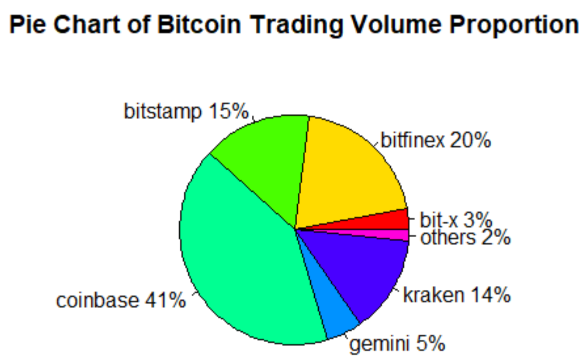

Figure 1: US Bitcoin Trading Volume Proportion
Notes: This figure shows the US Bitcoin trading volume proportion for different cryptocurrency exchanges between the period July 15, 2019 and June 13, 2022.

2020, more than \$5.4 trillion was invested in assets tied to the performance of the index.[9] The short-term bond chosen is the 3-month treasury bill secondary market rate on discount basis, while the long-term bonds chosen is the market yield on U.S. treasury securities at 10-Year constant maturity, Quoted on an Investment Basis. All the data sources for the control variables are gathered from the Federal Reserve Economic Data (FRED), the database maintained by the Research division of the Federal Reserve Bank of St. Louis that consists of than 816,000 economic time series from various sources.[10] Their data labels are SP500, TB3MS, and DGS10 respectively. Note that the unit for stock market variable is the S&P 500 market index. And the units for both short term and long term bonds are their market interest rate. These units will be the default units when mentioned the control variables later on in the article. All the control variables' data sets are gathered in weekly format and in the main interest time period. It is safe to say that the data for the control

[9] From its official website and internet: Official Website
[10] FRED's official website: https://fredhelp.stlouisfed.org/fred/about/about-fred/what-is-fred/

variables are reliable.

Table 2 provides the summary statistics of all the data for our main results after being processed.

Table 2: Summary Statistics

Statistic	N	Mean	St.Dev	Min	Max
Total Number of Tweets about Bitcoin	153	220941	180655	25	910527
Number of Positive Tweets about Bitcoin	153	52991	45514	2	201020
Number of Negative Tweets about Bitcoin	153	12990	13185	2	83082
Bitcoin GitHub Commits	111	226313	24006	151767	265161
Bitcoin GitHub Stars	111	409514	57066	285009	501887
Bitcoin GitHub Open Issues	111	4107	290	2385	4461
US Bitcoin Trading Volume	153	272777	148840	121393	1077218
Bitcoin Volatility	153	0.555	0.216	0.188	1.380
S&P 500	153	3716	633	2376	4784
3-Month Treasury Bill	153	0.520	0.681	0.014	2.138
10-Year Treasury Securities	153	1.451	0.588	0.556	3.126

Notes: 153 means the main interest period July 15, 2019 and June 13, 2022 consist of 153 weeks. Only the Github data consist of a slightly shorter period of time

Another point worth mentioning is that the volatility data in cryptocurrency are usually divided into indicators that measure the 30-day or 60-day variations in price for a specific crypto asset. Both indicators were tried for the purpose of this project, but the results were similar. So the 30-day variation version is used by default, while the 60-day variation version is only used for robustness checks. In contrast, the stock markets, which usually consist of 252 trading days in a calendar year, the crypto markets are 24/7 and the annualization formula need to take into account all the 365 days. Thus, the volatility indicator is calculated as the standard deviation of the period's daily returns and annualizing the variation. Additionally, the volatility data was multiplied by 1000 when being used.

Generally speaking, volatility is a measure used to evaluate the riskiness of investing in a particular asset. High volatility essentially means high price swings (either on the upside or downside). Low volatility, on the contrary, points to stagnating price status. The volatility used in this project is no different. In crypto, historically, periods of low volatility tend to precede large spikes in volatility as can be seen in October 2018, April 2019 and July 2020 on the Bitcoin chart. As such, these could be interpreted as moments to trade in either direction. On the other hand, large spikes in volatility generally point to a trend-reversal, suggesting that the current trend may be over-extended.[11]

5 Results & Discussions

Table 3 presents the results of equations (1) and (2). Column A and B present the regression coefficients from equation (1), and column C and D present the regression coefficients from equation (2). Column A and C do not include any controls, while column B and D include full sets of control variables.

From the results in table 3, we see that the total number of tweets does have a statistically significantly positive correlation with both Bitcoin trading volume and volatility. More specifically, the model suggests that every 1 standard deviation increase in total number of tweets would increase trading volume by 0.261 standard deviations[12] when no controls are included. Similarly, every 1 standard deviation increase in total number of tweets would increase the volatility by 0.183 standard deviations[13] when no controls are included. After

[11] Volatility is extracted from this website: https://app.intotheblock.com/

[12] Recall in the summary statistics table, the standard deviation for total sentiment is 180,655 and the standard deviation for total volume is 148,840. So $0.261 = 180.655 * 0.215/148.840$.The following calculations follow the same logic.

[13] $0.183 = 0.219 * 180.655/216$. The following calculations follow the same logic.

Table 3: Regression results for Twitter sentiments on US Bitcoin data

	A	B	C	D
	Volume (in thousands)	Volume (in thousands)	Volatility	Volatility
Total number of tweets	0.215***	0.289***	0.219**	0.360***
(in thousands)	(0.006)	(0.074)	(0.110)	(0.108)
Ratio of Positive Tweets	-8.253***	-5.837**	-0.009***	-0.120***
	(2.558)	(2.513)	(0.003)	(0.004)
Ratio of Negative Tweets	17.166***	15.298*	0.022*	0.010
	(10.738)	(8.164)	(0.012)	(0.012)
S&P 500		-5.733***		-8.91***
		(1.315)		(1.918)
3-month Bills		-110.750***		-282.600***
		(32.828)		(47.880)
10-year Bonds		101.791***		292.500***
		(35.359)		(51.580)
R-squared	0.197	0.309	0.115	0.302
Adjusted R-squared	0.180	0.281	0.097	0.274
No. observations	153	153	153	153

Notes: This table presents the relationship between Twitter sentiments and the US Bitcoin trading volume, and Bitcoin volatility. Column A and B present the regression coefficients from equation (1), and column C and D present the regression coefficients from equation (2). Column A and C do not include any controls, while column B and D include full sets of control variables. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

control variables are added, the relationship trend between the total number of tweets and volume or volatility does not change qualitatively, but the magnitudes change a bit. To be specific, with the control variables in effect, every 1 standard deviation increase in total number of tweets would increase the trading volume by 0.351 standard deviations, while every 1 standard deviation increase in total number of tweets would increase the volatility by 0.301 standard deviations.

For the ratio of positive tweets, the results show that every one percentage point increase in the ratio of positive tweets generate a 0.055 standard deviation[14] decrease in trading volume when no controls are included. However, this effect on volatility is quite small and can be ignored when no controls are included. After adding control variables, the effect of the ratio of positive tweets on trading volume decrease in magnitude to almost one third of the previous value, and the coefficient is still economically insignificant on volatility.

When it comes to the ratio of negative tweets, the results are slightly more complicated. Before adding the control variables, the ratio of negative tweets has a negative correlation with the trading volume. Every one percentage point increase in the ratio of negative tweets increase the trading volume by 0.115 standard deviation. After the control variables are added, the correlation is every one percentage point increase in the ratio of negative tweets increase the trading volume by 0.103 standard deviation. Thus, when the ratio of negative tweets increase, the Bitcoin trading volume tend to increase.

For control variables, S&P 500 and 3-month bills both have negative correlation with the volume and the volatility, while the 10-year bonds has positive association with the volume

[14] $0.055 = 8.253/148.84$. The following calculations follow the same logic.

and volatility. All the estimates in the control variables are statistically significant and have a decent magnitude.

Overall, the total number of Tweets has a positive correlation with Bitcoin trading volume and volatility. The ratio of positive tweets has a negative correlation with the trading volume while the ratio of negative tweets has a positive correlation with the trading volume. Both ratios' correlation with volatility are economically insignificant. And the market factors used as control variables also have correlation with both the volume and volatility.

Table 4, table 5, and table 6 demonstrate the results for how different GitHub activities correlate with US Bitcoin trading volume and volatility according to equation (3). Column A and C do not include any controls, while column B and D include full sets of control variables.

Table 4: Regression Results for GitHub Commit activities on Bitcoin

	A Volume (in thousands)	B Volume (in thousands)	C Volatility	D Volatility
Commits	-2.314***	-5.148***	0.924	-2.941**
(in thousands)	(0.485)	(0.973)	(0.721)	(1.356)
S&P 500		1.458		-0.857
		(2.082)		(2.900)
3-month Bills		-153.043		-572.399***
		(98.833)		(137.653)
10-year Bonds		137.408**		325.218***
		(55.712)		(77.595)
R-squared	0.173	0.275	0.015	0.242
Adjusted R-squared	0.165	0.248	0.006	0.213
No. observations	111	111	111	111

Notes: This table presents the results for how GitHub commits activity correlate with US Bitcoin trading volume and volatility according to equation (3). The key independent variable is GitHub commits activity in week t. Column A and C do not include any controls, while column B and D include full sets of control variables. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Table 5: Regression Results for GitHub Stars activities on Bitcoin

	A Volume (in thousands)	B Volume (in thousands)	C Volatility	D Volatility
Stars	-0.958***	-2.621***	0.401	-1.498**
(in thousands)	(0.205)	(0.455)	(0.303)	(0.644)
S&P 500		2.610		-0.198
		(2.117)		(2.996)
3-month Bills		-129.167		-558.752***
		(97.349)		(137.760)
10-year Bonds		154.877***		335.204***
		(55.058)		(77.913)
R-squared	0.167	0.302	0.016	0.246
Adjusted R-squared	0.160	0.276	0.007	0.218
No. observations	111	111	111	111

Notes: This table presents the results for how GitHub stars activity correlate with US Bitcoin trading volume and volatility according to equation (3). The key independent variable is GitHub stars activity in week t. Column A and C do not include any controls, while column B and D include full sets of control variables. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Table 6: Regression Results for GitHub Open Issues activities on Bitcoin

	A Volume (in thousands)	B Volume (in thousands)	C Volatility	D Volatility
Open Issues	-169.22***	-237.203***	50.300	-217.276**
(in thousands)	(41.09)	(66.045)	(59.960)	(86.007)
S&P 500		-0.281		-0.263
		(2.215)		(2.885)
3-month Bills		-208.864**		-607.597***
		(104.536)		(136.132)
10-year Bonds		72.537		279.226***
		(59.063)		(76.914)
R-squared	0.135	0.183	0.006	0.253
Adjusted R-squared	0.127	0.152	-0.003	0.225
No. observations	111	111	111	111

Notes: This table presents the results for how GitHub open issues activity correlate with US Bitcoin trading volume and volatility according to equation (3). The key independent variable is GitHub open issue activity in week t. Column A and C do not include any controls, while column B and D include full sets of control variables. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

The results in table 4, table 5, and table 6 show that no matter which GitHub action we are focusing on, the regression results are similar. Therefore, these tables suggest that when the GitHub activities increase, the trading volume in Bitcoin tends to decrease. When it comes to volatility, the trend is the same as it is in the volume. The only difference is that the result is only statistically significant in the model with the full controls.

The results on Twitter and GitHub do have some differences. One obvious difference is that the relationship between the GitHub activities and trading volume is negative, while the relationship between the Twitter total sentiments and trading volume is positive. This is reasonable since these two platforms serve different purposes and consist of different groups of users. Twitter as a social media platform is likely to include a wider variety of users than GitHub.

6 Robustness Checks & Limitations

The main results only focus on the US and Bitcoin data. However, Twitter is a global platform and contains users from outside US. Also, other than Bitcoin, Ether is another influential cryptocurrency that deserves attention. Thus, the robustness checks aim to check whether the global data and Ether data share the same trends. This first part of the robustness checks includes applying equation (1) with the dependant variable changing from the US Bitcoin trading volume to global Bitcoin trading volume. The second part of the robustness checks includes applying equation (1) and equation (2) with the dependant variable changing from the US Bitcoin trading volume to global Ether trading volume; the independent variables changing from the Bitcoin sentiments to Ether sentiments. The third part of the robustness checks includes applying equation (1) and equation (2) to different variations of the US bitcoin data and adding time trends to double check the results. The purpose is to test and see whether the trend in the results are still the same as the results in table 3.

Table 7 shows the summary statistics for the data used in the robustness checks and not listed in the main result section.

Table 7: Summary Statistics for robustness check

Statistic	N	Mean	St.Dev	Min	Max
Global Bitcoin Trading Volume	153	2.454e+11	1.120e+11	9.772e+10	7.667e+11
Total Ether Twitter Sentiment	153	113033	102138	27	406595
Positive Ether Twitter Sentiment	153	31718	29281	4	126640
Negative Ether Twitter Sentiment	153	5601	7273	2	54404
Global Ether Trading Volume	153	1.264e+11	6.883e+10	4.004e+10	4.007e+11
Ether Volatility	153	0.683	0.239	0.313	1.450

Table 8: Regression Results for Twitter sentiments on global Bitcoin data

	A Volume (in trillions)	B Volume (in trillions)
Total number of tweets	0.277***	0.280***
(in thousands)	(0.052)	(0.051)
Ratio of Positive Tweets	-1.026	-6.375***
	(1.290)	(1.723)
Ratio of Negative Tweets	10.591*	-1.521
	(5.697)	(5.597)
S&P 500		-1.884**
		(0.901)
3-month Bills		-140.226***
		(22.054)
10-year Bonds		96.512***
		(24.239)
R-squared	0.264	0.427
Adjusted R-squared	0.249	0.403
No. observations	153	153

Notes: Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Table 8 represents the regression results for the relationship between Twitter sentiment and global Bitcoin trading volume. The key difference between table 3 and table 8 is the dependent variable: the former trading volume is US based and the latter one is global. The global Bitcoin results indicate that total number of tweets has a positive correlation with global Bitcoin trading volume. The ratio of positive tweets has a negative correlation with the global Bitcoin trading volume. Note that this effect is only statistically significant after adding the control variables. For the ratio of negative tweets, the association is positive. Also, the coefficient became statistically insignificant after the control variables are added. The control variables' correlation with the global Bitcoin trading data are also demonstrated. The S&P 500 has a negative association with the trading volume; the 3-month bills also has a negative correlation with trading volume; the 10-year bonds has a positive effect on the trading volume. In general, we found similar trends in how Twitter sentiments and control variables changing the global Bitcoin trading volume and this suggests that the results in the main results section are solid.

The volatility results for global Bitcoin data are the same as the US Bitcoin data since the volatility is measured by Bitcoin price and the Twitter sentiments are still the same as the previous ones. Thus, the volatility results part are not shown here.

Table 9 represents the results of applying equations (1) and (2) on global Ether data. The major differences between this table and table 3 is the dependant variable changing from US Bitcoin trading volume to global Ether trading volume. In addition , the dependant variables changing from Bitcoin Twitter sentiments to Ether Twitter sentiments. Column A and B present the regression coefficients from equation (1), and column C and D present

Table 9: Regression results for Twitter sentiments on Global Ether data

	A Volume (in trillions)	B Volume (in trillions)	C Volatility	D Volatility
Total number of tweets	0.086	-0.102	-0.409*	-0.521*
(in thousands)	(0.059)	(0.064)	(0.225)	(0.266)
Ratio of Positive Tweets	0.787	-2.765***	1.189	-7.350**
	(0.666)	(0.800)	(2.538)	(3.298)
Ratio of Negative Tweets	13.953***	6.561**	30.005**	13.270
	(3.246)	(3.167)	(12.380)	(13.053)
S&P 500		0.792		-3.864
		(0.586)		(2.413)
3-month Bills		-73.275***		-278.100***
		(13.917)		(57.354)
10-year Bonds		33.581**		162.066***
		(14.522)		(59.850)
R-squared	0.204	0.419	0.042	0.183
Adjusted R-squared	0.188	0.395	0.022	0.149
No. observations	153	153	153	153

Notes: Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

the regression coefficients from equation (2). Column A and C do not include any controls, while column B and D include full sets of control variables. The results suggest that the total number of tweets does not have a statistically significant correlation with the global Ether trading volume. The ratio of positive tweets in Ether has a negative association with the global Ether trading volume after the control variables are included. Before adding the control variables, the result is not significant. When it comes to the ratio of negative tweets, the association is positive both with or without the control variables. In control variables, the coefficient for S&P 500 is not significant, the 3-month bills has a negative association, and the 10-year bonds has a positive correlation with trading volume. In the volatility columns, the total number of tweets has a negative correlation with the volatility, both with or without the control variables. The ratio of positive tweets has a negative correlation with the volatility with the control variables, and the result is not significant without the control variables. The ratio of negative tweets has a positive association without the control variables, and the coefficient is not significant with the control variables. In control variables, the S&P 500 does not provide significant coefficient, the 3-month bills has a negative correlation with volatility, and the 10-year bonds has a positive effect on volatility.

In summary, the results on Ether data do have some differences with the results on the Bitcoin data. With respect to trading volume, they share similar results in how the ratio of positive tweets and negative tweets impact the trading volume, 3-month bill, and 10-year bond impact the trading volume. however, the total number of tweets does not have an correlation with the trading volume in Ether. While in Bitcoin, this effect is significant both in US based data and global data. In addition, unlike the Bitcoin data, the S&P 500 index does not have a association with the volume in Ether data. In the volatility columns, the differences are even more obvious. They also share similarities in the results in the ratio of positive tweets, ratio of negative tweets, 3-month bills, and 10-year bonds. But the total number of tweets has a negative association in the Ether data, which is in sharp contrast to the positive association in the Bitcoin data. Additionally, the S&P 500 index here is also not significant.

From the previous comparison between the Bitcoin and Ether, it is apparent that the effect of the total number of tweets and S&P 500 on both crypocurrencies' trading volume and volatility are different. However, the rest of the variables have similar results on both crypocurrencies. Therefore, we can conclude that some of the features in Bitcoin are the same as the features in Ether due to their similar natures. Nevertheless, there also exist some differences between the two results and how their trading volumes are associated with our independent variables. There are many possible explanations that can explain these differences. For example, Bitcoin is more decentralized than Ether since Ether is still being impacted by its developer company, Ether Foundation. In many cases, they are used for different purposes: Ether can be used in smart contracts and Bitcoin cannot.

The motivation behind doing this third robustness test is that the Us bitcoin data shows some anomalous trends between March 2021 and May 2021. To be more specific, the Twitter total sentiments in the week of April 19, 2021 is 58, and the Twitter total sentiments in the week of May 31, 2021 is 25. In addition, the weeks between March 22, 2021 to April 12, 2021 contains total Twitter total sentiments numbers at magnitudes of tens of thousands, while all the other Twitter total sentiments numbers in the data set are at one hundred thousand magnitudes. Thus, the data set used in this robustness check removed all the weeks with abnormal Twitter total sentiments.

Table 10 represents the results of applying equations (1) and (2) on the newly edited US bitcoin data. After the abnormal periods are removed, the results still show the same trends, with similar coefficient magnitudes, in all the four models compared to the main results in table 3. Thus, this robustness check shows that the mains results are valid.

Lastly, linear time trend variables are added in this robustness check. This project uses OLS models, which have some underlying assumptions about the data used. Especially when it comes to the time series data, where various variables can have different trends and potentially causes some problems. In the main model section, when using OLS models, we only regress time series data on time series data to avoid similar problems. This linear time

Table 10: Regression results for Twitter sentiments on edited US Bitcoin data

	A Volume (in thousands)	B Volume (in thousands)	C Volatility	D Volatility
Total number of tweets	0.120	0.222**	0.234*	0.354***
(in thousands)	(0.083)	(0.088)	(0.126)	(0.126)
Ratio of Positive Tweets	-9.041***	-6.233**	-0.009***	-0.133***
	(1.810)	(2.745)	(0.003)	(0.004)
Ratio of Negative Tweets	39.682***	32.228***	0.021	0.019
	(11.138)	(11.069)	(0.017)	(0.016)
S&P 500		-5.401***		-8.69***
		(1.358)		(2.006)
3-month Bills		-100.050***		-300.300***
		(34.546)		(51.120)
10-year Bonds		96.370***		296.100***
		(35.595)		(52.770)
R-squared	0.247	0.340	0.105	0.299
Adjusted R-squared	0.231	0.311	0.086	0.269
No. observations	146	146	146	146

Notes: This table presents the relationship between Twitter sentiments and the edited US Bitcoin trading volume, and Bitcoin volatility. Column A and B present the regression coefficients from equation (1), and column C and D present the regression coefficients from equation (2). Column A and C do not include any controls, while column B and D include full sets of control variables. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

trends check will check whether it is the case.

Table 11: Regression results for Twitter sentiments on US Bitcoin data with time trends

	A Volume (in thousands)	B Volume (in thousands)	C Volatility	D Volatility
Total number of tweets	211***	0.303***	0.229**	0.344***
(in thousands)	(0.073)	(0.076)	(0.110)	(0.110)
Ratio of Positive Tweets	-6.546**	-6.887**	-0.014***	-0.101**
	(2.514)	(2.756)	(0.004)	(0.004)
Ratio of Negative Tweets	18.969**	15.077*	0.0167	0.010
	(8.128)	(8.171)	(0.012)	(0.012)
S&P 500		-6.439***		-7.641***
		(1.519)		(2.212)
3-month Bills		-73.584		-349.300***
		(51.710)		(75.310)
10-year Bonds		62.908		362.300***
		(54.748)		(79.740)
Linear Time Trend	-0.353	0.919	1.006	-1.648
	(0.364)	(0.988)	(0.550)	(1.438)
R-squared	0.202	0.314	0.134	0.308
Adjusted R-squared	0.180	0.280	0.111	0.275
No. observations	153	153	153	153

Notes: This table presents the relationship between Twitter sentiments and the US Bitcoin trading volume, and Bitcoin volatility. Column A and B present the regression coefficients from equation (1), and column C and D present the regression coefficients from equation (2). Column A and C do not include any controls, while column B and D include full sets of control variables. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Table 11 represents the results after adding linear time trends variables to equations (1) and (2) on the US bitcoin data. The results demonstrate that the linear time trend variables added in all four models come out to be insignificant. That is a good signal because it means that the time trend variables are not impacting the regression results, which also proves that the results in the main results section is valid.

There are some limitations in this project. For example, the trading volume data used in the project might not be complete. For example, in the data, certain trading volumes may be missing due to "dark pool". These are trading venues for anonymously-trading cryptocurrencies. Exchanges such as Kraken have offered dark pools for cryptocurrency trading. Also, Singapore-based Republic Protocol launched the first decentralized platform for dark pool trading in 2018[15]. Large organizations or institutional investors can trade huge volumes of coins, anonymously and discreetly, with the help of dark pools or other methods available to them. An estimated 15% of all trading volume in the American market takes place in dark pools, with some estimates putting the volumes as high as 40%[16]. On the other side, many exchanges are being accused for conducting wash trading practices in order to inflate trading volume and they are incentivized to report inflated volumes in order to attract traders. This can also bring uncertainty to the trading volume data. The same logic is also valid for the global trading volume data used in the project.

In terms of the global trading volume data, there are also some drawbacks. The major problem is that the magnitude of the data is enormous and the more detailed descriptions of the data, such as the proportional contribution from individuals, corporations, and governments are missing. That is the reason why the results can potentially be inaccurate or some trends can not be noticed in the first place. Also, after the trends are found, we are not able to further analyse the mechanisms behind it due the lack of understanding of these huge data sets.

For future work, we could focus on gathering more detailed and comprehensive data sets. In this way, it is then possible to test more specific associations between certain data groups and study further on the mechanisms behind them. Another important step would be to use even more reliable data sources, such as CoinMarketCap and Coinbase Pro. For the wash

[15] Extracted from: Investopedia.com

[16] Extracted from: https://ambcrypto.com/is-there-a-bright-side-to-dark-pools-in-Bitcoin-trading/: :text=Large%20organizations%20or%20institutional%20investors,volumes%20as%20high%20as%2040%25.

trading practices mentioned above, we could ”penalize” data from exchanges that engage in conducting wash trading activities. Additionally, future work could explore more types of cryptocurrency, such as XRP, EOS, and Litecoin.

7 Conclusions

To summarize, this project asks the question about whether social media sentiments relate to the trading volume and volatility of cryptocurrency. To investigate this question, this project brought together sentiment analysis and regression models. We used data on trading volume from the US and from all over the world, Twitter sentiments, GitHub activities, volatility, stock index, and bond interest rates. We showed that Twitter sentiments and GitHub activities are significantly associated with trading volume and volatility of Bitcoin and Ether, both globally and in the United States. More specifically, Twitter total discussion volume has is positively related to both Bitcoin trading volume and volatility. Moreover, Twitter positive sentiment tends to decrease the trading volume, and Twitter negative sentiment tends to increase the trading volume. Additionally, when GitHub activities increase, the trading volume and volatility in cryptocurrency tend to decrease.

This article also mentioned the existing limitations on the data accuracy, data structures, and data magnitude related to crypto markets. Limitations exist in not having enough understanding in some of the mechanisms and the participants in the cryptocurrency market. Future work can focus on testing the results using data from different sources, expanding the domain to include more types of cryptocurrency, or providing more designs to study these mechanisms.

References

Abraham, Jethin, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. "Cryptocurrency price prediction using tweet volumes and sentiment analysis." *SMU Data Science Review*, 1(3): 1.

Babkin Alexander, V, D Burkaltseva Diana, W Pshenichnikov Wladislav, and S Tyulin Andrei. 2017. "Cryptocurrency and blockchain-technology in digital economy: development genesis." π*-Economy*, 67(5): 9–22.

Bollen, Johan, and Huina Mao. 2011. "Twitter mood as a stock market predictor." *Computer*, 44(10): 91–94.

Chen, Cathy Yi-Hsuan, and Christian M Hafner. 2019. "Sentiment-induced bubbles in the cryptocurrency market." *Journal of Risk and Financial Management*, 12(2): 53.

Chohan, Usman W. 2017. "Cryptocurrencies: A brief thematic review." *Available at SSRN 3024330*.

Dimitrova, V, M Fernández-Martínez, MA Sánchez-Granero, and JE Trinidad Segovia. 2019. "Some comments on Bitcoin market (in) efficiency." *PloS one*, 14(7): e0219243.

Gil-Alana, Luis Alberiko, Emmanuel Joel Aikins Abakah, and María Fátima Romero Rojo. 2020. "Cryptocurrencies and stock market indices. Are they related?" *Research in International Business and Finance*, 51: 101063.

Hassan, M Kabir, Md Bokhtiar Hasan, Zairihan Abdul Halim, Neal Maroney, and Md Mamunur Rashid. 2022. "Exploring the dynamic spillover of cryptocurrency environmental attention across the commodities, green bonds, and environment-related stocks." *The North American Journal of Economics and Finance*, 101700.

Heimbach, I, and O Hinz. 2012. "How smartphone apps can help predicting music sales."

Hughes, Scott D. 2017. "Cryptocurrency Regulations and Enforcement in the US." *W. St. UL Rev.*, 45: 1.

Kraaijeveld, Olivier, and Johannes De Smedt. 2020. "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices." *Journal of International Financial Markets, Institutions and Money*, 65: 101188.

Krafft, Peter M, Nicolás Della Penna, and Alex Sandy Pentland. 2018. "An experimental study of cryptocurrency market dynamics." 1–13.

Molina, Brett. 2017. "Twitter overcounted active users since 2014, shares surge on profit hopes." *USA Today*.

Nofer, Michael, and Oliver Hinz. 2015. "Using twitter to predict the stock market." *Business & Information Systems Engineering*, 57(4): 229–242.

Omane-Adjepong, Maurice, Paul Alagidede, and Nana Kwame Akosah. 2019. "Wavelet time-scale persistence analysis of cryptocurrency market returns and volatility." *Physica A: Statistical Mechanics and its Applications*, 514: 105–120.

Rao, Tushar, and Saket Srivastava. 2013. "Modeling movements in oil, gold, forex and market indices using search volume index and twitter sentiments." 336–345.

Stylianou, Konstantinos, Leonhard Spiegelberg, Maurice Herlihy, and Nic Carter. 2021. "Cryptocurrency Competition and Market Concentration in the Presence of Network Effects." *Ledger*, 6: 81–101.

Treiblmaier, Horst. 2018. "The impact of the blockchain on the supply chain: a theory-based research framework and a call for action." *Supply chain management: an international journal*.

Wilson, Claire. 2019. "Cryptocurrencies: The Future of Finance?" In *Contemporary Issues in International Political Economy*. 359–394. Springer.

Zhu, Feng, and Xiaoquan Zhang. 2010. "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics." *Journal of marketing*, 74(2): 133–148.