What is behind investor sentiment about Bitcoin return and volume?

A topic modeling approach

Ehsan Khaksar
Department of Finance, Operations and Information Systems
Goodman School of Business
Brock University, Canada
Email: ek21ej@brocku.ca

Ernest Biktimrov
Department of Finance, Operations and Information Systems
Goodman School of Business
Brock University, Canada
Email: ebiktimirov@brocku.ca

Anteneh Ayanso
Department of Finance, Operations and Information Systems
Goodman School of Business
Brock University, Canada
Email: AAyanso@brocku.ca

Tatyana Sokolyk
Department of Finance, Operations and Information Systems
Goodman School of Business
Brock University, Canada
Email: tsokolyk@brocku.ca

What is behind investor sentiment about Bitcoin return and volume?

A topic modeling approach

Abstract:

This paper examines Bitcoin related discussions on Bitcointalk.com over the 2013-2022 period. Using Latent Dirichlet Allocation (LDA) topic modeling algorithm, we discover eight distinct topics: *Mining*, *Regulation*, *Investment/trading*, *Public perception*, *Bitcoin's nature*, *Wallet*, *Payment*, and *Other*. Importantly, we find differences in relations between different topics' sentiment, disagreement (proxy for uncertainty) and hype (proxy for attention) on one hand and Bitcoin return and trading volume on the other hand. Specifically, among all topics, only the sentiment and disagreement of *Investment/trading* topic have significant contemporaneous relation with Bitcoin return. In addition, sentiment and disagreement of several topics, such as *Mining* and *Wallet*, show significant relationships with Bitcoin return only on the tails of the return distribution (bullish and bearish markets). In contrast, sentiment, disagreement, and hype of each topic show significant relation with Bitcoin volume across the entire distribution. Interestingly, whereas hype has a positive relation with trading volume in a low-volume market, this relation becomes negative in a high-volume market.

Keywords: Bitcoin; disagreement; LDA, sentiment; topic modeling; text analysis

JEL Classification: G12 · G40 · G41

1. Introduction

Since its inception in 2009 as the first cryptocurrency, Bitcoin has grown from a shadowy novelty to the most widely used and valued form of digital money worldwide. As the role of bitcoin market in the financial world has been fast expanding, a growing body of academic literature examines the relation between Bitcoin price and fundamental factors, such as S&P 500 returns (e.g., López-Cabarcos et al., 2021), gold returns (e.g., Panagiotidis et al., 2018; Jareño et al., 2020), the Chicago Board Option Exchange Volatility Index (VIX) (e.g., Aalborg et al., 2019; Koutmos, 2020), and the US dollar exchange rate (e.g., Rajput et al., 2022). Given the decentralized nature of Bitcoin, an increasing number of studies also analyze the relation between behavioral factors, such as media and investor sentiments, and Bitcoin price behavior (e.g., Philippas et al., 2019; Caferra, 2020; Gurdgiev and O'Loughlin, 2020; Ahn and Kim, 2021; Guégan and Renault, 2021; Huynh, 2021; López-Cabarcos et al., 2021).

Using topic modeling analysis – an emerging tool in finance research (e.g., Lowry et al., 2020; Bellstam et al., 2021) – Kim et al. (2017), Linton et al. (2017), Phillips and Gorse (2018), and Poongodi et al. (2018) identify discussion topics about Bitcoin and show that the use of these topics improves the predictive power of their models for Bitcoin return. Loginova et al. (2021) not only extract the topics but also measure the sentiment for each topic for directional prediction of Bitcoin returns.

In this study, we analyze comments posted on Bitcointalk.org, the most popular Bitcoin-related forum, over the 2013-2022 period. We discover eight distinct topics: *Mining*, *Regulation*, *Investment/trading*, *Public perception*, *Bitcoin's nature*, *Wallet*, *Payment*, and *Other*. Importantly, we find differences in relations between different topics' sentiment, disagreement (proxy for uncertainty) and hype (proxy for attention) on one hand and Bitcoin return and trading volume on the other hand. Specifically, among all topics, only the sentiment and disagreement of *Investment/trading* topic have significant contemporaneous relation with Bitcoin return. In addition, sentiment and disagreement of several topics, such as *Mining* and *Wallet*, show significant relationships with Bitcoin return only on the tails of the return distribution (bullish and bearish markets). In contrast, sentiment, disagreement, and hype of each topic show significant relation with Bitcoin volume across the entire distribution. Interestingly, whereas hype has a positive relation with trading volume in a low-volume market, this relation becomes negative in a high-volume market.

We extend the literature in three ways. First, we perform a more robust topic modeling analysis by using different approaches, including Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), Joint Topic Sentiment (JST), and Topic Sentiment Mixture (TSM) to discover topics. In addition, to determine the optimal number of topics, we use both subjective (analyzing the top twenty words and posts for each topic) and objective (coherence scores) approaches. Second, to better understand the importance of each topic, we consider not only the sentiment of each topic, but also its hype and disagreement. Moreover, we analyze the topics' relations to both Bitcoin return and trading volume. Third, using quantile regression models, we examine relations between topics' sentiment/hype/disagreement and Bitcoin return/volume under different market conditions.

2. Literature review

2.1 Investors' attention

Researchers have used different proxies, including Google search trends, Wikipedia, and the number of social media posts, to measure the impact of investors' attention on both stock and the Bitcoin market (Kristoufek, 2013, 2015; Yelowitz, 2015; Athey, 2016).

Kristoufek (2013, 2015) find a correlation between Google searches for Bitcoin-related terms and the price of the cryptocurrency as well as weekly total transaction volume. In addition, relevant Wikipedia views have been suggested as a further signal that could offer a digital footprint of new users learning about a cryptocurrency (Glaser, 2014), these views show a bidirectional association with price (Kristoufek, 2013).

Ibikunle et. al. (2020) examine the impact of heightened attention on the price discovery mechanism of bitcoin. They find that the noisy element of bitcoin pricing is influenced by elevated levels of attention. This suggests that when attention is high, there is an increase in uninformed trading in the bitcoin market. On the other hand, informed trading activity in the market is driven by arbitrage rather than attention.

Dewan (2002) compared the strong relationship between investors' attention and Bitcoin price to that of web visits and firm equity value. Several other studies showed a significant relationship between the number of social media posts as well as Google search with Bitcoin price (Nie and Ji, 2014; Wolk, 2020; Ciaian, 2016; Liu and Tsyvinski, 2018, Abraham, 2018). For example, Smales (2022) uses Google Search Volume (GSV) to show that higher investor attention is associated with increased Bitcoin return.

On the other hand, Urquhart (2018) and Shen (2019) find that while Google trends and the number of tweets significantly influence the trade volume and realized volatility the next day, they are unable to predict the future returns. Using both Google and Wikipedia trend as proxy for popularity (search intensity), Panagiotidis et al. (2019) show that the impact of investors' attention on Bitcoin return has decreased over time. More particularly, they state that a shock in search intensity is now less likely to aggravate Bitcoin bubbles compared to the past. Tumarkin and Whitelaw (2001) show that stock prices cannot be predicted by message board activity; rather, it appears that causality flows the other way, the market can predict the forums' activity. While both Google trends and the number of social media posts are popular proxies for investors' attention, knowledgeable investors who are familiar with cryptocurrency do not Google it; instead, they might discuss it on social media. Therefore, based on Shen's (2019) study, compared to Google Trends, which is a measure of uninformed investor attention, the volume of Bitcoin tweets is a more accurate indicator of investor attention.

In conclusion, to gauge the attention toward Bitcoin, we utilize an innovative measure called Hype (Biktimirov et al., 2021). Hype captures both the intensity and breadth of attention, by considering the number of documents on a specific topic and the weight of topics in each article.

2.2 Investors' sentiment

The main objective of sentiment analysis, also known as opinion mining, is to categorize unstructured text as having a positive, negative, or neutral sentiment polarity score through the computational study of people's attitudes, opinions, and emotions toward events, issues, individuals, entities, and topics (Liu and Zhang, 2012).

Since there is no agreement on how cryptocurrencies should be defined (e.g., White et al., 2020), and there are no obvious underlying fundamentals, it is anticipated that non-fundamental factors like mood and sentiment will play a major role in determining the value of cryptocurrency (Aharon, 2022; Naeem, 2021). In addition, new investors, despite being reasonable will depend more on publicly available information than on their own information, which encourages herd behavior (Banerjee, 1992; Shleifer, 1990). It has been shown that in the existence of herding behavior, sentiment predicts future return (Banerjee, 1992). Thus, as most of Bitcoin investors are young enthusiasts, not necessarily well-informed traders (Yelowitz and Wilson, 2015), and herding behavior matters in the cryptocurrency (Naeem, 2021; Bouri, 2017a) we estimate a significant impact of investor sentiment on cryptocurrency (Naeem, 2021).

In the case of news sentiment, Karalevicius (2018) used lexicon-based sentiment analysis to construct a trading strategy. The results show that news from particular sources, including CoinDesk, NewsBTC, and CoinTelegraph can forecast short-term changes in the price of Bitcoin. However, they also showed that a trader cannot achieve abnormal profits by taking advantage of these market movement patterns. They also confirm the investors' overreaction to news resulting in a pricing pattern where the price moves first in line with the mood before being slightly corrected.

A growing number of studies examine the impact of social media sentiment on cryptocurrency return (e.g., Georgoula, 2015; Garcia, 2015; Kristoufek, 2013; Kraaijeveld, 2020; Matta, 2015; Lamon, 2017). The price variations of Bitcoin were predicted using Twitter sentiment analysis by Georgoula et al. (2015) utilizing a Support Vector Machine (SVM) and several regression models. The authors only discovered a short-term association between the price of Bitcoin and positive sentiment on Twitter. Garcia (2015) employs a lexicon-based method with a Vector Autoregressive (VAR) model and Granger-causality testing to discover that rises in Twitter sentiment polarity precede swings in the price of Bitcoin.

Matta et al. (2015) utilize the lexical-based sentiment analysis method known as 'SentiStrength' to gauge sentiment in tweets. They find that positive tweets forecast changes in the price of Bitcoin in their sample period. Guegan (2021) report that the impact of sentiment on returns is primarily felt during the time of the Bitcoin bubble. This outcome is in line with the existence of irrational emotion-driven noise traders over this time (De Long et al., 1990). Some other studies focus on the sentiment of comments posted by forum users (Kim, 2015; Cohen-Charash, 2013; Bollen, 2011). On the other hand, Abraham (2018) shows that tweets sentiment cannot forecast changes in the price of Bitcoin and Ethereum.

Kraaijeveld et al. (2020) study the predictive power of Twitter sentiment on returns of nine major cryptocurrencies: Bitcoin, Ethereum, XRP, Bitcoin Cash, EOS, Litecoin, Cardano, Stellar, and TRON. For sentiment purposes, they use Loughran and McDonald financial corpus and VADER algorithm and consider three classes of positive (buy), neutral (hold), and negative (sell). They confirm that lexicon-based sentiment analysis generally has a positive bias. By using Granger causality on daily frequency data, they show that Twitter sentiment can significantly predict the return of Bitcoin, Bitcoin Cash, and Litecoin. They also show that this predictive power could be extended to EOS and Tron by considering the Bullishness ratio. For hourly analysis, they find that Twitter measures merely respond to the market rather than causing it.

Using Vector Error Correction Models (VECMs, an extension of VAR) and Granger causality, Mai et al. (2018) show that social media sentiment significantly and positively predicts the Bitcoin price. Their other important finding is that sentiments of forum messages have more predictive power of future Bitcoin prices than tweets do. Finally, they show that the sentiment of posts by the silent majority has a stronger impact on Bitcoin price than that of the vocal minority. Guegan et al. (2021) use posts on StockTwits to see the impact of investor sentiment in social media on Bitcoin returns at various time frequencies (from 1 minute up to 24 hours). In order to measure the sentiment, they use the messages on StockTwits that were self-classified as bullish or bearish and calculated the difference between the number of bullish and bearish messages divided by the total number of massages in that interval. The results of OLS analysis show that the impact of investor sentiment on Bitcoin returns is statistically significant up to 15 minutes intervals, which is supported by Granger causality for the same frequency. However, the coefficients are too small that making it practically impossible to make any profit by trading on investment sentiment. As a robustness check, they show that considering investor attention does not influence the relationship between sentiment and return. Finally, they consider two subsamples, bubbles and post bubbles, and they showed that investor sentiment is significant only during the bubble period.

Wolk (2020) examines the impact of the number and sentiment of tweets and Google trend on short-term performance of Bitcoin, Electroneum, Ethereum, Monero, Ripple, and Zcash using uses VADER and other methods. The researcher finds that the ensemble method provides a better performance than linear regression. In addition, tweets' sentiment and price have a significant negative correlation, and there is also a significant relationship between Google trends, Tweet frequency, and crypto data.

Naeem et al. (2021) study the relationship between the Twitter happiness index and FEARS index (Google search based) and the return of Bitcoin (BTC), Litecoin (LTC), Ripple (XRP), Ethereum (ETH), Monero (XMR), and Dash (DASH). Using linear Granger causality, Happiness only Granger causes ETH and Dash, and FEARS only Granger causes LTC returns. Using quantiles regressions, the authors find that Happiness can predict almost all the cryptocurrencies in this study in the lower and higher quantiles. However, while the FEARS index can also predict the price, this predictability is weaker, heterogeneous across the cryptocurrencies, and mainly in the short horizon.

Yasir et al. (2020) incorporate the sentiment of some mega events into deep learning, linear regression, and support vector regression (SVR) to enhance the performance of these models in predicting the price of Bitcoin, Litecoin, Dash, Monero, and Stellar. They use the Tweets on five mega events, Gaza Attack 2014, Brexit 2016, Hong Kong Protest 2014, Refugee Welcome 2015, and Lahore Blast 2016, to calculate the sentiment. First, they run the models without incorporating the sentiment and they show that considering all the coins, SVR has the most accurate result, while linear regression has the worst performance. Then, considering the sentiment, they show that the performance of the Deep learning model has significantly improved, which shows the importance of sentiment in analyzing cryptocurrencies' prices.

Using high frequency intra-day data, Rognone et al. (2020) show that Bitcoin return react positively to both negative and positive Unscheduled and Bitcoin news, while traditional currencies usually experience a increase (decrease) in return after positive (negative) news. They use this difference as an additional proof to categorize Bitcoin as an asset rather than a currency. Dias et al. (2022) use VIX as a proxy for investor sentiment and Bitcointalk.org merit system as a proxy for investor attention and show that sentiment and attention are significant predictors of bitcoin returns. Then using moments quantile regression, they show the nonlinearity of the relationship between bitcoin returns and sentiment/attention as the coefficients change based on the market conditions.

Based on the aforementioned studies, we apply sentiment analysis using the lexicon-based approach called Valence Aware Dictionary and Sentiment Reasoner (VADER) (Gilber and Hutto, 2014) in combination with Loughran & McDonald financial corpus (Loughran and McDonald, 2011) for more accuracy. It has been shown that the VADER algorithm can outperform both human annotators and most other classifiers (Gilbert and Hutto, 2011).

2.3 Investors' disagreement

Uncertainty can be measured as the dispersion of opinions, forecasts and sentiments across individuals. For example, Bomberger and Frazer (1981) argue that forecast error standard deviation measures the dispersion of opinions among individual forecasts of inflation. Thus, it is a proxy of inflation uncertainty. Ackert and Athanassakos (1997) define the standard deviation of earnings forecast as a proxy for uncertainty where they study the relation between analysts' over-optimism and uncertainty. Poncela and Senra (2017) define uncertainty in the premise of survey forecasts as "the variance of the future outcome of the target indicator conditional to the available information".

Suardi et al. (2022) and Jiang et al. (2022) show that sentiment disagreement positively contributes to Bitcoin returns volatility. In accordance with the aforementioned studies, we use the daily standard deviation of sentiment as the measure for disagreement. We compute the disagreement for each topic separately.

2.4 Topic modeling

A text mining technique known as topic modeling (Blei et al., 2003; Lee & Seung, 1999) generates a list of dominant topics and pertinent keywords from a huge document corpus. By giving readers an immediate overview of the corpus through this thematic information, the need to go through comments, which would otherwise be a tiresome and time-consuming procedure, is eliminated. Most studies attempt to simply extract sentiments or opinions, while the fact that sentiments and opinions are voiced about different aspects or topics is missing (Nguyen, 2015). Considering the growing popularity of cryptocurrencies, understanding the debate issues that influence price would be helpful (Philips, 2018).

Topic modeling techniques have recently been applied to stock and cryptocurrency-related discussions. Twitter posts were trained using the Continuous Dirichlet Process Mixture (CDPM) model to forecast the stock market (Si et al., 2013). Linton (2017) uses dynamic topic modeling, which allows for tracking of subjects and their word' changes over time. The findings demonstrated how the conversations surrounding specific topics have evolved throughout time. Kim et al. (2017) apply topic modeling on posts on Bitcointalk to predict the fluctuation in Bitcoin value. They used non-negative matrix factorization for topic modeling and end up with 50 topics. Using Granger causality, they failed to predict the fluctuation in Bitcoin, while it shows some relationships between some topics and Bitcoin transaction volume and price. They also used Pearson correlation to show that there is a significant correlation between most of the topics and Bitcoin prices. Finally, they got more than 80% accuracy for both price and volume prediction when they included the topics in their deep learning algorithm.

Using dynamic topic modeling, Phillips et al. (2018b) examine what topics on social media (Reddit) have predictive power for Bitcoin and Ethereum pieces. They find that while topics are lagging, which means the market reaction is quicker than social media, some topics precede certain types of price movement.

Poongodi et al (2021) used topic modeling on Reddit, Twitter, and Bitcointalk in order to predict the Bitcoin price movement. After constructing the Document-Term matrix, they ran an LDA topic modeling and found 15 distinct topics. Finally, using a neural network and comparing different models, they showed that including LDA topics increase the R-squared of the model. This model can predict the rise and fall of Bitcoin price very well at the 5 minutes interval.

To sum up, researches find that including the topics of social media discussions can provide useful insights for predicting cryptocurrencies price movements. However, despite the existence of valuable aforementioned studies on topic modeling in cryptocurrency field, there is still a lot to do to fully explore its potential. This gap becomes even more visible when we consider that topics can be used in combination with other textual features such as sentiment and hype.

2.3 Hypotheses Development

In our previous discussion, we noted that there is no clear consensus in the current academic literature regarding the relationship between sentiment, attention, disagreement, and Bitcoin's return/trading volume. While some studies find clear links between these variables, others do not identify such strong connections. Additionally, among the studies that do find significant relationships, there is disagreement about the direction of these links.

Furthermore, certain research highlights the importance of considering various discussion topics when exploring the connection between social media posts and Bitcoin market trends (Loginova et al., 2021; Kim et al., 2017; Phillips and Gorse, 2018; Poongodi et al., 2021). This suggests that investor reactions might vary depending on the specific topic, meaning some discussions might heavily influence investment decisions, while others might have minimal impact.

Given these varying findings, we propose our initial hypothesis:

H1: *The level of significance, magnitude, and direction of the relationship between investors' sentiment/attention/disagreement and Bitcoin's returns/trading volume varies among different topics.*

In addition, Tversky and Kahneman (1979) state that investors act differently in times of depression and terror compared to calm and peaceful times. For instance, investors often join the market when results are positive rather than after a downturn. Research has shown that the predictability of investors’ sentiment and attention on Bitcoin returns and trading volume changes depending on the market situation (e.g., Ma et al., 2018; Garcia, 2013). In addition, several recent research emphasizes the need to use quantile methods or similar techniques to study the extremes of Bitcoin return patterns. This is because Bitcoin return behaviors can change with market conditions (Dias et al., 2022; Aharon et al., 2022; Naeem et al., 2021). So, we suggest:

H2: *The relationship between investors' sentiment/attention/disagreement and Bitcoin return/trading volume change across the distribution of dependent variables, which represent different states of the market*

3. Data

3.1 Online posts

Using the Scrapy package in Python, we collect posts from *BitcoinTalk*,[1] which is a public online forum for discussions about bitcoin, cryptocurrencies, and blockchain in general, for the period from September 2014 to September 2022.[2] Among different forums, we focus on *BitcoinTalk* for two reasons. First, *BitcoinTalk* is the oldest and one of the largest bitcoin-related forums created by the legendary Satoshi Nakamoto, the inventor of bitcoin, in November 2009 (e.g., Bitcointalk, n.d.; Thellmann, 2017). As another evidence of its popularity, *BitcoinTalk* appears at the top of the community section of the official Bitcoin website.[3] Second, Loginova et al. (2021) find that *BitcoinTalk* dataset outperforms *Reddit* and *CryptoCompare* for predicting

[1] https://bitcointalk.org
[2] Our period begins in September 2014, because bitcoin price and trading volume data are available on Yahoo Finance starting from September 17, 2014.
[3] https://bitcoin.org/en/community

directional bitcoin returns. The authors attribute this result to longer and more frequent comments on *BitcoinTalk.*

The Bitcointalk.org message boards are divided into five main sections: ‘Bitcoin’, ‘Economy’, ‘Other’, ‘Alternative cryptocurrencies’, and ‘Local’. Each of these sections includes three or more subsections. Specifically, the section ‘Bitcoin’ consists of ‘Bitcoin Discussion’, ‘Development & Technical Discussion’, ‘Mining’, ‘Bitcoin Technical Support’, and ‘Project Development’ subsections. Similar to Kim et al. (2017) and Mai et al. (2018), we scrap the ‘Bitcoin Discussion’ subsection, where comments appear most frequently. We download the post itself, the date when that post was made, and the thread that it belongs to. After cleaning the missing values, our sample contains 1,954,166 posts. Figure 1 demonstrates the daily frequency of posts from November 22, 2009, the first available post, to September 25, 2022.

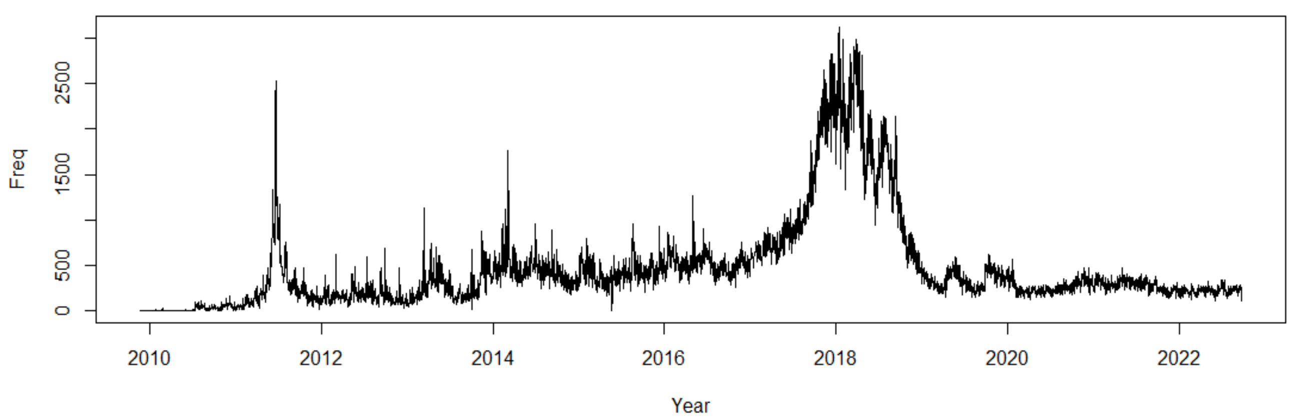

Fig. 1. The daily frequency of posts on Bitcointalk.com.
Note: This figure shows the total number of daily posts under the “Bitcoin Discussion” subsection of Bitcointalk.com from November 22, 2009, the first available post, to September 25, 2022.

3.2 Textual data preprocessing

Social media data are renowned for its high levels of noise and lack of structure. To prepare the data for subsequent topic modeling and sentiment analyses, we perform a series pre-processing steps. First, we remove from the replies the duplicate sentences that quoted earlier posts. Second, we detect the language of each post and keep only posts written in English. Third, we remove URLs, numbers, and whitespaces as they do not carry any sentiment or semantic value. Forth, we expand contractions (e.g., "I'm" to "I am") and negations (e.g., "hasn't" to "has not"), apply case-folding by reducing all letters to lower case (e.g., "SELL" to "sell"), and remove punctuations and stop words, which are irrelevant words with no sentiment or topic significance, using seven stop-word lists suggested by Loughran and McDonald.[4] Finally, we tokenize (e.g., split the text into individual words), select only nouns, and lemmatize (e.g., group words according to their root form) the text.

3.3 Financial data

We download daily Bitcoin return and trading volume data from Yahoo Finance from September 18, 2014, the first date with available data, to September 25, 2022.[5] We calculate Bitcoin return as the first logarithmic difference of closing prices. To approximate normal distribution, we apply log transformation to trading volume. To control for economic variables that may affect Bitcoin price and volume, we use Bloomberg to collect the daily data on the S&P 500 index return, gold return, USD/EUR exchange rate, and VIX index. In addition, we use longtrend package in Python to download Google search trends on word "Bitcoin."

---

[4] "Generic," "auditor," "currencies," "datesandnumbers," "genericlong," "geographic," and "names" stop-word lists are available at https://sraf.nd.edu/textual-analysis/stopwords/. We add additional stop words during the topic modeling analysis stage.

[5] Yahoo Finance provides Bitcoin price and trading volume data agregated over all exchanges from coinmarketcap.com.

Table 1 presents descriptive statistics for the financial variables.

Table 1
Descriptive statistics

Variable	Bitcoin Returns	Bitcoin Volume	S&P 500 Returns	Gold Returns	USD/EUR	VIX	Google Trend
Mean	0.0013	21.572	0.0003	0.0001	0.882	18.460	8.617
Std. Dev.	0.0390	2.860	0.0117	0.0087	0.042	7.785	9.667
Median	0.0019	22.592	0.0006	0.0003	0.885	16.280	5.419
Min	–0.465	15.593	–0.128	–0.059	0.774	9.140	0.898
Max	0.225	26.584	0.090	0.050	1.032	82.690	100
Skewness	–0.765	–0.557	–0.924	–0.211	0.163	2.531	2.792
Kurtosis	10.724	–1.280	16.578	3.602	0.126	11.781	12.714

4. Methods

4.1 Sentiment analysis

Two methods for quantifying text sentiment are lexicon-based (or dictionary-bnased) and machine learning approaches. The lexicon-based method uses a pre-defined list of words or phrases, and each word is assigned a score depending on the emotion associated with it. On the other hand, to quantify feelings, machine learning approach relies on creating intricate models and training them with a huge amount of data.

In our study, we use a robust lexicon-based approach by combining two lexicons. First, we use the Valence Aware Dictionary and Sentiment Reasoner (VADER) algorithm which is a lexicon and rule-based sentiment analysis model. Gilbert and Hutto (2014) demonstrate that VADER can surpass not only most classifier benchmarks, such as SentiWordNet, and Linguistic Inquiry and Word Count (LIWC), but also human experts. By extracting additional sentiment from emoticons, slangs, punctuations, acronyms, and degree modifiers, VADER is particularly suitable for sentiments expressed in social media (Kraaijeveld et al., 2020).

Second, similar to Kraaijeveld (2020), Chen et al. (2021), Zhang et al. (2020), we add the Loughran and McDonald's (2011) dictionary to the VADER lexicon. To combine the dictionaries, we assign the average score of all the positive (negative) words in VADER dictionary to the words in positive (negative) class of Loughran and McDonald's dictionary. VADER calculates a normalized weighted composite compound score between –1 and 1 using the valence scores. According to this polarity score, a social post is either negative (≤ –0.05), neutral (> –0.05 and < 0.05), or positive (≥ 0.05). For our analyses, we use the specific polarity score rather than positive, negative, or neutral classifications.

4.2 Topic modeling

Because text documents consists of words, a topic discussed in several documents can be conveyed by combining words that to appear together. Each document can contain several topics, and the purpose of topic modeling is to uncover the latent topics in documents (Jong et al., 2019). Thus, the results of topic modeling analysis are the probability and distribution of topics in each document and words in each topic (Philips, 2018).

Among several models for extracting topics from a corpus of documents, Latent Dirichlet Allocation (LDA) model (Blei et al, 2003), tends to produce the most interpretable results (Chiru et al., 2014; Qomariyah et al., 2019) and its application has been growing in finance (Philips, 2018; Loginova, 2021; Kim, 2017; Nguyen, 2015; Poongodi, 2021). Another common approach for topic modeling is Latent Semantic Analysis (LSA) introduced by Deerwester et al. (1990). As more recent developments, Topic Sentiment Mixture (TSM) proposed by Mei et al. (2007) and Joint Topic Sentiment (JTS) suggested by Lin and He (2009) extract topic and sentiment simultaneously. In our study, we apply all four models–LDA, LSA, TSM, and JST–to our dataset. We present the results generated by the LDA model, as it produces the most distinct and interpretable topics.

In topic modeling analysis, the number of topics needs to be determined by researchers. Therefore, we run topic modeling analysis for the number of topics ranging from five to fifteen. We select the the optimal number of topics as the one that produces the most distinct and interpretable topics by analyzing the top twenty words and top twenty documents of each topic. In addition, as objective criteria, we use four coherence scores, which are a computational tool that rates topics based on semantic similarity between high scoring words in the topic: the UCI score, C_{UCI}, (Newman et al., 2010), Umass score, C_{UMass}, (Mimno et al., 2011), NMPI score, C_{NMPI}, (Stevens et al., 2012), and V score, C_V, (Roder et al., 2015).

4.3 Sentiment proxies

This study uses three sentiment proxies: sentiment, disagreement, and hype. We compute the daily values of these proxies for each topic and in general, without differentiation among topics. To compute the sentiment and disagreement for each topic, we first need to determine the sentiment regarding each topic in each post. Therefore, we define the sentiment of topic *j* in post i, $TopicSent_{j,i}$, as:

$$TopicSent_{j,i} = Sentiment_i \times W_{j,i}, \tag{1}$$

where $Sentiment_i$ refers to the sentiment score of post *i*, and $W_{j,i}$ indicates the weight of topic *j* in post *i*.

4.3.1 Sentiment

We determine the sentiment of topic *j* on day *t*, $DailyTopicSent_{j,t}$, as the average of that topic sentiment of all posts published on that day:

$$DailyTopicSent_{j,t} = \frac{\sum_{i=1}^{n} TopicSent_{j,i}}{n}, \tag{2}$$

where *n* is the total number of posts publised on day *t*.

Accordingly, the daily general sentiment on day *t*, $DailyGenSent_t$, is calculated as:

$$DailyGenSent_{j,t} = \frac{\sum_{i=1}^{n} Sentiment_i}{n}, \quad (3)$$

where $Sentiment_i$ is the sentiment of post *i* and *n* is the total number of posts on day *t*.

4.3.2 Disagreement

Following Jiang and Marneffe (2022) and Suardi et al. (2022), we compute the disagreement of topic *j* on day *t*, $DailyTopicDisagr_{j,t}$, as the standard deviation of sentiments of topic *j*, $TopicSent_{j,i}$:

$$DailyTopicDisagr_{j,t} = \sqrt{\frac{\sum_{i=1}^{n}(TopicSent_{j,i} - \mu)^2}{n-1}}, \quad (4)$$

where μ is the average of $TopicSent_{j,i}$ on day *t* (or $DailyTopicSent_{j,t}$), and *n* is the total number of posts on that day. Similarly, we compute the general disagreement on day *t*, $DailyGenDisagr_t$, as:

$$DailyGenDisagr_t = \sqrt{\frac{\sum_{i=1}^{n}(Sentiment_i - \mu)^2}{n-1}}, \quad (5)$$

where μ is the average of $Sentiment_i$ on day *t* (or $DailyGenSent_{j,t}$).

4.3.3 Hype

To capture both intensity and breadth of discusions, we consider the number of posts on a specific topic and the weight of topics in each post. Therefore, we compute the the hype of topic *j* on day *d*, $DailyTopicHype_{j,t}$, as the summation of the weight of topic *j* in all posts, *i*, $W_{j,i}$, published on that day:

$$DailyTopicHype_{j,t} = \sum_{i=1}^{n} W_{j,i}, \tag{6}$$

where *n* is the total number of posts published on day *t*.

Finally, we determine the general hype on day *t*, $DailyGenHype_t$, as the total number of posts, *n*, published on that day.

4.4 Regression analyses

To examine relations between Bitcoin return/volume and sentiment proxies, we estimate the following ordinary least squares (OLS) regression model for each topic *j*:

$$Bitcoin_t = a + \beta_1 DailyTopicSent_{j,t} + \beta_2 DailyTopicDisagr_{j,t} + \beta_3 DailyTopicHype_{j,t} + \beta_4 X_t + e_t, \tag{7}$$

The dependent variable, $Bitcoin_t$, is either Bitcoin return, computed as the first logarithmic difference of daily closing prices, or Bicoin trading volume, defined as the natural log of Bicoin trading volume on day *t*. Independent variables include daily topic *j* sentiment, $DailyTopicSent_{j,t}$, disagreement, $DailyTopicDisagr_{j,t}$, and hype, $DailyTopicHype_{j,t}$, scores. and which are sentiment, hype, and disagreement of each topic. Vector X_t is a set of control variables that have shown significant relations with the Bitcoin market in prior studies. It consists of S&P 500 index returns, $S\&P\ 500_t$, the USD/EUR exchange rate, $USD/EUR_{i,t}$, gold returns, $Gold_t$, the CBOE Volatility Index, VIX_t, and the relative value of Google search trend for word "Bitcoin," $Trend_t$. To estimate the significance of regression coefficients, we use the Newey-West heteroscedasticity and autocorrelation-consistent (HAC) standard errors.

To examine the relations across the different parts of the distribution, we also use the quantile regression model. Introduced by Koenker and Bassett (1978), the quantile regression has been widely used in finance research (e.g., Chay et al., 2015; Dias et al. 2022; Singh and Kannadhasan, 2020). Compared to OLS, the quantile regression offers two main advantages. First, quantile regression directly addresses asymmetric effects by estimating lower (higher) quantiles associated with lower (higher) levels of Bitcoin returns and trading volume. Second, quantile regressions are robust to outliers, heteroskedasticity, and skewness of dependent variables (Koenker and Hallock, 2001; Koenker, 2005).

In addition to performing regression analyses at the topic level, we repeat them at the general level as well.

5. Empirical Findings

5.1 Topic modeling results

To determine the number of topics, we examine the top 20 words and top 20 posts associated with each topic for the LDA models ranging in the number of topics from five to fifteen. We select the LDA model with eight topics, which produces the most meaningful, interpretable, and distinct topics. The examination of coherence scores provides another confirmation for the choice of eight topics. Specifically, Figure 2 presents the values for the NMPI, UCI, Umass, and V coherence scores for the number of topics ranging from eight to fifteen. A higher coherence score is associated with better interpretability of the topics, and the NMPI and UCI scores reach the second highest value for eight topics.

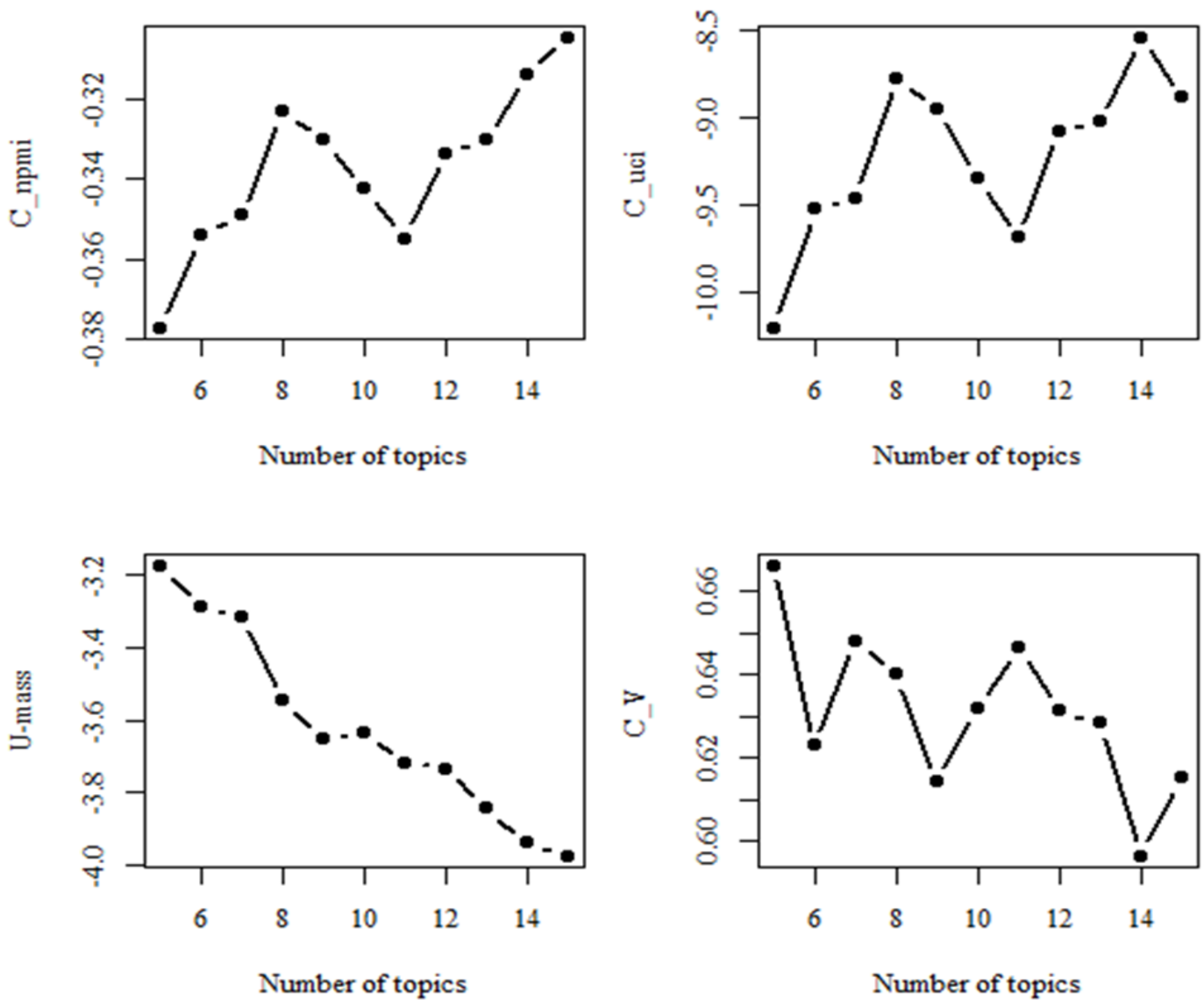

Fig. 2. Coherence scores
Note: This figure shows NMPI, UCI, Umass, and V coherence scores for the number of topics ranging from five to fifteen produced by the LDA model.

We label the topics by examining the top 20 words and top 100 posts related to each of the eight topics. Table 2 shows the eight identified topics, the top 10 words of each topic, and the percentage frequency of each topic based on the dominant topic in each post. Listed from the most to least frequent, the topics are: *Investment/Trading*, *Public Perception*, *Regulation*, *Wallet*, *Bitcoin's Nature*, *Payment*, *Mining*, and *Other*. *Investment/Trading* appears as the main topic in 26.03 percent of posts. In contrast, *Other*, which is the only one topic does not exhibit a coherent theme, has the largest weight only in 3.42 percent of posts.

Table 2
Major topics in Bitcoin posts

Topic	Top 20 words	Frequency
Investment/Trading	crypto, market, time, coin, people, investment, profit, invest, increase, trading, network, future, term, amount, buying, altcoins, wait, rise, demand, end	26.03%
Public Perception	bank, payment, investor, company, business, accept, service, system, project, pay, site, trader, purpose, online, share, atm, method, simple, store, privacy	16.45%
Regulation	government, country, people, risk, news, control, stop, learn, tax, ban, loss, regulation, reason, source, fact, state, kind, system, activity, income	14.35%
Wallet	currency, world, fiat, exchange, country, asset, digital, hold, adoption, future, agree, internet, medium, op, economy, article, el, answer, game, result	12.36%
Bitcoin's Nature	transaction, wallet, address, coin, security, fund, user, amount, access, data, time, send, number, paper, store, computer, secure, node, bought, blockchain	9.15%
Payment	people, time, problem, understand, life, work, community, change, idea, knowledge, future, decision, benefit, job, trust, start, earn, forum, technology, opinion	9.14%
Mining	cryptocurrency, mining, blockchain, satoshi, scam, technology, issue, reach, cost, energy, altcoin, cryptos, mine, hardware, electricity, token, ethereum, plan, space	9.10%
Other	account, experience, usd, show, topic, scammer, growth, paypal, elon, exist, afraid, human, time, musk, die, correct, forum, hear, prediction, eth	3.42%

Note: This table presents eight major topics discussed on on Bitcointalk.com along with the top 20 words and the percentage percentage frequency of each topic based on the dominant topic in each post.

To examine the relations between topics, Table 3 presents Pearson correlation matrices for sentiment (Panel A), disagreement (Panel B), and hype (Panel C). As shown in Panel A, sentiment scores between all topics have positive correlations. They range from 0.243 between

Wallet and *Investment/Trading* to 0.708 between *Bitcoin's Nature* and *Public Perception*. In contrast, Panel B depicts both negative and positive correlations between disagreement scores of topics. Specifically, *Other* and *Investment/Trading* show the strongest negative correlation of –0.453, whereas *Bitcoin's Nature* and *Investment/Trading* exhibit the largest positive correlation of 0.404. Compared to sentiment and disagreement, the hype scores of topics show very strong positive correlations ranging from 0.890 between *Wallet* and *Investment/Trading* to 0.986 between *Bitcoin's Nature* and *Regulation*. Taken together, the correlation analysis highlights strong positive relations between topics in terms of hype, modest positive relations in terms of sentiment, and mixed or no relations in terms of disagreement.

Table 3
Correlation matrices of sentiment, disagreement, and hype scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Sentiment								
(1) Investment/Trading	1							
(2) Public Perception	0.642	1						
(3) Regulation	0.550	0.689	1					
(4) Wallet	0.243	0.359	0.377	1				
(5) Bitcoin's Nature	0.695	0.708	0.702	0.3258	1			
(6) Payment	0.347	0.479	0.503	0.490	0.510	1		
(7) Mining	0.570	0.605	0.591	0.355	0.589	0.439	1	
(8) Other	0.464	0.579	0.603	0.448	0.482	0.530	0.510	1
Panel B: Disagreement								
(1) Investment/Trading	1							
(2) Public Perception	0.214	1						
(3) Regulation	0.098	0.187	1					
(4) Wallet	–0.413	–0.197	–0.226	1				
(5) Bitcoin's Nature	0.404	0.311	0.402	–0.301	1			
(6) Payment	–0.335	–0.082	–0.097	0.217	–0.085	1		
(7) Mining	–0.134	–0.105	–0.056	–0.019	–0.206	–0.121	1	
(8) Other	–0.453	–0.163	–0.187	0.148	–0.432	0.045	0.070	1
Panel C: Hype								
(1) Investment/Trading	1							
(2) Public Perception	0.974	1						
(3) Regulation	0.967	0.977	1					
(4) Wallet	0.890	0.924	0.906	1				
(5) Bitcoin's Nature	0.979	0.984	0.985	0.915	1			
(6) Payment	0.952	0.971	0.970	0.944	0.976	1		
(7) Mining	0.957	0.971	0.964	0.937	0.969	0.967	1	
(8) Other	0.948	0.971	0.955	0.950	0.959	0.965	0.970	1

5.2 Regression analysis results

Table 4 reports the OLS regression results examining the relation between Bitcoin return (Panel A) or Bitcoin trading volume (Panel B) and the sentiment, disagreement, and hype scores of eight topics and all the posts (*General*) controlling for a set of economic variables. Newey-West heteroscedasticity and autocorrelation-consistent (HAC) standard errors are presented in parentheses. As shown in Panel A, among eight topics and general sentiment, only one topic–*Investment/Trading*–exhibits significant relations with Bitcoin return. Specifically, *Investment/Trading* has a positive sentiment coefficient and a negative disagreement coefficient that are significant at the 1% level. Hype coefficients are not significant for any of the topics. Among control variables, coefficients for the S&P 500 index and gold returns are positively related to Bitcoin return at the 1% level. Taken together, only one topic –*Investment/Trading*– seems to matter to Bitcoin investors. Its optimism is associated with higher Bitcoin returns, whereas its disagreement is related to lower returns.

As reported in Panel B, the sentiments of all topics are positively and significantly related to Bitcoin volume. Somewhat surprisingly, the hype of all topics, except *Regulation,* exhibit a negative relation to Bitcoin volume. Disagreement coefficients show mixed results. Specifically, four topics, *Investment/Trading*, *Regulation*, *Bitcoin's Nature*, and *General*, have positive and significant disagreement coefficients. In contrast, three topics, *Wallet*, *Payment*, and *Other*, have negative and significant coefficients. The disagreements of *Public Perception* and *Mining* do not exhibit significant relations with Bitcoin volume.

Table 4
Regressions of Bitcoin return and trading volume on topic sentiment, disagreement, and hype scores

Panel A: Bitcoin return									
	Topics								
	Investment/ Trading	*Public Perception*	*Regulation*	*Wallet*	*Bitcoin's Nature*	*Payment*	*Mining*	*Other*	*General*
Sentiment	0.183***	0.030	0.082	0.094	0.139	0.048	0.138	0.075	0.018
	(0.059)	(0.072)	(0.076)	(0.070)	(0.092)	(0.106)	(0.095)	(0.108)	(0.013)
Disagreement	–0.159***	0.081	0.019	–0.004	–0.023	0.121	–0.048	0.044	–0.014
	(0.051)	(0.074)	(0.052)	(0.044)	(0.073)	(0.086)	(0.061)	(0.082)	(0.030)
Hype	0.00000	–0.00001	0.00000	0.00001	0.00000	0.00001	0.00000	0.00001	0.00000
	(0.00001)	(0.00002)	(0.00002)	(0.00003)	(0.00002)	(0.00002)	(0.00002)	(0.00004)	0.00000
S&P 500	0.779***	0.767***	0.770***	0.774***	0.773***	0.772***	0.769***	0.771***	0.772***
	(0.132)	(0.131)	(0.132)	(0.132)	(0.132)	(0.132)	(0.132)	(0.131)	(0.131)
Gold	0.418***	0.428***	0.429***	0.429***	0.426***	0.433***	0.431***	0.427***	0.428***
	(0.126)	(0.129)	(0.129)	(0.129)	(0.129)	(0.129)	(0.129)	(0.129)	(0.128)
USD/EUR	0.009	–0.008	–0.006	–0.004	–0.006	–0.008	–0.002	–0.003	–0.003
	(0.027)	(0.026)	(0.026)	(0.025)	(0.027)	(0.025)	(0.026)	(0.026)	(0.028)
VIX	0.00020	–0.00002	–0.00001	0.00003	0.00002	0.00005	0.00000	0.00002	0.00004
	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)
Trend	0.00010	–0.00004	–0.00005	–0.00003	–0.00010	0.00003	–0.00005	–0.00003	–0.00004
	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)	(0.00020)
Constant	0.01	0.00	0.00	0.00	0.01	–0.01	0.00	0.00	0.01
	(0.022)	(0.025)	(0.023)	(0.023)	(0.023)	(0.025)	(0.023)	(0.026)	(0.023)
N	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018
R^2	0.055	0.054	0.06	0.055	0.055	0.054	0.056	0.054	0.055

Table 4 continued

Panel B: Bitcoin trading volume									
	Topics								
	Investment/ Trading	*Public Perception*	*Regulation*	*Wallet*	*Bitcoin's Nature*	*Payment*	*Mining*	*Other*	*General*
Sentiment	26.802***	59.814***	59.280***	30.609***	65.052***	71.271***	102.161***	35.681***	15.046***
	(3.677)	(5.633)	(5.407)	(7.051)	(6.335)	(9.830)	(8.111)	(12.385)	(0.870)
Disagreement	35.461***	8.056	44.543***	–27.674***	43.704***	–44.950***	3.447	–87.493***	41.947***
	(3.385)	(5.603)	(4.705)	(3.794)	(6.192)	(7.587)	(6.060)	(6.936)	(2.061)
Hype	–0.004***	–0.004***	–0.001	–0.009***	–0.005***	–0.004*	–0.004*	–0.015***	0.0002**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.0001)
S&P 500	6.355*	9.889***	12.557***	9.531***	12.452***	11.546***	12.326***	11.001***	4.791*
	(3.316)	(3.210)	(3.219)	(3.647)	(3.500)	(3.768)	(3.321)	(3.400)	(2.495)
Gold	0.62	6.53	7.09	6.96	7.25	6.67	6.63	9.886*	–0.37
	(4.673)	(5.186)	(5.045)	(5.227)	(4.742)	(5.645)	(5.130)	(5.070)	(3.386)
USD/EUR	–2.685	–1.101	–0.241	0.286	–3.172	–0.210	–0.003	–2.889	–8.442***
	(2.344)	(3.244)	(3.310)	(3.747)	(2.867)	(3.695)	(3.536)	(3.288)	(1.441)
VIX	0.084***	0.105***	0.102***	0.084***	0.103***	0.106***	0.106***	0.085***	0.053***
	(0.014)	(0.015)	(0.015)	(0.018)	(0.015)	(0.018)	(0.016)	(0.016)	(0.009)
Trend	0.078***	0.137***	0.130***	0.147***	0.126***	0.137***	0.141***	0.140***	0.065***
	(0.012)	(0.018)	(0.016)	(0.019)	(0.015)	(0.020)	(0.019)	(0.017)	(0.008)
Constant	14.684***	16.274***	12.010***	21.750***	15.262***	21.329***	15.795***	27.909***	–1.78
	(2.138)	(3.101)	(3.069)	(3.387)	(2.632)	(3.444)	(3.235)	(2.956)	(1.725)
N	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018
R^2	0.537	0.561	0.639	0.523	0.58	0.488	0.471	0.555	0.786

Note: The table presents the coefficient estimates from the OLS regressions of Bitcoin return and trading volume with Newey-West heteroscedasticity and autocorrelation-consistent (HAC) standard errors (in parentheses). The dependent variable is Bitcoin return, computed as the first logarithmic difference of daily closing prices, or Bicoin trading volume, defined as the natural log of daily Bicoin trading volume. Independent variables include daily topic sentiment, disagreement, and hype scores, and a set of control variables: S&P 500 index returns, gold return, the USD/EUR exchange rate, VIX – the CBOE Volatility Index, and Trend – the relative value of Google search trend for word "Bitcoin." ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using a two-tail test.

4.3.2 Quantile regression

Compared to OLS regression, quantile regression examines the relations across the different parts of the distribution, we also use the quantile regression model.

We run the quantile regression separately on Bitcoin return and Bitcoin trading volume as dependent variables and sentiment, disagreement, and hype of the same day as independent variables for each of the eight topics and all the posts. Consequently, we have 18 different quantile regressions that show the relationship among our variables across 10 deciles (.1 to .9) of the dependent variables (Bitcoin return and Bitcoin trading volume).

Table 5 presents the results of quantile regression across the 10 deciles of Bitcoin returns for independent variables consisting of Sentiment, Disagreement, and Hype associated with *vestment/Trading* and *Public Perception* topic, which are the two most frequent topics.

Table 5

Quantile regression for Bitcoin return in panel a: Investment/Trading and panel b: Public Perception

The table reports a summary of the quantile regression results across 10 deciles of the Bitcoin return distribution for the topics Investment/Trading and Public Perception considering sentiment, disagreement, and hype as independent variables. In addition, we considered the following control variables; S&P 500: The daily return on the S&P 500 index, USD/EUR: The daily USD/EUR exchange rate, Gold: the daily return on gold, VIX: the daily volatility index, and Trend: which is the relative value of Google trend for term "Bitcoin". The coefficients on sentiment, disagreement, hype, and control variables and also summary statistics for all the quantiles are displayed below: the dependent variable (Bitcoin return) regressed on independent variables (sentiment, disagreement, and hype) and control variables of each topic throughout all the 10 quantiles. The dependent variable (Bitcoin return) is calculated as the first logarithmic difference of Bitcoin's daily closing price. Sentiment/Disagreement is the daily average/standard deviation of the sentiment of posts of each topic; Hype is the sum of probabilities of each topic in all the posts in a day. The standard errors are in parentheses.

Panel a: Investment/Trading

	Quantiles								
	10th	***20th***	***30th***	***40th***	***50th***	***60th***	***70th***	***80th***	***90th***
Sentiment	0.416***	0.226***	0.112**	0.085*	0.047	0.032	-0.041	-0.131*	-0.110
	(0.073)	(0.057)	(0.049)	(0.048)	(0.044)	(0.045)	(0.054)	(0.077)	(0.109)
Disagreement	-0.316***	-0.127**	-0.076*	-0.086**	-0.079**	-0.101***	-0.096**	-0.08	-0.10
	(0.068)	(0.056)	(0.040)	(0.037)	(0.035)	(0.036)	(0.043)	(0.057)	(0.084)
Hype	-0.00005***	-0.00003***	-0.00002	0.00000	0.00001	0.00002**	0.00003***	0.00004**	0.00005***
	(0.00002)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00002)
S&P 500	0.944***	0.566***	0.557***	0.495***	0.433***	0.348***	0.376***	0.378***	0.29
	(0.109)	(0.074)	(0.067)	(0.065)	(0.057)	(0.054)	(0.068)	(0.092)	(0.194)
USD/EUR	0.058*	0.046*	0.012	0.010	0.021	-0.003	0.006	0.001	0.006
	(0.034)	(0.027)	(0.019)	(0.017)	(0.018)	(0.018)	(0.022)	(0.025)	(0.045)
Gold	0.324**	0.206*	0.275***	0.248***	0.251***	0.315***	0.354***	0.485***	0.762***
	(0.144)	(0.106)	(0.079)	(0.072)	(0.071)	(0.069)	(0.086)	(0.082)	(0.182)
VIX	0.0003	0.0001	0.0000	0.0001	0.0001	0.0001	0.0000	0.0001	(0.0001)
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)
Trend	-0.001***	-0.001***	- 0.001***	-0.0002	0.0000	0.0004**	0.001***	0.001***	0.002***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Constant	-0.05	-0.042*	-0.01	0.00	-0.01	0.02	0.02	0.03	0.04
	(0.031)	(0.024)	(0.017)	(0.015)	(0.016)	(0.016)	(0.020)	(0.021)	(0.039)
Observation	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018

Note: *p<.01; **p<0.05; ***p<0.01

Panel b: Public Perception

	Quantiles								
	10th	20th	30th	40th	50th	60th	70th	80th	90th
Sentiment	0.231**	0.076	0.025	0.001	-0.016	-0.046	-0.089	-0.178*	-0.353***
	(0.104)	(0.072)	(0.054)	(0.054)	(0.041)	(0.052)	(0.063)	(0.095)	(0.123)
Disagreement	0.084	0.105	0.064	0.039	0.04	-0.044	-0.055	-0.047	0.043
	(0.115)	(0.071)	(0.047)	(0.047)	(0.042)	(0.046)	(0.057)	(0.086)	(0.126)
Hype	-0.0001***	-0.0001***	-0.00003	0	0.00001	0.00003**	0.00003**	0.0001**	0.0001**
	(0.00003)	(0.00002)	(0.00002)	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00002)	(0.00003)
S&P 500	0.899***	0.570***	0.534***	0.479***	0.449***	0.359***	0.375***	0.394***	0.22
	(0.141)	(0.077)	(0.063)	(0.063)	(0.043)	(0.056)	(0.062)	(0.108)	(0.177)
USD/EUR	0.025	0.038	0.009	-0.005	0.016	-0.008	0.006	-0.02	-0.018
	(0.037)	(0.026)	(0.017)	(0.016)	(0.016)	(0.018)	(0.020)	(0.030)	(0.043)
Gold	0.364**	0.272**	0.260***	0.257***	0.283***	0.334***	0.326***	0.426***	0.583***
	(0.174)	(0.107)	(0.072)	(0.071)	(0.022)	(0.071)	(0.077)	(0.121)	(0.188)
VIX	-0.0001	-0.00001	-0.00003	-0.00004	0.0001	0.0001	-0.00002	0.00005	-0.00004
	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)
Trend	-0.001***	-0.001***	-0.001***	-0.0003*	-0.0001	0.0003	0.001***	0.001***	0.002***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Constant	-0.061*	-0.060**	-0.021	-0.003	-0.02	0.017	0.013	0.042	0.048
	(0.035)	(0.024)	(0.016)	(0.015)	(0.015)	(0.016)	(0.018)	(0.028)	(0.040)
Observation	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018

Note: *p<.01; **p<0.05; ***p<0.01

The first important conclusion from the results of quantile regression is that the relationship between hype and Bitcoin return, which was not significant for any of the models in contemporaneous regression, is now significant on both tails of the distribution (lower and upper quantiles) for all the models. More interestingly, we can observe the asymmetric attribute of this relationship, as in lower deciles hype is negatively related to Bitcoin return while the relationship is positive in upper quantiles. In other words, these results indicate that in bearish markets more conversation is associated with a decrease in Bitcoin return, while more discussion of the matter in a bullish market suggests an increase in the return. This conclusion is consistent over all the models.

Regarding the disagreement, the only significant relationship with Bitcoin return in the contemporaneous model is for *Investment/Trading*, and by looking at different quantiles we can see that this negative relationship is mostly derived from the lower deciles (up to 70^{th} percentile). In addition, quantile regression exhibits the same asymmetric pattern in other models that does not show significant relationship between disagreement and Bitcoin return in the contemporaneous regression. Specifically, results of *Payment, and Wallet* suggest a positive relationship only on the lower tail of the Bitcoin return, *Regulation* has significant positive relationship on lowest quantile and negative relationship on the 60^{th} and 70^{th} percentile, while *Other* and *General* indicate significant relationships in the right tail. These results confirm the nonlinear and asymmetric relationship between disagreement and Bitcoin return reported in Aharon et al. (2022).

We also find interesting results regarding the relationship between sentiment and Bitcoin return in some other models. Particularly, in the *General* model, sentiment is positively related to Bitcoin return only in the two lower deciles, while this relationship is negative on the higher tail, which is consistent with Naeem et al. (2021). In addition, *Investment/Trading* results suggest that there is a positive relationship between sentiment and Bitcoin return on the left tail of the distribution. Moreover, among the models that do not show any significant relationship between sentiment and Bitcoin return in the contemporaneous regression, all models, except for mining, show a significant relationship either on lower tail, higher tail, or both.

Finally, looking at Table 7 we can confirm that the relationships are different among different topics. For example, while sentiment has a significant relationship in five of the quantiles of *Investment/Trading* this number is only two for *Public Perception*. This fact is even more obvious in the case of disagreement as none of the quantiles show a significant relationship in *Public Perception*, whereas seven quantiles show a significant relationship for Investment/Trading topic.

Table 6 shows the results of quantile regression across the 9 deciles of Bitcoin trading volume for independent variables consisting of Sentiment, Disagreement, and Hype associated with *Investment/Trading* and *Public Perception*.

Table 6

Quantile regression for Bitcoin trading volume in panel a: Investment/Trading and panel b: Public Perception

The table reports a summary of the quantile regression results across 10 deciles of the Bitcoin trading volume distribution for the topics Investment/Trading and Public Perception considering sentiment, disagreement, and hype as independent variables. In addition, we considered the following control variables; S&P 500: The daily return on the S&P 500 index, USD/EUR: The daily USD/EUR exchange rate, Gold: the daily return on gold, VIX: the daily volatility index, and Trend: which is the relative value of Google trend for term "Bitcoin". The coefficients on sentiment, disagreement, hype, and control variables and also summary statistics for all the quantiles are displayed below: the dependent variable (Bitcoin trading volume) regressed on independent variables (sentiment, disagreement, and hype) and control variables of each topic throughout all the 10 quantiles. The dependent variable (Bitcoin trading volume) is calculated as the logarithm of aggregated Bitcoin's daily trading volume over all the exchanges. Sentiment/Disagreement is the daily average/standard deviation of the sentiment of posts of each topic; Hype is the sum of probabilities of each topic in all the posts in a day. The standard errors are in parentheses.

Panel a: Investment/Trading

	Quantiles								
	10th	***20th***	***30th***	***40th***	***50th***	***60th***	***70th***	***80th***	***90th***
Sentiment	26.810***	29.790***	28.914***	31.327***	29.205***	27.807***	23.892***	8.110***	3.902***
	(3.741)	(3.361)	(3.103)	(3.093)	(3.033)	(2.191)	(3.246)	(1.535)	(0.506)
Disagreement	32.640***	37.130***	39.894***	40.342***	40.522***	40.187***	29.163***	8.849***	5.737***
	(2.681)	(2.450)	(2.251)	(2.349)	(2.328)	(1.660)	(3.510)	(1.636)	(0.677)
Hype	-0.001	-0.002***	-0.002***	-0.003***	-0.003***	-0.004***	-0.004***	-0.003***	-0.003***
	(0.0010)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.0003)	(0.0002)	(0.0001)
S&P 500	1.973	2.952	0.151	8.769*	11.189**	9.816***	13.629***	5.922***	3.726***
	(4.908)	(4.540)	(4.434)	(4.809)	(5.436)	(3.328)	(3.632)	(2.250)	(0.726)
USD/EUR	-4.191***	-4.697***	-2.891**	-2.524**	-2.880**	-3.406***	-2.887**	-0.549	0.05
	(1.211)	(1.260)	(1.143)	(1.269)	(1.299)	(1.050)	(1.391)	(0.805)	(0.312)
Gold	-1.567	-0.22	-4.418	-1.687	4.134	1.154	0.448	-0.554	-0.205
	(4.761)	(5.373)	(4.832)	(5.210)	(5.616)	(5.192)	(6.769)	(2.802)	(1.008)
VIX	0.075***	0.071***	0.077***	0.089***	0.097***	0.106***	0.099***	0.042***	0.029***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)	(0.006)	(0.010)	(0.005)	(0.001)
Trend	0.066***	0.070***	0.076***	0.077***	0.080***	0.076***	0.072***	0.056***	0.053***
	(0.010)	(0.005)	(0.006)	(0.006)	(0.004)	(0.002)	(0.004)	(0.002)	(0.001)
Constant	14.209***	14.571***	13.008***	12.701***	13.424***	14.386***	16.959***	21.208***	21.984***
	(1.073)	(1.101)	(0.994)	(1.149)	(1.205)	(1.011)	(1.421)	(0.916)	(0.309)
Observation	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018

Note: *p<.01; **p<0.05; ***p<0.01

Panel b: Public Perception

	Quantiles								
	10th	***20th***	***30th***	***40th***	***50th***	***60th***	***70th***	***80th***	***90th***
Sentiment	64.861***	69.563***	74.126***	75.374***	65.151***	44.529***	19.202***	8.651***	7.297***
	(3.888)	(4.448)	(4.605)	(4.486)	(3.970)	(5.327)	(2.968)	(1.238)	(1.203)
Disagreement	7.405**	9.392***	10.886**	10.044**	7.796	5.077	1.765	3.724***	3.541***
	(2.991)	(2.161)	(4.490)	(4.520)	(5.203)	(3.461)	(2.127)	(1.301)	(1.117)
Hype	-0.001	-0.001	-0.002***	-0.003***	-0.004***	-0.005***	-0.005***	-0.005***	-0.005***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0004)	(0.0003)	(0.0004)	(0.0002)	(0.0001)
S&P 500	9.682**	10.651**	13.129***	12.488**	10.036***	12.128***	7.581***	4.697***	3.433***
	(4.583)	(5.000)	(5.002)	(5.069)	(3.890)	(2.801)	(2.284)	(0.979)	(0.812)
USD/EUR	2.248	3.051**	2.082	-0.651	-3.897***	-5.822***	-2.205**	-0.901**	-0.367
	(1.447)	(1.326)	(1.518)	(1.353)	(1.268)	(1.111)	(1.017)	(0.417)	(0.329)
Gold	4.76	5.294	4.672	5.684	0.763	3.093	0.622	1.212	1.988
	(5.142)	(5.822)	(6.448)	(6.855)	(6.514)	(3.142)	(3.724)	(1.517)	(1.617)
VIX	0.113***	0.116***	0.125***	0.125***	0.122***	0.109***	0.058***	0.041***	0.027***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.006)	(0.002)	(0.002)
Trend	0.137***	0.158***	0.164***	0.160***	0.151***	0.118***	0.083***	0.066***	0.056***
	(0.011)	(0.009)	(0.006)	(0.004)	(0.007)	(0.008)	(0.005)	(0.003)	(0.002)
Constant	10.170***	9.560***	10.537***	13.686***	18.116***	22.428***	22.764***	22.625***	22.836***
	(1.377)	(1.211)	(1.440)	(1.255)	(1.416)	(1.139)	(0.878)	(0.387)	(0.337)
Observation	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018	2,018
Note:								*p<.01; **p<0.05; ***p<0.01	

Considering all the results, we can make two interesting conclusions. First, the disagreement for *Public Perception and mining*, which is not significant at the 5% level in the contemporaneous model, suggests some significant relationship at the 1% level on the lower and upper quantiles. Furthermore, we can confirm the existence of asymmetry in the relationship between hype and Bitcoin trading volume in *Mining*, *Regulation*, *Payment*, and general model. In fact, hype tends to have a significant positive relationship with Bitcoin trading volume in the left tail and negative one in the right tail. In other words, these results indicate that for these models in bearish markets more conversation is associated with an increase in Bitcoin return, while more discussion of the matter in bullish markets suggest a decrease in the trading volume. In the *General* model, this relationship stays positive in all the quantiles.

The asymmetric results almost do not exist for sentiment. In other words, sentiment of all the topics (except for the *Wallet* that is only significant on the lower tail) are mostly significant in all the quantiles and do not change sign across the trading volume distribution.

5. Discussion

The main purpose of this study is twofold. First, we discover the main topics regarding Bitcoin that have been discussed on Bitcointalk.com, one of the most popular sources of Bitcoin related online conversations. Second, we examine relationships between investors' online discussions (sentiment, disagreement, and hype of each topic) and Bitcoin market (Bitcoin return and trading volume).

Using the Latent Dirichlet allocation (LDA) topic modeling method, we discover eight distinct topics: *Mining*, *Regulation*, *Investment/trading*, *Public perception*, *Bitcoin's nature*, *Wallet*, *Payment*, and *Other*. To measure the sentiment of online postings, we complement VADER sentiment analysis model with the Loughran-McDonald finance specific dictionary.

We find that contemporaneous sentiment and disagreement of *Investment/trading* topic have a *significant* relationship with Bitcoin return. In addition, quantile regression suggests that some of formerly insignificant coefficients in contemporaneous model show significance in some topics (In case of Sentiment: *Wallet*, Disagreement: *Mining, Public Perception, Wallet, Payment, General*, Hype: all the topics and *General* model) in some parts of the return distribution (mostly in extremes, which indicate bullish and bearish market.).

In addition to asymmetric significance attribute, quantile regression highlights the direction asymmetry in these relationships. In particular, it suggests that disagreement and hype for most of the topics have a different direction of relationship with Bitcoin return in bullish and bearish markets. For example, in all models, hype is negatively related to Bitcoin return in lower tails, while this relationship is positive in higher ones; it means that more discussion in a bearish (bullish) is associated with lower (higher) return.

Regarding the Bitcoin trading volume, the results show a significant relationship between trading volume and sentiment, disagreement, and hype of almost all the topics, for most quantiles. Another important observation, is the asymmetric attribute of hype based on the quantile regression results in some of the topics. Specifically, hype is positively related to trading volume in a low volume market, while this relationship is negative in high volume market. moreover, we found asymmetric attributes in the relationship between sentiment, disagreement and Bitcoin trading volume in some of the topics.

Taken together, we contribute to existing literature by showing the importance of considering different topics discussed in social media, as different topics show different levels of significance, magnitude, and direction in their relationship with Bitcoin return and trading volume. In addition, we show the existence of nonlinearity and asymmetric relationships for some topics.

6. Conclusion

Bitcoin and other cryptocurrencies in general, offer many unique benefits by removing the existing payment barriers, especially in the case of global trading, and reducing the transaction fee, which can provide the economy with social welfare and wealth. However, to fully exploit these potentials there should exist a deep comprehension of the market activity and price fluctuation. Therefore, in this thesis, we explained the relationship between social media metrics and the Bitcoin market.

Using a robust finance-specific lexicon-based sentiment analysis and Latent Dirichlet Allocation (LDA) topic modeling combined with simple contemporaneous and quantile regression analysis; we show the existence of significant relationships between sentiment, disagreement, and hype of posts from Bitcointalk.com on one hand and Bitcoin return and trading volume on the other hand over the course of September 2014 to September 2022.

The most important conclusions of this thesis are as follows: First, we show the importance of considering different topics of discussion in social media, as different topics show different levels of significance, magnitude, and direction in their relationship with Bitcoin return and trading volume. Second, we demonstrate the nonlinearity and asymmetric attribute of the relationships via quantile regression analysis.

The findings have implications at individual, business, and government levels. First, these relationships indicate the value of information available on social media, especially when we consider different topics, different market conditions, and delayed relationships. This kind of information can influence future returns through the price-formation process, which can help investors discern the future value of bitcoin. In addition, this greater predictability makes Bitcoin a more reliable part of investment portfolios. Second, by understanding market movement patterns, businesses can make better decisions about adopting Bitcoin or launching their own cryptocurrency (Initial Coin Offering (ICO)). Finally, monitoring social media can help governments curb the potential systematic risk associated with Bitcoin in a timelier manner.

Finally, we extract topics and sentiment of posts separately but future studies can try to extract them simultaneously to generate more reliable results. As a caution, they should find ways other than the conventional methods considered in this thesis (i.e., JST and TSM), which did not produce interpretable and distinct topics. A possible solution for this purpose can be found in recent deep and machine learning techniques. Second, considering other cryptocurrencies can potentially add new insights. Moreover, we limit our data to only-English posts on the forum; as the Bitcoin market consists of investors and contributors all around the world. Considering and comparing posts in other languages may lead to insights about the potential effects of cultural differences. Finally, in addition to Bitcoin return and trading volume, Bitcoin return volatility can be examined as another dependent variable.

References

Aalborg, H. A., Aarhus, P., Molnár, P., & de Vries, J. E. (2019). What can explain the price, volatility and trading volume of Bitcoin? *Finance Research Letters, 29*, 255-265.

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

Aharon, D. Y., Demir, E., Lau, C. K. M., & Zaremba, A. (2022). Twitter-Based uncertainty and cryptocurrency returns. *Research in International Business and Finance, 59*, 101546.

Ahn, Y., Kim, D., 2021. Emotional trading in the cryptocurrency market. Finance Res. Lett. 42, 101912. https://doi.org/10.1016/j.frl.2020.101912.

Alagidede, P., & Panagiotidis, T. (2012). Stock returns and inflation: Evidence from quantile regressions. *Economics Letters, 117*(1), 283-286.

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

Badshah, I. U. (2013). Quantile regression analysis of the asymmetric return-volatility relation. *Journal of Futures Markets, 33*(3), 235-265.

Baur, D. G., Dimpfl, T., & Jung, R. C. (2012). Stock return autocorrelations revisited: A quantile regression approach. *Journal of Empirical Finance, 19*(2), 254-265.

Bellstam, G., Bhagat, S., Cookson, A.J., 2021. A Text-Based Analysis of Corporate Innovation. Manage. Sci. 67, 4004-4031. https://doi.org/10.1287/mnsc.2020.3682.

Bitcointalk (n.d.) https://coinmarketcap.com/alexandria/glossary/bitcointalk

Bouri, E., Jalkh, N., Molnár, P., & Roubaud, D. (2017). Bitcoin for energy commodities before and after the December 2013 crash: Diversifier, hedge or safe haven? *Applied Economics, 49*(50), 5063-5073.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Finance Research Letters, 20*, 192-198.

Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting, 37*(1), 28-43.

Chiang, T. C., & Li, J. (2012). Stock returns and risk: Evidence from quantile regression analysis. *Journal of Risk and Financial Management, 5*(1), 1-130.

Chiru, C. G., Rebedea, T., & Ciotec, S. (2014, April). Comparison between LSA-LDA-Lexical Chains. In *WEBIST (2)* (pp. 255-262).

Ciaian, P., Rajcaniova, M., & Kancs, D. (2016). The economics of BitCoin price formation. *Applied Economics, 48*(19), 1799-1815.

Dastgir, S., Demir, E., Downing, G., Gozgor, G., & Lau, C. K. M. (2019). The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. *Finance Research Letters, 28*, 160-164.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Demir, E., Gozgor, G., Lau, C. K. M., & Vigne, S. A. (2018). Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters, 26*, 145-149.

Demirer, R., Pierdzioch, C., & Zhang, H. (2017). On the short-term predictability of stock returns: A quantile boosting approach. *Finance Research Letters, 22*, 35-41.

Dias, I. K., Fernando, J. R., & Fernando, P. N. D. (2022). Does investor sentiment predict bitcoin return and volatility? A quantile regression approach. *International Review of Financial Analysis, 84*, 102383.

Garcia, D., & Schweitzer, F. (2015). Social signals and algorithmic trading of Bitcoin. *Royal Society Open Science*, *2*(9), 150288.

Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D., & Giaglis, G. M. (2015). Using time-series and sentiment analysis to detect the determinants of bitcoin prices. *Available at SSRN 2607167*.

Guégan, D., & Renault, T. (2021). Does investor sentiment on social media provide robust information for Bitcoin returns predictability?. *Finance Research Letters, 38*, 101494.

Gurdgiev, C., O'Loughlin, D., 2020. Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty. J. Behav. Exp. Finance. 25, 100271. https://doi.org/10.1016/j.jbef.2020.100271.

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, *28*(3), 791-837.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media, 8*(1), 216-225.

Huynh, T.L.D., 2021. Does Bitcoin react to Trump's tweets? J. Behav. Exp. Finance. 31, 100546. https://doi.org/10.1016/j.jbef.2021.100546.

Jareño, F., González, M., Tolentino, M., Sierra, K. Bitcoin and gold price returns: A quantile regression and NARDL analysis. Resour. Policy. 67, 101666. https://doi.org/10.1016/j.resourpol.2020.101666.

Jiang, N. J., & de Marneffe, M. C. (2022). Investigating Reasons for Disagreement in Natural Language Inference. *Transactions of the Association for Computational Linguistics, 10*, 1357-1374.

Jin, X., Zhou, Y., & Mobasher, B. (2005). A maximum entropy web recommendation system: Combining collaborative and content features. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 612-617).

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance, 19*(1), 7-17.

Kim, Y. B., Lee, J., Park, N., Choo, J., Kim, J. H., & Kim, C. H. (2017). When Bitcoin encounters information in an online forum: Using text mining to analyze user opinions and predict value fluctuation. *PloS one, 12*(5), e0177630.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society, 33*(1), 33-50.

Koutmos, D., 2020. Market risk and bitcoin returns. Ann. Oper. Res. 294, 453–477. https://doi.org/10.1007/s10479-019-03255-6.

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

Krestel, R., Fankhauser, P., & Nejdl, W. (2009). Latent Dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on Recommender systems* (pp. 61-68).

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific reports, 3*(1), 1-7.

Li, M. Y., & Wu, J. S. (2014). Analysts' forecast dispersion and stock returns: A quantile regression approach. *Journal of Behavioral Finance*, *15*(3), 175-183.

Lin, C., & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 375-384).

Linton, M., Teo, E. G. S., Bommes, E., Chen, C. Y., & Härdle, W. K. (2017). Dynamic topic modelling for cryptocurrency community forums. In *Applied quantitative finance* (pp. 355-372). Springer, Berlin, Heidelberg.

Loginova, E., Tsang, W. K., van Heijningen, G., Kerkhove, L. P., & Benoit, D. F. (2021). Forecasting directional Bitcoin price returns using aspect-based sentiment analysis on online text data. *Machine Learning*, 1-24.

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 Res. Lett. 38, 101399. https://doi.org/10.1016/j.frl.2019.101399.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance, 66*(1), 35-65.

Lowry, M., Michaely, R., Volkova, E., 2020. Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO. Rev. Financ. Stud. 33, 5510–5554. https://doi.org/10.1093/rfs/hhaa007.

Ma, C., Xiao, S., & Ma, Z. (2018). Investor sentiment and the prediction of stock returns: a quantile regression approach. *Applied Economics*, *50(50)*, 5401-5415.

Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, R. H. L. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems, 35*(1), 19-52.

Matta, M., Lunesu, I., & Marchesi, M. (2015). Bitcoin Spread Prediction Using Social and Web Search Media. In *UMAP workshops* (pp. 1-10).

Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C. (2007). Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web* (pp. 171-180).

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 262-272).

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

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In Human language technologies: *The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 100-108).

Nguyen, T. H., & Shirai, K. (2015). Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1354-1364).

Nie, S., & Ji, Q. (2014). Feature learning using Bayesian linear regression model. In *2014 22nd International Conference on Pattern Recognition* (pp. 1502-1507). IEEE.

Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of Bitcoin returns: A LASSO approach. *Finance Research Letters, 27,* 235-240.

Pedersen, T. Q. (2015). Predictable return distributions. *Journal of Forecasting*, *34*(2), 114-132.

Philippas, D., Rjiba, H., Guesmi, K., Goutte, S., 2019. Media attention and Bitcoin prices. Finance Res. Lett. 30, 37–43. https://doi.org/10.1016/j.frl.2019.03.031.

Phillips, R. C., & Gorse, D. (2018). Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PloS one*, *13*(4), e0195200.

Phillips, R. C., & Gorse, D. (2018). Mutual-excitation of cryptocurrency market returns and social media topics. In *Proceedings of the 4th international conference on frontiers of educational technologies* (pp. 80-86).

Polasik, M., Piotrowska, A. I., Wisniewski, T. P., Kotkowski, R., & Lightfoot, G. (2015). Price fluctuations and the use of Bitcoin: An empirical inquiry. *International Journal of Electronic Commerce*, *20*(1), 9-49.

Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2021). Global cryptocurrency trend prediction using social media. *Information Processing & Management*, *58*(6), 102708.

Poyser, O. (2017). Exploring the determinants of Bitcoin's price: An application of Bayesian Structural Time Series. *arXiv preprint arXiv:1706.01437*.

Poyser, O. (2019). Exploring the dynamics of Bitcoin's price: A Bayesian structural time series approach. *Eurasian Economic Review*, *9*(1), 29-60.

Qomariyah, S., Iriawan, N., & Fithriasari, K. (2019). Topic modeling twitter data using latent dirichlet allocation and latent semantic analysis. *AIP Conference Proceedings*, *2194*(1), 020093.

Rajput, S.K.O., Soomro, I.A., Soomro, N.A., 2022. Bitcoin sentiment index, bitcoin performance and US dollar exchange rate, J. Behav. Finance 23 (2), 150–165, https://doi.org/10.1080/15427560.2020.1864735.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).

Rognone, L., Hyde, S., & Zhang, S. S. (2020). News sentiment in the cryptocurrency market: An empirical comparison with Forex. *International Review of Financial Analysis, 69,* 101462.

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

Smyth, L. (2013). The demographics of Bitcoin (Part 1 updated). *Simulacrum (https://spacedruiddotcom. wordpress. com/2013/03/04/the-demographics-of-bitcoinpart-1-updated/, last accessed 18 January 2018)*.

Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from Bitcoin, ethereum, dash, litcoin, and monero. *Journal of Economics and Financial Analysis, 2*(2), 1-27.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning* (pp. 952-961).

Suardi, S., Rasel, A. R., & Liu, B. (2022). On the predictive power of tweet sentiments and attention on Bitcoin. *International Review of Economics & Finance, 79*, 289-301.

Thellmann, P. (2017) From Nakamoto to 'HODL': 5 Intriguing Facts About Iconic Bitcointalk Forum. Retrieved from: https://cointelegraph.com/news/from-nakamoto-to-hodl-5-intriguing-facts-about-iconic-bitcointalk-forum

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters, 148,* 80-82.

Urquhart, A. (2018). What causes the attention of Bitcoin? *Economics Letters, 166*, 40-44.

Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 448-456).

Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from Twitter posts. In *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 231-238). Springer.

Wołk, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems, 37*(2), e12493.

Xing, D., & Girolami, M. (2007). Employing latent Dirichlet allocation for fraud detection in telecommunications. *Pattern Recognition Letters, 28*(13), 1727-1734.

Zhu, Y., & Dickinson, D. (2017). Analysis on the influence factors of Bitcoin's price based on VEC model. *Financial Innovation, 3*(3), 1-13.