UNIVERSITY OF COPENHAGEN

MSc in Social Data Science

Yanzhen Luo

Analyzing the Relationship between Cryptocurrencies Volatility and Social Media Activities

Date: 31 May 2023

Advisor: Nikolaj Arpe Harmon, Lau Lilleholt Harpviken

Institute:	Copenhagen Center for Social Data Science
Department	Department for Social Science
Author(s):	Yanzhen Luo
Title and subtitle:	Analyzing the Relationship between Cryptocurrencies Volatility and Social Media Activities
Abstract:	This study investigates the relationship between cryptocurrency volatility and social media activities, and the applicability of prospect theory. Analyzing financial and social media data of the top ten cryptocurrencies from November 2022 to March 2023, using sentiment and regression analysis, this research explore the association between Return, Volatility, Gain, and Loss variables and social media engagement. Results show a positive correlation between Volatility and social media active frequency, indicating increased engagement during higher volatility. Moreover, this association diminishes as the time window narrows. However, no significant correlation is found between volatility and sentiment. Surprisingly, a positive correlation is observed between user activity and financial gains, suggesting risk-loving behavior among cryptocurrency supporters. Further research is needed to validate prospect theory's applicability and examine this behavior in real-world scenarios. These findings contribute to understanding the interplay between cryptocurrencies and social media activities, emphasizing volatility's role in driving online discussions and the response to financial gains. Future research should consider additional factors and longer time periods for a more comprehensive analysis.
Advisor:	Nikolaj Arpe Harmon, Lau Lilleholt Harpviken
Date:	31 05 2023

Table of content

1 Introduction

In recent years, blockchain technology has gained increasing attention, and one of its most well-known applications is in the realm of cryptocurrencies, which have become popular among individual investors and financial institutions. However, cryptocurrencies have also faced concerns and criticism due to their potential risks, particularly the extreme fluctuations in their market prices. In traditional financial markets, such severe volatility often triggers panic among investors, leading to sell-offs and potential market crashes. Despite these concerns and some incidents that have indeed occurred, the cryptocurrency market remains active, with numerous cryptocurrencies boasting substantial market capitalization and increasing trading volumes. These contradictions with traditional markets emphasize the importance of understanding people's psychology and behavior in the cryptocurrency market. One potential way for understanding is through analyzing the expressions and sentiments of individuals in the cryptocurrency community on social media platforms when facing market volatility of cryptocurrencies. Thus, this study aims to explore the relationship between cryptocurrencies volatility and social media activities.

While industry and academia have shown growing interest in blockchain and cryptocurrencies, existing research predominantly focuses on sentiment analysis of social media data as an indicator for predicting cryptocurrency prices. However, limited attention has been given to understanding the psychology and behavior of individuals within the cryptocurrency community, leaving gaps in our theoretical understanding of these phenomena.

To address this research gap, this study aims to help gain a deeper understanding of the psychology and behavior of people in the cryptocurrency market in the face of market volatility. By analyzing the market volatility of ten cryptocurrencies and the corresponding discussions on the social media platform Reddit from November 2022 to March 2023, employing sentiment analysis and regression analysis, this study aims to investigate the relationship between user activity on social media and cryptocurrency market volatility. Through this investigation, the study aims to reveal insights into the psychology and behavior of cryptocurrency users.

Overall, this research strives to advance our understanding of the psychology and behavior exhibited by people in the cryptocurrency market by examining the relationship between cryptocurrency market volatility and users' social media activity. In addition, it seeks to make potential contributions to the analysis of social media data and explores the application of social data science methods to the study of human psychology and behavior in specific contexts. Furthermore, the study aims to promote theoretical innovation by testing whether economic theories such as prospect theory can explain people's social media activities in the evolving cryptocurrency market. This leads to the research question: How is users' social media activity related to the volatility of cryptocurrencies?

For a focused investigation, the following research questions were defined:

1. What is the relationship between the volatility of cryptocurrencies and the frequency of tweets on social media?

2. How does the general sentiment expressed on social media correlate with the volatility of cryptocurrencies?

The remainder of this paper is structured as follows. Section 2 provides a concise review of the literature, exploring the current research about social media and cryptocurrency volatility. It also includes a review of prospect theory and sentiment analysis methods based on social media data. Section 3 presents an overview of the data used in this study and provides an explanation of the adopted methodology. Section 4 presents the key findings and results obtained from the analysis. Section 5 discusses the limitations of this paper and suggests potential improvements for further research. Finally, Section 6 concludes the paper, summarizing the main findings and implications.

2 Literature Review

Cryptocurrencies, with Bitcoin as the most prominent example, have emerged as a new and transformative asset class, captivating both speculators and technology-fluent investors (Lee et al., 2020). The concept of digital cash was introduced even before the advent of Bitcoin, with proposals like Wei Dai's B-money in 1998 (Wei, 1998) and Adam Back's Hashcash in 2002 (Back, 2002), which laid the foundation for decentralized cryptocurrencies. Bitcoin, introduced in 2008 by Satoshi Nakamoto, became the pioneering decentralized cryptocurrency, utilizing blockchain technology and decentralized ledgers (Nakamoto, 2008). Its market capitalization has exceeded 500 billion by May 2023 (CoinMarketCap, 2023), showcasing its immense success.

The cryptocurrency landscape has rapidly expanded, with CoinMarketCap listing over 24,864 cryptocurrencies, totaling more than 1.122 trillion dollars as of May 27, 2023 (CoinMarketCap, 2023). However, due to their notorious price volatility, cryptocurrencies are often considered more as speculative assets rather than alternative currencies or mediums of exchange (Fry, 2018).

The rapid growth of the cryptocurrency market, coupled with its recognition as a speculative financial asset, has aroused great interest among industry professionals and academics. As a result, there has been a growing literature focused on understanding and analyzing the performance of cryptocurrencies in financial markets.

2.1 Cryptocurrency Market Volatility and Social Media

Several academic studies have investigated the association between social media activity and the volatility of cryptocurrencies, focusing on the correlation between cryptocurrency volatility and sentiment expressed on social media platforms. For instance, Nikolaos Kyriazis et al. explored the impact of social media sentiment on cryptocurrency returns and volatility during the COVID-19 pandemic, revealing a significant influence of Twitter-derived sentiment on the analyzed cryptocurrencies (Kyriazis et al., 2023). Similarly, M. Ángeles López-Cabarcos et al. investigated the relationship between sentiment expressed in Stocktwits messages and Bitcoin volatility in stable and unstable periods (López-Cabarcos et al., 2021), affirming the effect of social network sentiment on cryptocurrency market volatility in stable periods.

Additionally, some studies have focused on employing social media data to predict cryptocurrency prices. Mason McCoy and Shahram Rahimi, for example, utilized sentiment analysis of tweets to forecast cryptocurrency prices (McCoy & Rahimi, 2020). Similarly, Chahat Tandon et al. utilized Twitter data, specifically considering Elon Musk's tweets, in combination with an AutoRegressive Integrated Moving Average (ARIMA) model to predict Bitcoin's price (Tandon et al., 2021).

Existing studies primarily focus on utilizing social media data to analyze and forecast fluctuations in cryptocurrency prices. In contrast, this study places a stronger emphasis on understanding people' behavior and psychology within the cryptocurrency market by investigating the relationship between social media activities and cryptocurrency volatility. By adopting this approach, the aim is to gain valuable insights into the various factors that influence user engagement in the cryptocurrency space.

2.2 Prospect Theory and Application

In traditional economic and financial theory, such as the efficient market hypothesis and expected utility theory, investors are often assumed to be rational and consistent decision-makers (Fama, 1970). These theories propose that investors maximize their expected utility and base their decisions on objective information and probabilities. However, prospect theory challenges these assumptions by demonstrating that investor behavior is influenced by subjective factors, including expectations, sentiment, and cognitive biases (Kahneman & Tversky, 1979). Prospect theory emphasizes the role of loss aversion and nonlinear probability weighting, suggesting that individuals may exhibit different risk preferences when facing gains and losses.

However, it is important to note that prospect theory has primarily been studied in controlled laboratory settings, and its applicability outside of such environments has been a topic of debate (Barberis, 2013). Nonetheless, recent empirical studies have explored the correlation between sentiment and stock market dynamics, providing insights into the potential influence of psychological factors on financial markets (Goodell et al., 2023). In the cryptocurrency market, Rongxin Chen et al. found that investor behavior aligns with the predictions of prospect theory (Chen et al., 2022).

Given the increasing popularity and unique characteristics of evolving cryptocurrency markets, there is a research interest in understanding whether prospect theory can explain people's behavior in this context. By examining the relationship between social media activities and cryptocurrency volatility, this study aims to contribute to the understanding of how prospect theory would be applied to individuals behavioyr in cryptocurrency markets.

2.3 Social Media and Sentiment Analysis

Scholars have increasingly turned to social media data to study people's discussion in the realm of cryptocurrencies. As mentioned earlier, researchers like Mason McCoy, Chahat Tandon, have utilized sentiment analysis of tweets to predict cryptocurrency prices (McCoy & Rahimi, 2020; Tandon et al., 2021). Similarly, Stephen et al. employed sentiment analysis of Reddit discussions to observe price changes in the cryptocurrency market (Wooley et al., 2019). Those reveal that Twitter and Reddit are widely used social media platforms by scholars in researching cryptocurrency communities.

Furthermore, there are various methods for conducting sentiment analysis on social media data. Giachanou and Crestani's study provides a summary of Twitter sentiment analysis methods, including supervised machine learning models that rely on a training corpus with manually annotated tweet sentiment (Giachanou & Crestani, 2017). Additionally, researchers have made pre-trained models such as Google AI's Bidirectional Encoder Representations from Transformers (BERT), which leverages deep learning algorithms and emphasizes contextual semantics (Devlin et al., 2019). Other variants of the BERT algorithm, such as DistilBERT (Sanh et al., 2020) and Robustly Optimized BERT Pretraining Approach (RoBERTa) (Liu et al., 2019), have also been widely employed. There are pre-trained models based on emotional words, such as VADER (Hutto & Gilbert, 2014), which is quite useful and easy to interpret. These methods have provided valuable insights into people' emotional expressions and sentiment analysis.

In this study, the analysis focuses on data from Reddit. Reddit offers many subreddits dedicated to specific cryptocurrencies, making it easier and more focused to identify relevant discussions and users associated with specific cryptocurrencies.

This paper investigates the link between cryptocurrency market volatility and social media activity, specifically focusing on Reddit post frequency and general sentiment. Consistent with previous research and prospect theory, the hypothesis suggests a connection between social media activity and cryptocurrency volatility, highlighting the heightened sensitivity to losses among individuals in the cryptocurrency community.

3 Data and Methododology

3.1 Propose Method

The sample for this study consists of the top ten cryptocurrencies based on market capitalization, namely Bitcoin, Ethereum, Tether, BNB, USD Coin, XRP, Cardano, Dogecoin, Polygon, and Solana. However, due to relatively lower social media engagement, Tether was replaced with the 11th cryptocurrency in terms of market cap, which is TRON. The final sample includes these top ten cryptocurrencies, known for their financial performance and significant social media discussions.

To address the research question, financial data capturing the volatility of the cryptocurrency market and social media data are required. Following established methodologies in financial research (Campbell et al., 1998), the collection of close price data is necessary to calculate daily returns and volatility. These indicators will be examined to determine their potential influence on users' social media activities. Furthermore, considering the principles of prospect theory, the study aims to explore individuals' differential responses to gains and losses. For this purpose, the close price data will be utilized to create variables representing gain or loss as well. Yahoo Finance, a reputable online platform offering comprehensive financial information, including daily close prices for most cryptocurrencies, will serve as the primary source of financial data for this research.

Reddit, a social media platform with dedicated communities centered around specific interests, including a wide range of cryptocurrencies, offers a suitable social media data source for this study. Through its organized structure of subreddits, each major cryptocurrency possesses its own dedicated community where information and discussions related to the specific cryptocurrency take place (Phillips & Gorse, 2017). Consequently, Reddit serves as an ideal platform to capture the target audience and discussions pertaining to individual cryptocurrencies.

The methodology employed in this study involves sample selection, data collection, sentiment analysis, and regression analysis, as illustrated in Figure 1.

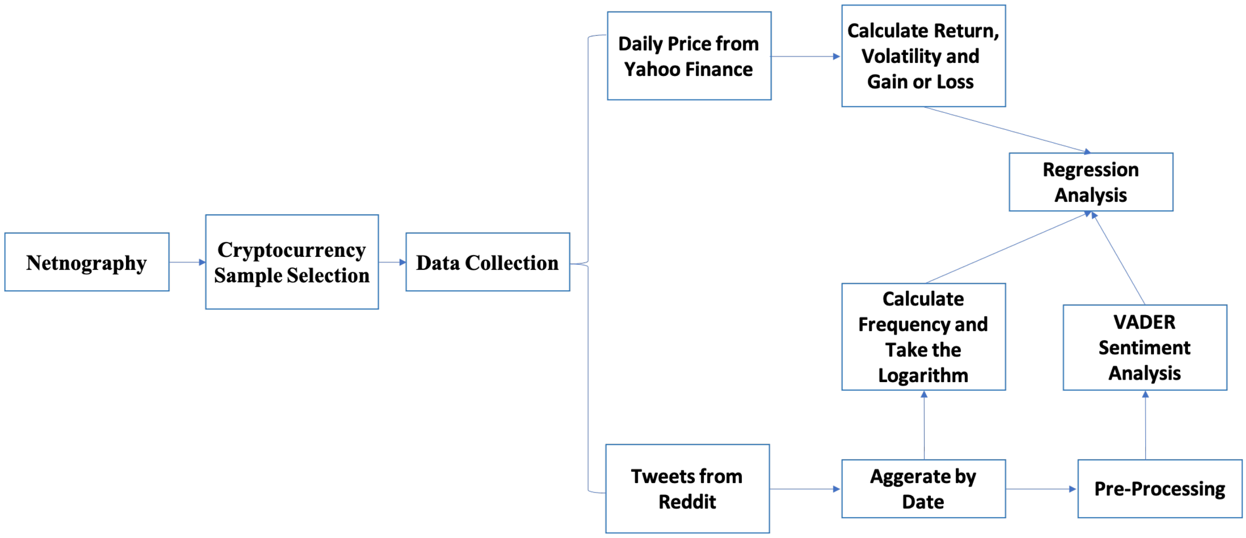


Figure 1: Propose Method

3.2 Financial Data and volatility indicators

The data collection started from Yahoo Finance's cryptocurrency market section, specifically focusing on the USD exchange rate (YahooFinance, 2023). The data encompasses a period of 141 days, from November 2, 2022, to March 23, 2023, covering the 10 chosen cryptocurrencies. Panel data analysis was performed using this dataset, which contains a total of 14,100 data points.

The daily return and period return are generated based on the close price, following definitions commonly used in the financial market (Phillips & Gorse, 2017). The simple net return, denoted as R_t, is calculated as the ratio of the close price at date t to the close price at date $t-1$, minus 1, which is presented as:

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

140 days of daily returns across 10 cryptocurrencies are obtained, as first day returns were not available (requires close price from previous day).

Returns for various time periods, including two-day return, three-day return, four-day return, five-day return, six-day return, and seven-day return, are also calculated based on established methods (Phillips & Gorse, 2017). The net return over the most recent k periods, denoted as $R_t(k)$, is computed as the ratio of the price at date t to the price at date $t-k$, minus 1, which is illustrated as:

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

This allows for the calculation of returns over different time periods. The return and related indicators are obtained, and their distributions are presented in Figure 2.

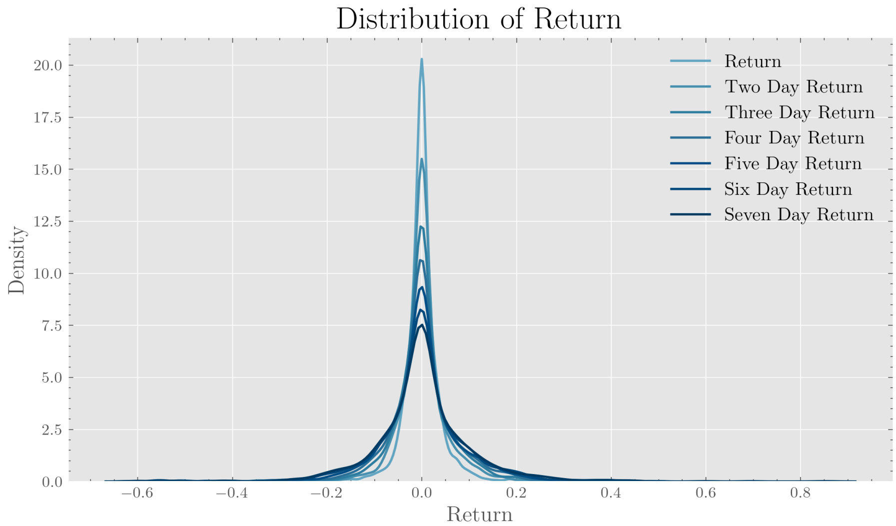


Figure 2: Distribution of Return

Figure 2 shows that the daily returns exhibit a distribution pattern resembling a normal distribution.

Moving volatility, representing the standard deviation of daily returns over a specified number of previous days, is calculated following standard financial analysis practices. This yields variables such as Three Day Volatility, Four Day Volatility, Five Day Volatility, Six Day Volatility, and Seven Day Volatility, which capture volatility measures over different time periods.

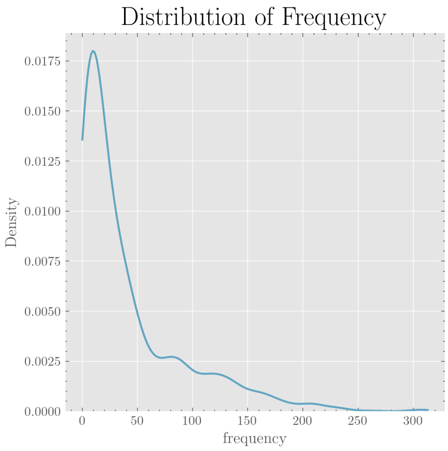
Additionally, the study examines individuals' sensitivity to gains and losses. Four variables are created: Gain, Loss, Three Day Gain, and Three Day Loss. These variables indicate whether there was a gain or loss on the current day and the previous three days. Gains and losses are quantified using a threshold approach, where returns exceeding two times the standard deviation are classified as gains, and returns below minus two times the standard deviation are categorized as losses. Returns within this range are considered neither gains nor losses.

3.3 Social Media Data and Activity indicators

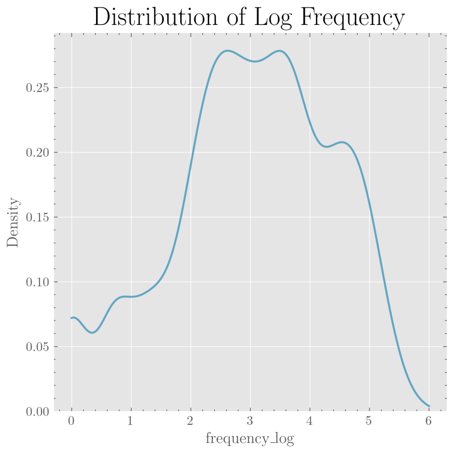
3.3.1 Active frequency

The research collected a total of 59,351 tweets from the ten subreddits of the cryptocurrencies using a Reddit API (Pushshift) from November 3, 2022, to March 23, 2023, in total

140 days. The tweets were aggregated by date for each cryptocurrency, providing the daily frequency of tweets. To address skewness and normalize the frequency data, a natural logarithmic transformation was applied. The distributions of the frequency data before and after the logarithmic transformation can be seen in Figure 3.



(a) Distribution of Frequency



(b) Distribution of Log Frequency

Figure 3: Logarithmic Transformation of Frequency

3.3.2 Sentiment Analysis

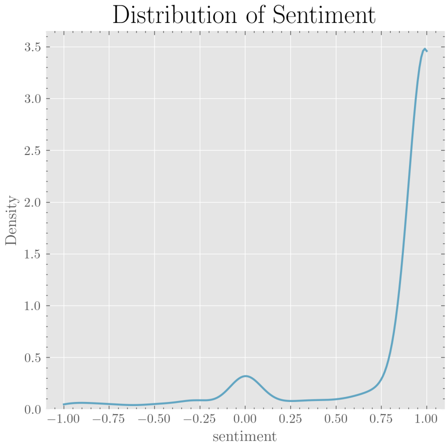
In addition to activity frequency, the study investigated the relationship between cryptocurrency market volatility and users' sentiment. Sentiment analysis was conducted on the aggregated tweets using the Valence Aware Dictionary and Sentiment Reasoner (VADER). Pre-processing steps were implemented, including the removal of mentions, links, and stop words.

The Sentiment Intensity Analyzer (SIA), a component of VADER, was employed to classify the text data for each cryptocurrency on a daily basis, generating sentiment compound scores. These scores were integrated into a result set, transformed into a data frame, and merged with the financial data set.

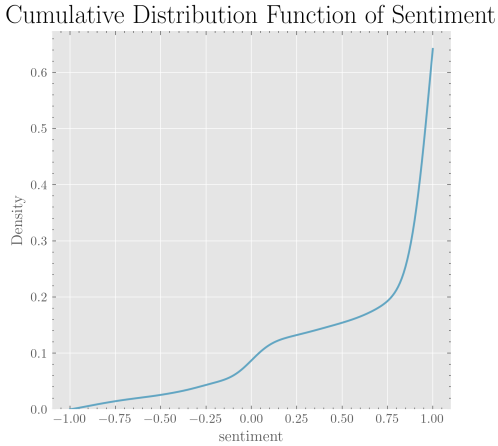
VADER is a lexicon and rule-based sentiment analysis tool specifically designed for analyzing sentiment in social media text (Hutto & Gilbert, 2014). It utilizes a word-based language model and predefined rules to calculate sentiment scores ranging from -1 to 1, reflecting negative, neutral, and positive sentiment, respectively (Hutto & Gilbert, 2014).

The chosen scoring range aligns well with the research objectives, helping to interpret and capture overall sentiment for each day. Additionally, aggregation and pre-processing steps naturally mitigate the contextual information in the content, making context-based deep learning models such as BERT less feasible in this study than word-based methods such as VADER.

The Distribution and Cumulative Distribution Function of the final sentiment scores are presented in Figure 4.



(a) Distribution of Sentiment



(b) CDF of Sentiment

Figure 4: Distribution of Sentiment Score by VADER

3.4 Regression Analysis

The summary statistics for the main collected data and newly generated variables are provided in Table 1.

Table 1: Summary Statistics

	Observations	Yes	No	Percemtage	
Current Day Gain	1400	36	1364	0.026	
Current Day Loss	1400	28	1372	0.020	
	Observations	Mean	SD	Min	Max
Current Day Return	1400	0.001	0.044	-0.423	0.383
Three Day Volatility	1400	0.028	0.035	0.000	0.351
frequency	1400	41.533	49.580	0.000	313.000
log frequency	1400	3.019	1.352	0.000	5.749
sentiment	1400	0.795	0.436	-1.000	1.000

Note: The table shows the main summary statistics for the variables in the dataset, omitting the variables like Two Day Return, Four Day Volatility, etc. See the companion paper for complete summary statistics.

Besed on the collected data, regression analysis was conducted to investigate the relationship between market volatility and users' social media activities, specifically frequency and sentiment. The dependent variables in the regression were log frequency and sentiment, while the independent variables included return, volatility, gain, and loss.

Equations 1 and 2 were used to examine the correlation between return and log frequency, and return and sentiment, respectively. These equations (and others listed below) accounted for cryptocurrency and time fixed effects.

$$log(Frequency) = \beta_0 + \beta_1 \cdot Return + \sum_{l=1}^{9} \gamma_l \cdot Cryptocurrency_l + \sum_{t=1}^{T} \lambda_t \cdot Time_t + \varepsilon \quad (1)$$

$$Sentiment = \beta_0 + \beta_1 \cdot Return + \sum_{l=1}^{9} \gamma_l \cdot Cryptocurrency_l + \sum_{t=1}^{T} \lambda_t \cdot Time_t + \varepsilon \quad (2)$$

Equations 3 and 4 were utilized to examine the relationship between volatility and log frequency, as well as sentiment, respectively.

$$log(frequency) = \beta_0 + \beta_1 \cdot Volatility + \sum_{l=1}^{9} \gamma_l \cdot Cryptocurrency_l + \sum_{t=1}^{T} \lambda_t \cdot Time_t + \varepsilon \quad (3)$$

$$sentiment = \beta_0 + \beta_1 \cdot Volatility + \sum_{l=1}^{9} \gamma_l \cdot Cryptocurrency_l + \sum_{t=1}^{T} \lambda_t \cdot Time_t + \varepsilon \quad (4)$$

Furthermore, Equations 5 and 6 were employed to investigate the association between gain or loss and log frequency, as well as sentiment.

$$log(frequency) = \beta_0 + \beta_1 \cdot Gain + \beta_2 \cdot Loss + \sum_{l=1}^{9} \gamma_l \cdot Cryptocurrency_l + \sum_{t=1}^{T} \lambda_t \cdot Time_t + \varepsilon \quad (5)$$

$$sentiment = \beta_0 + \beta_1 \cdot Gain + \beta_2 \cdot Loss + \sum_{l=1}^{9} \gamma_l \cdot Cryptocurrency_l + \sum_{t=1}^{T} \lambda_t \cdot Time_t + \varepsilon \quad (6)$$

The coefficients in these regression equations provide insights into the correlation between daily return, market volatility, gain or loss, and users' social media activities and sentiment in the cryptocurrency market. The analyses shed light on the relationship between market volatility and users' psychology and behavior, as well as the sensitivity to gains and losses in the cryptocurrency domain. See the next Section for the regression results.

4 Results

4.1 Return, Active frequency and Sentiment

The regression results for the independent variables daily return and period returns of Eqation 1 and 2 are presented in Table 2.

Table 2: Return, Social Media Active Frequency among Cryptocurrencies

	Dependent variable: Log frequency Index			*Dependent variable: Sentiment*		
	(1)	(2)	(3)	(4)	(5)	(6)
Current Day Return	-0.176	-0.239	0.004	0.080	0.056	0.073
	(0.692)	(0.430)	(0.405)	(0.313)	(0.297)	(0.310)
Two Day Return	0.147	0.026	0.364	0.122	0.090	0.138
	(0.521)	(0.311)	(0.296)	(0.151)	(0.134)	(0.136)
Three Day Return	0.047	-0.134	0.321	0.105	0.070	0.112
	(0.464)	(0.281)	(0.273)	(0.136)	(0.123)	(0.123)
Four Day Return	0.099	-0.104	0.302	0.153	0.119	0.138
	(0.408)	(0.236)	(0.232)	(0.114)	(0.101)	(0.108)
Five Day Return	0.078	0.135	0.269	0.081	0.046	0.057
	(0.359)	(0.199)	(0.200)	(0.097)	(0.086)	(0.093)
Six Day Return	0.056	-0.185	0.218	0.035	-0.006	0.003
	(0.330)	(0.181)	(0.186)	(0.091)	(0.082)	(0.093)
Seven Day Return	0.141	0.124	0.246	0.038	-0.008	-0.002
	(0.298)	(0.165)	(0.171)	(0.085)	(0.076)	(0.091)
Cryptocurrency fixed effects	No	Yes	Yes	No	Yes	Yes
Weekday fixed effects	No	No	Yes	No	No	Yes
Month fixed effects	No	No	Yes	No	No	Yes

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

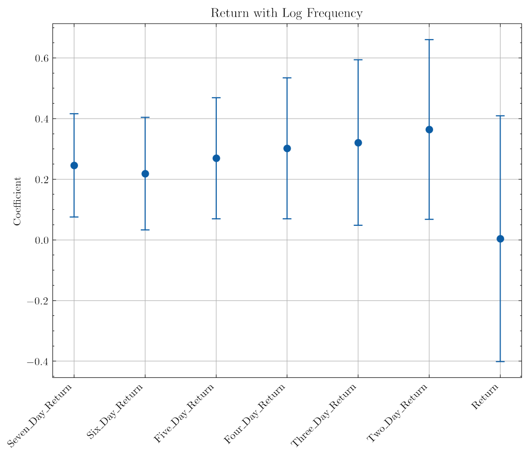
Variables such as Two Day Return and Three Day Return represent moving average returns over a specified number of days.

The regression analysis consisted of two sets examining the relationship between the logarithm of frequency and sentiment scores with each Return variable. Each set comprised three subgroups, with each subgroup assessing the relationship of a specific Return variable on the dependent variable. The regressions controlled for cryptocurrency fixed effects and time effects (weekday fixed effect and month fixed effect) sequentially.

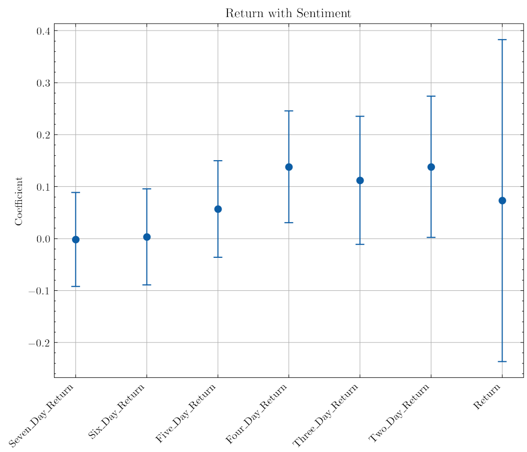
The analysis findings indicate no significant association between the return indicators (current day return, two day return, three day return, four day return, five day return, six day return, and seven day return) and the log frequency index or sentiment in cryptocurrencies. Even after controlling for cryptocurrency fixed effects, weekday fixed effects, and month fixed effects, the coefficients were not statistically significant, suggesting that any observed associations may be due to chance rather than genuine relationships.

In summary, the study's analysis did not find significant evidence supporting a relationship between return, social media active frequency (log frequency index), and sentiment in the selected cryptocurrencies.

The coefficients of all Return-related indicators, controlling for cryptocurrency fixed effects and time fixed effects, are shown in Figure 5.



(a) Coefficients of Returns (Log Frequency)



(b) Coefficients of Returns (Sentiments)

Figure 5: Coefficients of Returns

4.2 Volatility, Active frequency and Sentiment

Although no significant evidence was found to support the relationship between Return and active frequency and sentiment, it is still important to investigate the potential relationship between volatility and active frequency and sentiment on social media. Table 3 presents the regression results of the independent variables, namely different volatility indicators, based on Equation 3 and Equation 4.

Table 3: Volatility, Social Media Active Frequency and Sentiment among Cryptocurrencies

	Dependent variable: Log frequency Index			*Dependent variable: Sentiment*		
	(1)	(2)	(3)	(4)	(5)	(6)
Three Day Volatility	5.732***	2.845***	2.119***	1.465***	-0.132	-0.166
	(1.047)	(0.483)	(0.513)	(0.492)	(0.464)	(0.478)
Four Day Volatility	6.577***	3.262***	2.575***	1.872***	0.036	0.018
	(1.098)	(0.515)	(0.554)	(0.489)	(0.485)	0.500
Five Day Volatility	7.171***	3.615***	2.786***	2.117***	0.132	0.035
	(1.125)	(0.538)	(0.591)	(0.463)	(0.469)	(0.488)
Six Day Volatility	7.596***	3.879***	2.839***	2.302***	0.223	0.055
	(1.150)	(0.563)	(0.654)	(0.442)	(0.447)	(0.469)
Seven Day Volatility	7.811***	3.944***	2.737***	2.530***	0.390	0.162
	(1.170)	(0.573)	(0.688)	(0.457)	(0.466)	(0.498)
Cryptocurrency fixed effects	No	Yes	Yes	No	Yes	Yes
Weekday fixed effects	No	No	Yes	No	No	Yes
Month fixed effects	No	No	Yes	No	No	Yes

Note: *p<0.1; **p<0.05; ***p<0.01

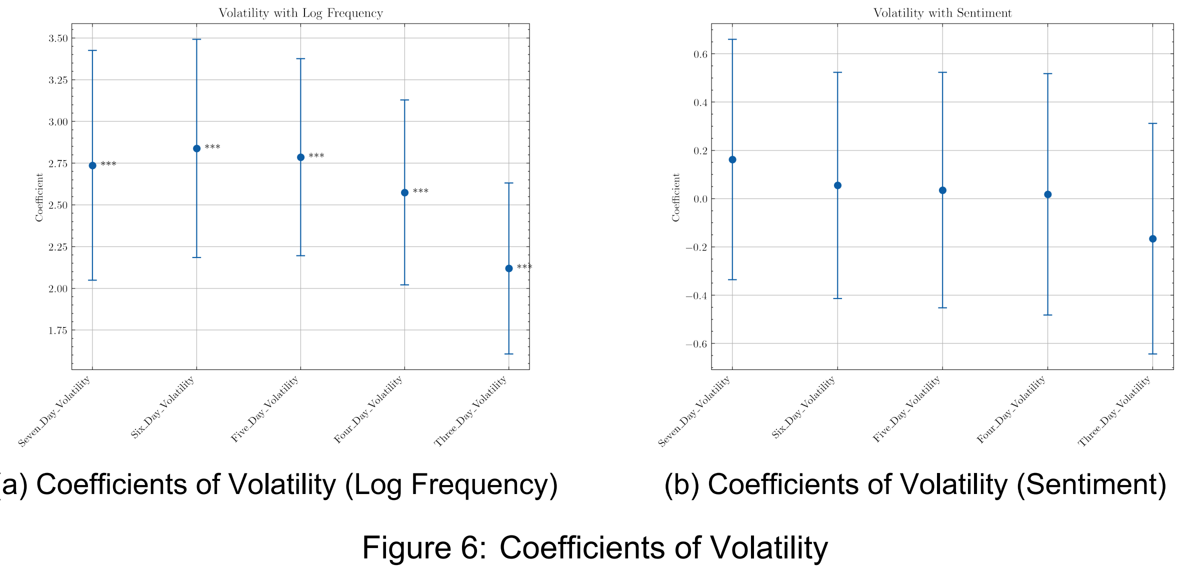
Variables such as Three Day Volatility and Four Day Volatility represent moving average volatilities over a specified number of days.

The above regression analysis revealed two sets of regressions analyzing the relationship between the logarithm of frequency and sentiment with each Volatility variable. Each set consisted of three subgroups, with each subgroup contains five regressions of each Volatility variable on the dependent variable. The regressions controlled for cryptocurrency fixed effects and time effects (weekday fixed effect and month fixed effect) in a sequential manner.

Overall, the results indicated a significant positive relationship between volatility and both the log frequency index and sentiment in the selected cryptocurrencies. However, after controlling for cryptocurrency fixed effects and time effects, the relationship between volatility and sentiment became statistically non-significant.

For instance, considering the independent variable ”Three Day Volatility,” the estimated coefficient was 2.119, which was statistically significant at a high level of confidence. This suggested a strong positive relationship between cryptocurrency volatility, measured by the moving standard deviation of their three-day returns, and the log frequency index. Specifically, for every unit increase in cryptocurrency volatility, the log frequency index was expected to increase by 2.119 units. This indicated that as cryptocurrency volatility rises, there is a corresponding increase in social media activity captured by the log frequency index. In simpler terms, higher volatility in cryptocurrency returns tends to lead to more social media activity, with individuals posting more tweets about cryptocurrencies.

The coefficients of all Volatility-related indicators, controlling for cryptocurrency fixed effects and time fixed effects, are shown in Figure 6.



(a) Coefficients of Volatility (Log Frequency)

(b) Coefficients of Volatility (Sentiment)

Figure 6: Coefficients of Volatility

Notably, the coefficients associated with the different volatility indicators demonstrate a gradual decrease as the time period decreases. This diminishing effect suggests that the link between volatility and social media active frequency diminishes when considering narrower time windows. Specifically, recent volatility within the past six days has a stronger connection on social media activity compared to volatility observed within the past three days. People on social media platforms exhibit higher engagement and responsiveness to discussions during periods of increased volatility in the cryptocurrency market over a longer time horizon, like four to seven days.

In summary, the analysis findings suggest a positive relationship between higher levels of volatility in the cryptocurrency market and an increased frequency of discussions on social media. Moreover, this association diminishes as the time window narrows. These findings underscore the importance of considering the temporal aspect of volatility when studying its relationship with social media activity. Overall, the study provides significant evidence supporting a positive relationship between volatility and the log frequency index for the selected cryptocurrencies.

4.3 Gain or Loss, Active frequency and Sentiment

While these findings imply that higher levels of volatility in the cryptocurrency market are associated with an increased frequency of discussions, they do not directly speak to the principles of prospect theory. Based on Equation 5 and Equation 6, further research is needed to explore how people respond differently to gains and losses, which is showed in Table 4.

Table 4: The correlation between frequency, sentiment and gain

	Dependent variable: Log frequency Index				*Dependent variable: Sentiment*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Current Day Gain	0.418^{*}	0.292^{*}			0.053	9.015		
	(0.215)	(0.154)			(0.071)	(0.073)		
Current Day Loss	0.455^{*}	0.071			-0.057	-0.115		
	(0.265)	(0.202)			(0.102)	(0.080)		
Three Day Gain			0.515^{***}	0.162^{**}			0.065	-0.016
			(0.129)	(0.071)			(0.044)	(0.041)
Three Day Loss			0.601^{***}	0.144			0.036	-0.039
			(0.155)	(0.102)			(0.059)	(0.054)
Cryptocurrency fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Weekday fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Month fixed effects	No	Yes	No	Yes	No	Yes	No	Yes

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Variables, Three Day Gain and Three Day Loss, represent whether the moving three day return is more than two standard deviations or less than negative two standard deviations.

The analysis showed a significant positive relationship between certain gain variables (current day gain, three day gain) and the log frequency index among cryptocurrencies. However, the association between gain variables and sentiment was less pronounced. On the other hand, no significant relationship was found between loss and both the log frequency index and sentiment.

In summary, the findings provide evidence supporting a positive relationship between gain variables (current day gain, three-day gain) and the log frequency index among the selected cryptocurrencies. This suggests that cryptocurrency users may exhibit a characteristic of being more sensitive to gains, as reflected in their increased social media activity.

These findings appear to contradict one of the principles from prospect theory, which suggests that people are generally more sensitive to losses than gains. According to prospect theory, individuals are expected to be more emotionally affected by losses and, therefore, may exhibit stronger responses in terms of social media activity and sentiment. However, in this particular study, the results indicate a stronger association between gains and social media activity rather than losses.

Additionally, the study found no significant relationships between the sentiments of cryptocurrency users and gain or loss variables. This implies that users may not demonstrate emotional sensitivity to short-term market volatility, potentially indicating a preference for long-term investments over immediate trading opportunities.

It is worth noting that these results appear to deviate from the predictions of prospect theory. Further research is necessary to explore and understand the underlying factors contributing to these observed relationships.

5 Discussions

Despite the initial findings of this research, several areas warrant further consideration and analysis to strengthen the study's conclusions.

Firstly, it is important to address data ethics. The research strictly relies on publicly available social media data and does not involve the collection of personal information. Additionally, data aggregation were employed to ensure compliance with data ethics requirements and considerations.

One limitation of the study pertains to data availability and the external validity of the findings. The current dataset focuses on the top ten cryptocurrencies and their corresponding communities based on market capitalization and social media activity. However, it is crucial to recognize that cryptocurrencies with lower attention and smaller market values have received limited attention in the research. While the inclusion of the ten cryptocurrencies in this study is reasonable given their larger volume and social media participation (accounting for 78% of the total cryptocurrency market capitalization by May 2023), expanding the research scope to include additional cryptocurrencies and longer time periods would enhance the generalizability of the findings. By increasing the data volume, a more comprehensive examination can be conducted, ensuring that research conclusions are not confined to specific objectives or time periods.

Additionally, the relatively short time span of the study poses another limitation. Significant events occurring during this period, such as the FTX bankruptcy case in the beginning of November, may have influenced the research results. To address this concern, extending the time frame and incorporating more data points are essential. This would be possible to control for event-driven impacts and bolster the reliability of the research conclusions.

In summary, while the research has provided initial insights, it is important to acknowledge the need for further analysis and consideration in various areas. Adhering to data ethics and expanding the data collection to encompass a broader range of cryptocurrencies and longer time periods to enhance the reliability and comprehensiveness of the research findings. By addressing these aspects, the research can provide a more robust understanding of relationship between cryptocurrencies financial performance and social media community activities.

6 Conclusion

This study aimed to examine the relationship between cryptocurrency market volatility and user behavior and psychology on social media. The research investigated the connection between Return, Volatility, Gain, and Loss variables and individuals' social media activities, as well as the applicability of prospect theory in this context. By analyzing financial and social media data of the top ten cryptocurrencies from November 2022 to March 2023 using sentiment analysis and regression analysis, the study obtained several key findings.

The results revealed a significant positive correlation between Volatility and the frequency of social media posts, indicating that as volatility increased, social media users became more active. Moreover, the association presents a gradual decrease as the time period decreases. However, no significant correlation was found between volatility and the sentiment expressed in these discussions. Surprisingly, the study uncovered a significant positive correlation between user activity (log frequency) and financial gains, but not with losses. This suggests that cryptocurrency supporters may exhibit risk-loving behavior and respond more strongly to financial gains. Further research is needed to explore this group's behavior and the applicability of prospect theory in real-world scenarios, requiring additional empirical evidence.

While this study demonstrated a significant positive correlation between log frequency and market volatility, as well as a positive correlation with financial gains. It is important to note that further research is necessary to gain a comprehensive understanding of the relationship between cryptocurrency market volatility and people's psychology and behaviours in social media. The findings should be interpreted within the specified study period and should not be extrapolated to other time periods. To delve deeper into this subject, it is crucial to consider the influence of other factors and explore the impact across different time periods.

In conclusion, this study contributes to our understanding of the interplay between cryptocurrencies volatility and social media activities. The findings emphasize the significant positive relationship between volatility and social media discussions and highlight the heightened response of users to financial gains. However, further investigations are required to validate and expand upon these findings, incorporating additional factors and extending the analysis to longer time periods. By addressing these areas, researchers can gain a more comprehensive understanding of the relationship between cryptocurrencies, social media, and people's behaviour and psychology.

References

Back, A. (2002). Hashcash - a denial of service counter-measure.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, *27*(1), 173–196. https://doi.org/10.1257/jep.27.1.173

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., & Whitelaw, R. F. (1998). THE ECONOMETRICS OF FINANCIAL MARKETS [Publisher: Cambridge University Press]. *Macroeconomic Dynamics*, *2*(4), 559–562. https://doi.org/10.1017/S1365100598009092

Chen, R., Lepori, G. M., Tai, C.-C., & Sung, M.-C. (2022). Explaining cryptocurrency returns: A prospect theory perspective. *Journal of International Financial Markets, Institutions and Money*, *79*, 101599. https://doi.org/10.1016/j.intfin.2022.101599

CoinMarketCap. (2023, May 27). *Cryptocurrency prices, charts and market capitalizations* [CoinMarketCap]. Retrieved May 27, 2023, from https://coinmarketcap.com/

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May 24). BERT: Pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/arXiv.1810.04805

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work [Publisher: [American Finance Association, Wiley]]. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Fry, J. (2018). Booms, busts and heavy-tails: The story of bitcoin and cryptocurrency markets? *Economics Letters*, *171*, 225–229. https://doi.org/10.1016/j.econlet.2018.08.008

Giachanou, A., & Crestani, F. (2017). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys*, *49*(2), 1–41. https://doi.org/10.1145/2938640

Goodell, J. W., Kumar, S., Rao, P., & Verma, S. (2023). Emotions and stock market anomalies: A systematic review. *Journal of Behavioral and Experimental Finance*, *37*, 100722. https://doi.org/10.1016/j.jbef.2022.100722

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text [Number: 1]. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk [Publisher: [Wiley, Econometric Society]]. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kyriazis, N., Papadamou, S., Tzeremes, P., & Corbet, S. (2023). The differential influence of social media sentiment on cryptocurrency returns and volatility during COVID-19. *The Quarterly Review of Economics and Finance*, *89*, 307–317. https://doi.org/10.1016/j.qref.2022.09.004

Lee, A., Li, M., & Zheng, H. (2020). Bitcoin: Speculative asset or innovative technology? *Journal of International Financial Markets, Institutions and Money*, *67*, 101209. https://doi.org/10.1016/j.intfin.2020.101209

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019, July 26). RoBERTa: A robustly optimized BERT pretraining approach. https://doi.org/10.48550/arXiv.1907.11692

López-Cabarcos, M. Á., Pérez-Pico, A. M., Piñeiro-Chousa, J., & Šević, A. (2021). Bitcoin volatility, stock market and investor sentiment. are they connected? *Finance Research Letters*, *38*, 101399. https://doi.org/10.1016/j.frl.2019.101399

McCoy, M., & Rahimi, S. (2020). Prediction of highly volatile cryptocurrency prices using social media [Publisher: World Scientific Publishing Co.]. *International Journal of Computational Intelligence and Applications*, *19*(4), 2050025. https://doi.org/10.1142/S146902682050025X

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Phillips, R. C., & Gorse, D. (2017). Predicting cryptocurrency price bubbles using social media data and epidemic modelling. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–7. https://doi.org/10.1109/SSCI.2017.8280809

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020, February 29). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. https://doi.org/10.48550/arXiv.1910.01108

Tandon, C., Revankar, S., Palivela, H., & Parihar, S. S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? insights from big data analytics. *International Journal of Information Management Data Insights*, *1*(2), 100035. https://doi.org/10.1016/j.jjimei.2021.100035

Wei, D. (1998). *B-money*. Retrieved May 27, 2023, from http://www.weidai.com/bmoney.txt

Wooley, S., Edmonds, A., Bagavathi, A., & Krishnan, S. (2019). Extracting cryptocurrency price movements from the reddit network sentiment. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 500–505. https://doi.org/10.1109/ICMLA.2019.00093

YahooFinance. (2023). *Crypto real time prices & latest news - yahoo finance*. Retrieved May 28, 2023, from https://finance.yahoo.com/lookup/