



Contents lists available at ScienceDirect

The British Accounting Review

journal homepage: www.elsevier.com/locate/bar

From whales to waves: Social media sentiment, volatility, and whales in cryptocurrency markets

Suwan(Cheng) Long^a, Ying Xie^b, Zhengyuan Zhou^c, Brian Lucey^{d,e}, Andrew Urquhart^{f,*}

^a IESEG School of Management, University of Lille, Lille, France

^b Cranfield School of Management, Cranfield University, College Road, Cranfield, MK43 0AL, United Kingdom

^c Stern School of Business, New York University, New York, NY 10012, United States

^d Trinity Business School, Trinity College Dublin, Dublin 2, Ireland

^e College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates

^f Birmingham Business School, University of Birmingham, United Kingdom

ARTICLE INFO

Keywords:

Sentiment analysis
Asymmetric TVP-VAR
Market dynamics
Cryptocurrency
Whales

ABSTRACT

This paper examines the relationship between cryptocurrency market dynamics and investor sentiment, employing advanced techniques like time-variant Granger causality and asymmetric time-varying parameter vector autoregression (TVP-VAR) frequency connectivity. We create unique sentiment analysis tools, including a custom cryptocurrency sentiment lexicon, to deeply analyze content in the cryptocurrency domain, particularly focusing on investor discussions and viewpoints. Our findings demonstrate a significant, evolving link between market sentiment and cryptocurrency movements. A key observation is that the volatility of shock transmission is tightly connected to major market events, often influenced by large-scale investors, or “whales”. Our study indicates that market sentiment consistently affects both short- and long-term cryptocurrency volatility, underlining the crucial influence of investor sentiment in driving the dynamics of the cryptocurrency market. This underscores the importance of understanding investor sentiment for predicting and navigating the cryptocurrency market.

1. Introduction

Characterized by rapid growth and distinctive dynamics, cryptocurrency markets attract significant attention from investors and researchers. These markets present a uniquely suitable environment for examining the relationship between social media sentiment and market volatility due to several distinguishing features. Unlike traditional securities markets, which operate within established institutional frameworks, cryptocurrencies emerged without conventional valuation metrics or regulatory infrastructures. As noted by Yermack (2015) and Baur, Hong, and Lee (2018), cryptocurrencies have predominantly been perceived as speculative assets. This speculative nature creates an information environment where investor sentiment plays an outsized role compared to traditional markets.

Cryptocurrencies are further characterized by exceptional volatility, which poses significant challenges for investors and increases the importance of understanding underlying drivers (Saleh, 2019). Factors like market size, beta, and momentum have emerged as crucial determinants of expected returns in cryptocurrency markets (Chen & Zimmermann, 2019; Feng, Giglio, & Xiu, 2020). The momentum premiums observed in cryptocurrencies appear closely tied to investor behavior, particularly the tendency

* Corresponding author.

E-mail address: a.urquhart@bham.ac.uk (A. Urquhart).

<https://doi.org/10.1016/j.bar.2025.101682>

Received 1 July 2024; Received in revised form 20 March 2025; Accepted 29 April 2025

Available online 16 May 2025

0890-8389/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Please cite this article as: Suwan(Cheng) Long et al., *The British Accounting Review*, <https://doi.org/10.1016/j.bar.2025.101682>

towards overreaction — a phenomenon extensively explored within traditional finance (Barberis, Shleifer, & Vishny, 1998; Daniel, Hirshleifer, & Subrahmanyam, 1998). Liu and Tsyvinski (2020) further highlight that cryptocurrencies exhibit low correlation with traditional market risk factors, suggesting potential diversification benefits (Akyildirim, Corbet, Lucey, Sensoy, & Yarovaya, 2020; Corbet, Meegan, Larkin, Lucey, & Yarovaya, 2018; Platanakis & Urquhart, 2020).

Extending this exploration, studies by Cong, Li, and Wang (2020) explore dynamic user adoption models that can predict pricing and adoption patterns throughout a cryptocurrency's lifecycle. This research offers valuable insights into the long-term viability of different cryptocurrencies. Additionally, researchers like Pagnotta and Buraschi (2018), Pagnotta (2021), and Biais, Bisière, Bouvard, Casamatta, and Menkveld (2023) are exploring the influence of network effects, speculator sentiment, and the potential for multiple equilibria in cryptocurrency returns. These studies highlight the complex interplay of market forces and investor psychology that characterize the rapidly evolving landscape of cryptocurrency finance. Shifting the focus to the broader financial markets, the impact of institutional investors is a well-documented phenomenon, notably within traditional stock markets. Field and Lowry (2009) research suggests that the superior performance of these institutions derives from their ability to identify and hedge against underperforming firms, primarily by analyzing publicly available information. There is a substantial body of literature discussing the positive market returns on institutional investors' investments and the characteristics of their stock purchases, including aspects such as corporate governance (Parrino, Sias, & Starks, 2003; Leuz, Lins, & Warnock, 2009; Bushee & Miller, 2012), research and development (Daniel & Titman, 2006; Jiang, 2010) and corporate social responsibility (CSR) (Johnson & Greening, 1999; Oh, Chang, & Martynov, 2011; Dyck, Lins, Roth, & Wagner, 2019), among others. In contrast, retail investors are often considered uninformed or irrational, which may be due to psychological and cognitive biases that result in suboptimal investment decisions. These biases include overconfidence (Kyle & Wang, 1997; Daniel et al., 1998; Daniel, Hirshleifer, & Subrahmanyam, 2001), and representativeness bias (Tversky & Kahneman, 1974; De Bondt & Thaler, 1987; De Bondt, 1993; El-Gazzar, 1998). However, recent research suggests a shift in the power dynamics between institutional and retail investors. Kaniel, Liu, Saar, and Titman (2012) argues that retail investors tend to show positive excess returns after trading. Barber, Odean, and Zhu (2009) argues that trades initiated by retail investors can have a significant impact on the market. This shift was highlighted by the GameStop short squeeze in early 2021, which highlighted the potential impact of retail investors' collective behavior on stock market activity (Long, Lucey, Xie, & Yarovaya, 2023).

The increasing prominence of social media platforms is transforming how investors access and process market information. Market sentiment, as conceptualized by Baker and Wurgler (2007), captures collective investor mood and expectations, thus becoming an essential factor for traders attempting to predict market movements. Retail and institutional investors alike increasingly leverage social media to gauge market sentiment and forecast economic outcomes (Allon & Zhang, 2017; Jiao, Veiga, & Walther, 2020). While previous studies have employed online data like Google search volumes and product reviews to predict consumer behavior and equity values (Geva, Oestreicher-Singer, Efron, & Shimshoni, 2017; Luo, Zhang, & Duan, 2013; Wu & Brynjolfsson, 2009), cryptocurrency markets uniquely rely almost entirely on social media platforms for information dissemination. Cryptocurrency information and investor sentiment are predominantly shaped by online communities and social media rather than traditional financial reporting (Akyildirim, Aysan, Cepni, & Darendeli, 2021; Milas, Panagiotidis, & Dergiades, 2021). This phenomenon is intensified by the dominant role of retail investors, who heavily depend on sentiment-driven discussions on platforms such as Twitter and Reddit (Beckers, 2019; Smales, 2022).

Importantly, recent research indicates that social media sentiment exerts a stronger and more immediate influence on cryptocurrency prices compared to traditional financial assets (Rognone, Hyde, & Zhang, 2020). Social media enables real-time information dissemination, allowing investors to react instantly, which is critical in highly speculative markets. Twitter sentiment, for instance, has demonstrated predictive power for cryptocurrency price movements (Kraaijeveld & De Smedt, 2020), surpassing traditional news-based sentiment in predictive strength and accuracy for market timing strategies (Beckers, 2019). Additionally, the speculative nature and volatility of cryptocurrencies amplify investor herd behavior and tendency towards overreaction based on social media trends (Bowden & Gemayel, 2022; Smales, 2022).

Traditional sentiment analysis methods often rely on dictionaries of pre-defined words associated with specific emotions or themes. Researchers quantify the sentiment of a text by calculating the sum of sentiment scores based on the presence of dictionary words within the text. The simplicity and interpretability of these methods have made them particularly appealing for applications where transparency and explainability are crucial. Earlier research has commonly employed dictionaries like the General Inquirer (GI) built-in dictionary and DICTION for sentiment analysis (Tetlock, 2007; Engelberg, 2008; Feldman, Govindaraj, Livnat, & Segal, 2008; Tetlock, Saar-Tsechansky, & Macskassy, 2008; Henry & Leone, 2016). However, one major drawback of these general language dictionaries is their non-specificity to particular domains, rendering them potentially insufficient for deriving accurate results within specific contexts such as finance. Recognizing this, significant advancements have been made in developing specialized dictionaries tailored to particular domains. The Loughran and McDonald (2011) dictionary for financial sentiment analysis exemplifies this trend, demonstrating improved accuracy over general-purpose dictionaries in its domain. This specialization indicates the method's adaptability to the unique linguistic features of specific fields. These domain-specific tools have become increasingly prevalent in past studies, indicating their value and effectiveness within the financial domain (Doran, Peterson, & Price, 2012; Huang, Teoh, & Zhang, 2014; Jegadeesh & Wu, 2013; Chen, De, Hu, & Hwang, 2013; Liu & McConnell, 2013; Loughran & McDonald, 2013). The limitations of dictionary-based methods, particularly their lack of adaptability to new contexts, paved the way for machine learning approaches. These methods leverage annotated datasets to train algorithms capable of classifying text according to sentiment or thematic categories (Pang, Lee et al., 2008). The evolution from simple models, such as the bag-of-words, to more complex neural network-based models, like convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, marked a significant advancement in the field (Le & Mikolov, 2014; Tang, Qin, & Liu, 2015; Yang et al., 2016). These models are capable of capturing semantic relationships and contextual nuances, offering a more sophisticated understanding of textual data. Relevantly,

for the context of this study, sentence-level sentiment classification is employed, which involves discerning the sentiment expressed within a particular sentence (Wiebe, Bruce, & O'Hara, 1999).

The emergence of Large Language Models (LLMs) like Chat-GPT¹ has significantly impacted the field of textual analysis, heralding a new era where text generation, context understanding, and sentiment analysis are conducted with remarkable precision. This evolution, noted by researchers such as Bommasani et al. (2021), Yang and Menczer (2023), and Zhao et al. (2023), underscores the breadth of LLMs' capabilities, extending beyond specific tasks to general natural language processing and even domain-specific applications. Particularly in sentiment analysis, LLMs have shown to outperform traditional methods, with studies by Liang et al. (2022) and Zeng et al. (2022) demonstrating their high accuracy in detecting emotional inclinations in text. However, despite the growing literature, there remains an open need for comprehensive benchmarking and systematic evaluation of sentiment analysis techniques specifically tuned for cryptocurrency social media data. The contribution of this paper lies in its development of a machine learning model that merges a cryptocurrency-specific lexicon with cutting-edge text analysis techniques. This approach seeks to fill the existing gap in sentiment analysis tools by providing a more accurate and context-aware mechanism for evaluating sentiment among cryptocurrency traders and social media discussion.

Our research contributes to this growing literature by focusing specifically on the sentiment dynamics of cryptocurrency markets. We develop a machine learning-based sentiment analysis model designed to accurately identify the sentiment expressed in cryptocurrency traders' social media posts. This is achieved by incorporating human-labeled and Chat-GPT 3.5 labeled training data, allowing the model to learn the unique language patterns and context of cryptocurrency discourse. Our study extends the existing literature by integrating Granger causality (GC) and Time-Varying Parameter Vector Autoregression (TVP-VAR) methodologies to analyze the impact of sentiment on cryptocurrency price volatility. The GC approach tests whether lagged social media sentiment can predict current cryptocurrency price volatility, while the TVP-VAR method offers dynamic insights into the interconnectedness and directional influence of sentiment and market variables over time. This dual approach provides a robust framework for understanding the bidirectional relationship between sentiment and market volatility, capturing both immediate and evolving impacts. By blending our contributions and findings with previous literature, we highlight the novel application of sentiment analysis in the cryptocurrency market. Our development of a specialized sentiment lexicon tailored to cryptocurrency discourse and the advanced evaluation of our model against established tools like Vader and TextBlob demonstrates a significant improvement in accuracy. Achieving an 88.7% accuracy rate, our model sets a new standard for sentiment analysis in this field.

This research aims to fill a key gap in the current literature by providing investors and researchers with a valuable tool to effectively measure market sentiment. It aims to provide a detailed understanding of how sentiment expressed through social media platforms affects volatility and trends in the cryptocurrency market. In turn, this can facilitate more informed decision-making, risk assessment, and strategic planning for stakeholders in the cryptocurrency space. Furthermore, the research aims to contribute to the field of sentiment analysis by adapting and refining machine learning techniques to address the unique challenges posed by the cryptocurrency market.

2. Data

In order to collect the data we need, we use the web scrapping (specifically web crawling) as our primary tool. This allows us to delve into online platforms and extract valuable information. Our research covers a time span of January 1, 2018 to June 30, 2021. This period is full of events, market changes and evolving public opinion, making it a prime timeframe for our research. Reddit becomes one of our primary sources, and we delve into various Reddit subreddits to get different perspectives. Straight away, r/bitcoin and r/btc are our top choices. Given the growing importance of ethereum in the cryptocurrency conversation, we also look at r/eth. But to gain a broader understanding, we turn to r/cryptomarkets and r/Cryptocurrency. These Reddit subreddits are like town squares for the cryptocurrency community on Reddit. They help us understand the general feelings and attitudes of the Reddit community towards the crypto world, and serve as mirrors that reflect the collective sentiment.

Twitter, on the other hand, provides real-time reactions to the changes in cryptocurrency market. We continue to collect tweets that mention "bitcoin" and "btc". The platform is heavily discussed on a daily basis, providing a snapshot of what is going on in the cryptocurrency sentiment space on a day-to-day basis. In our collected dataset, we observe a large number of irrelevant entries, including placeholder messages such as "[deleted]" and "[removed]", as well as automated responses such as "[deleted]" and "[removed]", as well as automated responses like "I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.". To ensure data integrity and relevance for our analysis, we carefully cleaned it to remove these irrelevant entries. The generated dataset represents the refined usable data for our study. The cleaned data for both platforms is shown below (see Tables 1 and 2):

In our research, we focus on Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), and Dogcoin (DOGE) based on their roles played in the cryptocurrency market from 2018 to 2021. We obtained 30 min of trading data from: [firstratedata.com](https://www.firstratedata.com/), the leading provider of high-resolution intraday stock market, cryptocurrency, futures, and FX rate data. Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP) have huge market capitalizations and represent a wide range of technologies, adoption levels and investor interest. Despite its smaller market capitalization, we have chosen Dogcoin (DOGE) as a representative of the meme coins, which have gained significant cultural significance and influence in recent years. This selection provides a comprehensive view of the cryptocurrency market, including both major, established cryptocurrencies and novel, culturally influential cryptocurrencies.

¹ For more details, please visit: <https://chat.openai.com/>

Table 1
Reddit subchannels.

Subchannel	Number of Posts
r/bitcoin	3,653,207
r/btc	1,496,968
r/eth	738,678
r/cryptocurrency	7,998,010
r/cryptomarkets	269,937
Total	14,156,800

Table 2
Twitter Bitcoin&BTC Tweets.

Glossary	Number of Tweets
Bitcoin, Btc	55,889,578

Table 3
Descriptive statistics.
Rolling volatility measures and sentiment indicators.

	BTC Rolling Volatility	ETH Rolling Volatility	XRP Rolling Volatility	DOGE Rolling Volatility	Reddit Sentiments	Twitter Sentiments	All Sentiments
count	61 233	61 233	61 233	61 233	61 233	61 233	61 233
mean	0.029476	0.041053	0.058547	0.183936	0.101094	0.296862	0.261853
std	0.020728	0.024299	0.044393	0.530523	0.078412	0.072487	0.065066
min	0.003704	0.006216	0.005844	0.000000	-1.000000	0.000000	0.050691
25%	0.015551	0.025490	0.026760	0.000000	0.054455	0.233010	0.208411
50%	0.023921	0.034853	0.043131	0.000000	0.101010	0.308804	0.265345
75%	0.037401	0.049508	0.078045	0.073985	0.148746	0.348259	0.304749
max	0.265191	0.321201	0.381886	4.350546	1.000000	0.891304	0.750175

Notes: (1) Rolling volatility is measured at 30 min intervals for BTC, ETH, XRP, and DOGE.
(2) *rd_sentiments* and *tw_sentiments* capture sentiment data from Reddit and Twitter, respectively.
(3) *senti_all_sentiments* is an aggregated measure derived from both Reddit and Twitter sentiment signals.

For analytical purposes, we determine the frequency of price data to be 30 min. We carefully choose this timeframe to be consistent with our goal of studying the interaction between market sentiment and cryptocurrency market volatility. The 30 min interval is short enough to capture rapid changes in market sentiment, which is critical in the highly volatile and sentiment-driven cryptocurrency market. Compared to shorter time intervals such as 1, 5 or 10 min, 30 min data smooths out minute-by-minute price fluctuations, providing a clearer picture of market trends and sentiment changes. It also balances the need for a detailed view of intraday price movements with the practicality of managing a comprehensive data set for extensive statistical analysis. It is suitable for both short-term traders, who need intraday volatility insights, and long-term investors, who monitor broader price movements.

Table 3 presents the summary statistics for our main variables of interest. These include rolling volatility measures for four major cryptocurrencies (BTC, ETH, XRP, DOGE) and sentiment indicators (Reddit, Twitter, and aggregated sentiment).

3. Methodology

3.1. Sentiment analysis

In sentiment analysis, interpreting user-generated text is crucial. This study develops machine learning models to predict sentiment labels in financial market, and three categories of sentiment labels are considered, including positive, neutral and negative sentiment. This is a typical problem of multi-class classification. A feature based approach is adopted, which extracts key features from text data and then trains machine learning algorithms to predict sentiment. Logistic regression, random forest and XGBoost are chosen due to their superior performances in multi-class classification and computational efficiency, both essential factors for fast and accurate sentiment analysis. This subsection outlines the comprehensive steps involved in formulating the dataset used for training machine learning models to predict sentiment labels in financial market communications. The methodology is divided into two main parts: (1) the construction of the dataset itself, and (2) the processing required to prepare the data for machine learning models.

3.1.1. Training and testing dataset construction

To facilitate an effective evaluation of the selected machine learning models, it is imperative to build robust training and test datasets tailored for sentiment analysis. In this research, sentiments have been categorized into three distinct labels: positive, neutral, and negative. The process of dataset creation encompassed a series of methodological steps:

a. Noisy Labeling using Emojis: To begin with, a massive volume of 100,000 social media messages was processed using the 'noisy labeling' method. Noisy labeling, in this context, refers to the practice of determining the sentiment of a message based on

the emojis it contains. Emojis, due to their expressive nature, can often provide a heuristic to gauge the sentiment of a text. For instance, messages containing the “😊” emoji were earmarked as ‘positive’, whereas those with the “😞” emoji were designated as ‘negative’. In cases where a single message incorporated both positive and negative emojis, the sentiment was determined based on the majority presence. That is, the sentiment label would align with the emoji that appeared more frequently within that message.

b. Manual Annotation by Human Annotators: Subsequently, a subset of 10,000 social media messages was carefully annotated by a team of eight human annotators. These annotators, specifically hired for this task, underwent a comprehensive training schedule to ensure consistency and accuracy in their annotations. Their primary role was to read and categorize each message into one of the three sentiment labels.

c. Annotation by ChatGPT: Parallely, the identical set of 10,000 messages, previously annotated manually, was also processed by ChatGPT for sentiment labeling. This step aimed to juxtapose human intuition with the capabilities of an advanced language model in sentiment classification.

d. Segregation of Human-Annotated Data: From the corpus of 10,000 messages that were manually annotated, a subset of 3000 was curated and reserved as a test set. This test set would serve as a benchmark to evaluate the performance of machine learning models. The remaining 7000 messages from this corpus are designated as the training set, forming the foundation upon which the models would be trained.

e. Incorporation of ChatGPT-Annotated Data: From the 10,000 messages labeled by ChatGPT, the same 3000 messages earmarked in the previous step were harnessed to augment the training set. This strategy aimed to enrich the training dataset by introducing varied labeling perspectives. The residual 7000 messages, processed by ChatGPT, are integrated into the primary training set. This study combines human judgments and algorithmic accuracy to build a detailed dataset, laying the groundwork for thorough machine learning analysis.

3.1.2. Machine learning model development

The objective of this phase is to transform raw message data into actionable insights using a structured machine-learning pipeline. This involved a series of detailed steps designed to enhance the quality of the data and subsequently predict sentiments.

a. Pre-processing of Messages: The initial step was the refinement of the raw messages. This entailed several processes designed to enhance the clarity and relevance of the data. Symbols, numbers, and common words (often referred to as ‘stop words’) that lack significant semantic value were systematically eliminated. Subsequent to this cleansing process, the messages are tokenization. Tokenization, in this context, means breaking down the text into smaller units, such as individual words or combinations of words. These units were categorized into unigrams (single words), bigrams (two-word combinations), and trigrams (three-word combinations).

b. Text Representation and Feature Extraction: Once the messages were pre-processed, the next task was to represent the text in a manner conducive to machine learning. Two predominant techniques were employed for this: the ‘Bag of Words’ (BoW) model and the ‘Term Frequency-Inverse Document Frequency’ (TF-IDF). While BoW represents text data based on the frequency of words without considering their order, TF-IDF assigns importance to words based on how frequently they appear in a document relative to their occurrence across other documents. Both techniques transform textual data into numerical vectors, making them suitable for machine learning algorithms.

c. Deployment of Machine Learning Models: With the vectorized text data at hand, the next step is to feed this data into three distinct machine-learning models: Logistic Regression, Random Forest, and XGBoost. Each model was tasked with predicting the sentiments of the text messages. To optimize their performance, hyperparameter tuning was conducted using 5-fold cross-validation on the training dataset. This means the training data was divided into five subsets; each subset was used as a validation set while the others formed the training set. This process was repeated five times, ensuring comprehensive model evaluation. All three models were equipped to process both BoW and TF-IDF representations of unigrams, bigrams, and trigrams. From a vast array of over 160,000 potential units (unigrams, bigrams, and trigrams), a cap was set, and only the 30,000 most prevalent features (vectorized social messages) were chosen for model training.

d. Weighted Importance of Datasets: An innovative step was taken to assign differential importance to datasets. The datasets labeled by both ChatGPT and human annotators have explained a weight ten times greater than their original value. This strategic amplification was implemented to underscore their importance, especially when the size of the datasets labeled through noisy labeling methods is much greater than the human label. This approach was based on the premise that manually annotated and ChatGPT-labeled datasets exhibit higher reliability than those labeled using heuristic methods like noisy labeling.

3.1.3. Further adjustment

The result of the model is the average of the results of 10 random draws, strictly ensuring the mutual exclusivity between the data sets drawn each time. This approach was implemented to enhance the robustness of the findings, minimize potential bias, and ensure the generalizability of the results. After selecting the optimal model, an additional post-processing step is added to refine the assignment of emotion labels. This step involves calculating ratios to identify the dominance of a particular emotion. Specifically, if the ratio of the probability of being positive to the sum of the probabilities of being neutral and negative exceeds 1, the message is explicitly categorized as “positive”. Conversely, if the probability of negativity divided by the probability of the combination of positivity and neutrality exceeds 1, the message is explicitly categorized as “negative”. This approach ensures that sentiment labels are assigned only when there is a clear dominance of evidence in favor of the sentiment, thus improving the accuracy of the sentiment classification process.

$$\text{Categorize as "positive"} \Leftrightarrow \frac{P(\text{positive})}{P(\text{neutral}) + P(\text{negative})} > 1. \quad (1)$$

$$\text{Categorize as "negative"} \Leftrightarrow \frac{P(\text{negative})}{P(\text{positive}) + P(\text{neutral})} > 1. \quad (2)$$

In this approach, after the best performing model is selected, SHAP (SHapley Additive exPlanations) analysis is implemented for complex feature extraction. SHAP, originating from a game theory background, provides a sophisticated way to interpret the output of any machine learning model (Lundberg & Lee, 2017). This is done by assigning a specific predicted importance value to each feature, based on the concept of the Shapley value from cooperative game theory.

Using SHAP analysis, the state-of-the-art approach drills down into the model's output, providing detailed insights into how each feature contributes to the overall decision-making process. Characteristics are carefully divided into three different emotion groups: positive, negative, and neutral. This classification helps reveal the complex interplay between various types of emotions and their impact on cryptocurrency market communication. It allows the subtle ways in which different emotions influence market dynamics and investor behavior to be understood. To further refine the analysis, manual adjustments are performed. This critical step improves the accuracy of the approach and allows the sentiment dictionary to be tailored to work seamlessly with the complexity and unique characteristics of cryptocurrency market communications. By customizing this lexicon, market sentiment is ensured to be more accurately expressed and understood, capturing the volatile and often unpredictable nature of the cryptocurrency market. The refined sentiment lexicon can be accessed and downloaded from the following URL: <https://shorturl.at/iELS6>.

3.2. Granger causality and TVP-VAR

3.2.1. Granger causality

After constructing the sentiment variable, we analyze the impact of sentiment on cryptocurrency price volatility using the standard Granger causality approach (Granger, 1969). Granger causality testing is a statistical hypothesis test used to determine whether one time series helps to predict another. Essentially, we will test whether the lagged value of social media sentiment significantly explains the current value of cryptocurrency price volatility.

Specifically, if we denote cryptocurrency as P_t and social media sentiment as S_t , then the VAR model can be specified as follows:

$$P_t = \alpha + \sum_{i=1}^n \beta_i P_{t-i} + \sum_{i=1}^n \gamma_i S_{t-i} + \epsilon_t \quad (3)$$

$$S_t = \delta + \sum_{i=1}^n \phi_i S_{t-i} + \sum_{i=1}^n \theta_i P_{t-i} + \mu_t \quad (4)$$

where n is the number of lags, ϵ_t and μ_t are error terms, and $\alpha, \beta_i, \gamma_i, \delta, \phi_i, \theta_i$ are parameters to be estimated.

The null hypothesis for the Granger causality test from social media sentiments to cryptocurrencies' price volatility is that the coefficients of the lagged values of S_t (γ_i) in the first equation are jointly zero. If this hypothesis is rejected, it implies that social media sentiments Granger-cause cryptocurrencies' prices. We also test Granger causality from cryptocurrencies' prices volatility to social media sentiments. Both results are presented in Fig. 1.

3.2.2. TVP-VAR

In order to capture the dynamic relationship between the variables under study, we follow the approach in Long, Chatziantoniou, Gabauer, and Lucey (2024) and employ a 2nd order time-varying parameter vector autoregression (TVP-VAR(2)) model. The model allows the coefficients to vary over time, providing a flexible structure for identifying potential changes in the relationship. The model is defined as follows:

$$\mathbf{x}_t = \Phi_t \mathbf{x}_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, \Sigma_t) \quad (5)$$

$$\text{vec}(\Phi_t) = \text{vec}(\Phi_{t-1}) + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R}_t) \quad (6)$$

The sum of coefficients is a time-varying matrix and the error term follows a multivariate normal distribution with a time-varying covariance matrix.

To interpret the effects of shocks on the inter-variable, we use the GFEVD method, which was introduced by Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998). This method is rooted in the Wald representation theorem and requires transforming our TVP-VAR(2) into a TVP-VMA representation:

$$\mathbf{x}_t = \sum_{j=0}^{\infty} \Psi_{jt} \epsilon_{t-j} \quad (7)$$

The GFEVD is then calculated to understand the implications of a shock in variable j on variable i :

$$\theta_{ijt}(H) = \frac{(\Sigma_t)^{-1} \sum_{h=0}^H ((\Psi_{ht} \Sigma_t)_{ijt})^2}{\sum_{h=0}^H (\Psi_{ht} \Sigma_t \Psi'_{ht})_{ii}} \quad (8)$$

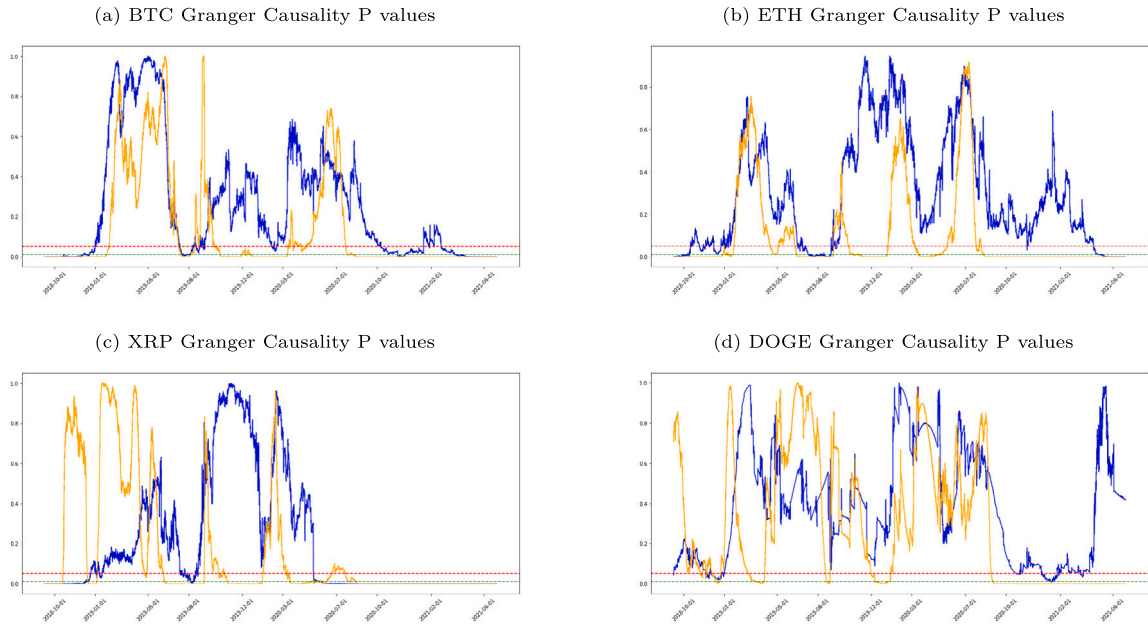


Fig. 1. Granger causality test results - Volatility to sentiments.

Note: The figure above displays the p -values obtained from the Granger causality test conducted on rolling windows of 250 days. The test aims to determine the causality between sentiments and the volatility of various cryptocurrency coins. Each point on the graph represents the p -value for a specific 250-day window, with the window moving forward in time across the dataset. The yellow line represents the p -values of the Granger causality test transitioning from sentiments to cryptocurrency volatilities. Conversely, the blue line illustrates the p -values from the Granger causality test flowing from cryptocurrency volatilities to social media sentiments. Lines indicating significance levels at 0.05 (red) and 0.01 (green) are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\tilde{\theta}_{ijt}(H) = \frac{\theta_{ijt}(H)}{\sum_{i=1}^N \theta_{ijt}(H)} \quad (9)$$

Following [Wiesen, Beaumont, Norrbom, and Srivastava \(2018\)](#), the GFEVD is preferred over its orthogonal counterpart, given its invariance to variable ordering.

Connectedness measures provide insights into the interdependence of variables. We compute these measures in both time and frequency domains. In the time domain:

$$NPDC_{ijt}(H) = \tilde{\theta}_{ijt}(H) - \tilde{\theta}_{jit}(H) \quad (10)$$

$$TO_{it}(H) = \sum_{i=1, i \neq j}^N \tilde{\theta}_{ijt}(H) \quad (11)$$

$$FROM_{it}(H) = \sum_{j=1, j \neq i}^N \tilde{\theta}_{ijt}(H) \quad (12)$$

$$NET_{it}(H) = TO_{it}(H) - FROM_{it}(H) \quad (13)$$

$$TCI_t(H) = \frac{N}{N-1} \sum_{i=1}^N TO_{it}(H) \quad (14)$$

Extending to the frequency domain, we employ the spectral decomposition method ([Stiassny, 1996](#)):

$$\theta_{ijt}(\omega) = \frac{(\Sigma_t)_{jj}^{-1} \left| \sum_{h=0}^{\infty} (\Psi_t(e^{-i\omega h}) \Sigma_t)_{ijt} \right|^2}{\sum_{h=0}^{\infty} (\Psi_t(e^{-i\omega h}) \Sigma_t \Psi_t(e^{i\omega h}))_{ii}} \quad (15)$$

$$\tilde{\theta}_{ijt}(\omega) = \frac{\theta_{ijt}(\omega)}{\sum_{k=1}^N \theta_{ijt}(\omega)} \quad (16)$$

Utilizing the connectedness measures detailed in [Diebold and Yilmaz \(2012, 2014\)](#), we focus on frequency connectedness, capturing spillovers within a specific frequency range d . To account for the overarching impact, [Baruník and Křehlík \(2018\)](#) recommends weighting these measures relative to the entire system. Finally, the relationship between time-domain and frequency-domain measures can be represented as follows.

$$NPDC_{ijt}(H) = \sum_d NPDC_{ijt}(d) \quad (17)$$

Table 4

Performance with different training data combinations (Initial Datasets).

The table shows macro-F1 scores for three classifiers (RF, XGBoost, Logistic Regression) trained on various mixes of noisy-labeled, human-labeled, and ChatGPT-labeled data. Scores are evaluated on the same 3000-message human-labeled test set.

		Noisy 50k + Human 7k + ChatGPT 0k	Noisy 50k + Human 3k + ChatGPT 4k	Noisy 50k + Human 0k + ChatGPT 7k
BoW	Random Forest	0.249	0.276	0.370
	XGBoost	0.408	0.470	0.522
	Logistic Regression	0.479	0.502	0.526
TF-IDF	Random Forest	0.274	0.292	0.368
	XGBoost	0.536	0.536	0.538
	Logistic Regression	0.606	0.585	0.574

Notes: (1) BoW = Bag-of-Words representation.

(2) TF-IDF = Term Frequency-Inverse Document Frequency representation.

(3) All values are macro-F1 scores, the harmonic mean of precision and recall across classes.

(4) “Noisy 50k” refers to data labeled automatically by a distant labeling method.

(5) “Human 7k” or “Human 3k” indicates the number of additional messages fully annotated by human experts.

(6) “ChatGPT 7k” or “ChatGPT 4k” indicates the number of additional messages labeled by ChatGPT.

$$TO_{it}(H) = \sum_d TO_{it}(d) \quad (18)$$

$$FROM_{it}(H) = \sum_d FROM_{it}(d) \quad (19)$$

$$NET_{it}(H) = \sum_d NET_{it}(d) \quad (20)$$

$$TCI_{it}(H) = \sum_d TCI_{it}(d) \quad (21)$$

Our analysis presents in a Connectedness Table, presenting the interdependence of variables over different time frames.

4. Results and discussion

4.1. Sentiment analysis model performance and lexicon

In the first set of experiments, different sets of training data were created by combining 50,000 noisy labeled datasets with a 7k random sampling of ChatGPT labeled dataset and manually labeled dataset. Firstly, a training dataset comprising 50,000 noisily labeled data supplemented by 7000 entries of human annotations. The subset of 3000 manually labeled messages was used as a unseen test set, as discussed in Section 3.1.1. The model is evaluated by its F1 score. The F1 score is a measure used to evaluate the accuracy of a classification model, which combines precision and recall into a single metric. High F1 scores are generally desirable, indicating that the model has good precision and recall balance. As shown in Table 4, for the Bag-of-Words (BoW) text representation, the models exhibit the following F1 scores: Random Forest at 0.249, XGBoost at 0.408, and Logistic Regression at 0.479. Switching to the TF-IDF representation, the F1 scores improve across the board, with Random Forest achieving 0.274, XGBoost reaching 0.536, and Logistic Regression leading the pack with 0.606.

Secondly, a dataset with 3000 human-labeled and 4000 ChatGPT-labeled entries were merged into the noisy 50,000 set. The BoW representation sees increases in F1 scores to 0.276 for Random Forest, 0.470 for XGBoost, and 0.502 for Logistic Regression. The TF-IDF representation scores for this mix are 0.292 for Random Forest, 0.536 for XGBoost, and 0.585 for Logistic Regression. Thirdly, replacing the human labels entirely with 7000 ChatGPT-labeled entries in the noisy 50,000 set, the BoW representation delivers a Random Forest F1 score of 0.370, XGBoost at 0.522, and Logistic Regression at 0.526. For the TF-IDF representation, the results show a Random Forest F1 score of 0.368, XGBoost at 0.538, and Logistic Regression at 0.574.

4.2. Performance analysis with enriched datasets

Building on the first experiment, each dataset configuration was enriched by incorporating additional labeled data. In the second set of experiments, the training datasets were augmented by incorporating an additional dataset consisting of 3000 unseen messages and their corresponding sentiment labels predicted by ChatGPT. This enrichment aimed to explore if the inclusion of ChatGPT-labeled data could enhance the predictive performance of the machine learning models. In the enriched BoW representation, as shown in Table 5, the F1 score of the Random Forest model improved from 0.249 (initial dataset) to 0.276 with the partial replacement of human-labeled datasets by ChatGPT-labeled datasets, and further to 0.370 with the complete replacement of human-labeled datasets by ChatGPT-labeled datasets. XGBoost's F1 score increased from 0.408 (initial dataset) to 0.470 with the partial replacement, and then to 0.522 with the complete replacement. The Logistic Regression model's F1 scores progressed from 0.479 (initial dataset) to 0.502 with the partial replacement, and then to 0.526 with the complete replacement. In the enriched TF-IDF

Table 5

Performance with Additional ChatGPT-Predicted Data (Enriched Datasets).

Macro-F1 scores for three classifiers trained on the same noisy/human/ChatGPT mixes as in Table 4, but now enriched with an additional 3000 ChatGPT-labeled messages ("X_{rtest}").

		Noisy 50k + Human 7k + ChatGPT 0k + X _{rtest}	Noisy 50k + Human 3k + ChatGPT 4k + X _{rtest}	Noisy 50k + Human 0k + ChatGPT 7k + X _{rtest}
BoW	Random Forest	0.245	0.297	0.472
	XGBoost	0.528	0.557	0.600
	Logistic Regression	0.661	0.658	0.665
TF-IDF	Random Forest	0.232	0.286	0.453
	XGBoost	0.581	0.579	0.582
	Logistic Regression	0.670	0.674	0.676

Notes: (1) *BoW* = Bag-of-Words representation.(2) *TF-IDF* = Term Frequency-Inverse Document Frequency representation.(3) *X_{rtest}* refers to an additional dataset of 3000 unseen messages labeled by ChatGPT.

(4) All scores are macro-F1 on the same 3000-message human-labeled test set used previously.

representation, the F1 score of the Random Forest model slightly increased from 0.274 (initial dataset) to 0.292 with the partial replacement, and then further increased to 0.368 with the complete replacement. XGBoost's scores remained relatively consistent, ranging from 0.536 (initial dataset) to 0.538 with the complete replacement. However, the Logistic Regression model's performance decreased slightly from 0.606 (initial dataset) to 0.585 with the partial replacement, and then further decreased to 0.574 with the complete replacement.

The optimal model performance is observed with the Logistic Regression model using the TF-IDF text representation. By leveraging a training set that excludes manually labeled data and utilizes 7000 ChatGPT-labeled entries alongside 3000 ChatGPT-predicted unseen labels, Logistic Regression yields the highest F1 score of 0.676. This superior result demonstrates the strength of Logistic Regression in sentiment analysis, particularly when the model is trained on a dataset rich with ChatGPT's algorithmic annotations and predictions. In addition, we assess our sentiment critically against two established tools, Vader (Hutto & Gilbert, 2014) and TextBlob,² using 2000 messages. analyzing randomly selected, previously unseen messages from machine learning models. These messages are first manually annotated to establish basic facts. Our model significantly outperforms the others with an accuracy of 88.7%, compared to 42.85% for Vader and 43.15% for TextBlob. This large accuracy gap not only emphasizes the effectiveness of our model in sentiment classification, but also sets a new standard for sentiment analysis by improving the reliability and accuracy of sentiment extraction from text data. Furthermore, we have developed a specialized sentiment analysis lexicon for the cryptocurrency market. It reflects Loughran and McDonald (2011)'s approach and includes various cryptocurrency-specific terms in various grammatical forms, such as lexical variations and variants. Covering more than 10,000 terms, this comprehensive lexicon is essential for accurately capturing the emotional detail of cryptocurrency-related texts, as well as their different grammatical forms. Its inclusivity ensures a detailed understanding of emotions, from basic forms to tenses and adjectival variants. To clearly demonstrate its utility, a sample of the dictionary is included in the appendix of our paper (see Table 9). The lexicon is an indispensable tool for deep sentiment analysis, catering to the diverse texts within the cryptocurrency domain and significantly improving the accuracy of sentiment assessments.

4.3. Network volatility dynamics and influence of market players

The relationship between market sentiment and cryptocurrency volatility is analyzed, revealing interesting dynamics as evidenced by the Granger causality test. The test plays a key role in understanding the predictive power of past sentiment values on current market volatility, which varies across cryptocurrencies. In some cases, the *p*-value is below the 0.01 critical value, indicating a statistically significant relationship between past sentiment values and current market volatility. However, in other cases, the *p*-value is found close to 1, indicating that past sentiment is a very small predictor of current market volatility. This pattern is consistent across cryptocurrencies.

The Granger causality test results are shown in Fig. 1. The figure below display the *p*-values obtained from the Granger causality test conducted on rolling windows of 250 days. The test aims to determine the causality between sentiments and the volatility of various cryptocurrency coins. Each point on the graph represents the *p*-value for a specific 250-day window, with the window moving forward in time across the dataset. Lines indicating significance levels at 0.05 (red) and 0.01 (green) are also shown.

Several factors are attributed to the volatile nature of these *p*-values. External shocks such as regulatory changes, technological advances, macroeconomic events, or geopolitical crises are found to have a considerable impact on the link between sentiment and market volatility. Another characteristic identified in the cryptocurrency market is the variation of volatility over time, swinging between peaks and valleys. The impact of sentiment on market volatility is seen to vary across time, necessitating the introduction of time-varying parameter vector autoregression (TVP-VAR) in the analysis. The results of the reverse Granger causality test are

² <https://textblob.readthedocs.io/en/dev/>

Table 6

Averaged total positive (negative) dynamic connectedness table.

	BTC	ETH	XRP	DOGE	Sentiments	FROM
BTC	48.37 (51.03)	27.35 (26.11)	17 (15.78)	4.52 (5.42)	2.76 (1.66)	51.63 (48.97)
ETH	29.95 (24.44)	44.09 (48.65)	19.27 (20.23)	4.04 (4.95)	2.66 (1.73)	55.91 (51.35)
XRP	15.91 (10.74)	14.91 (13.11)	61.7 (67.05)	4.42 (5.58)	3.06 (3.52)	38.3 (32.95)
DOGE	5.48 (5.95)	3.95 (3.59)	8.04 (8.11)	81.15 (81.45)	1.38 (0.9)	18.85 (18.55)
Sentiments	5.05 (2.51)	4.8 (3.02)	5.66 (3.46)	2.27 (1.73)	82.22 (89.29)	17.78 (10.71)
TO	56.39 (43.64)	51.01 (45.83)	49.96 (47.57)	15.25 (17.68)	9.86 (7.81)	182.47 (162.53)
Inc.Own	104.76 (94.67)	95.1 (94.48)	111.66 (114.62)	96.41 (99.13)	92.08 (97.09)	TCI
NET	4.76 (−5.33)	−4.9 (−5.52)	11.66 (14.62)	−3.59 (−0.87)	−7.92 (−2.91)	36.49(32.51)
NPT	3 (2)	1 (2)	4 (3)	2 (2)	0 (1)	

Notes: Results are based on a TVP-VAR model with a lag length of order two (BIC) and a 100-step ahead generalized forecast error variance decomposition. Values represent positive and negative - in parentheses - connectedness measures, respectively.

Table 7

Averaged long-term positive (negative) dynamic connectedness table.

	BTC	ETH	XRP	DOGE	Sentiments	FROM
BTC	45.02 (43.65)	30.44 (30.16)	16.82 (18.08)	4.04 (5.7)	3.67 (2.41)	54.98 (56.35)
ETH	30.63 (25.44)	41.48 (43.48)	17.56 (19.22)	5.08 (7.36)	5.25 (4.5)	58.52 (56.52)
XRP	20.28 (15.52)	20.13 (19.7)	50.04 (54.56)	5.57 (5.99)	3.98 (4.23)	49.96 (45.44)
DOGE	6.22 (8.69)	7.05 (8.79)	9.56 (8.06)	70.23 (66.18)	6.95 (8.28)	29.77 (33.82)
Sentiments	5.25 (4.54)	7.1 (4.73)	4.58 (3.9)	11.53 (16.34)	71.53 (70.49)	28.47 (29.51)
TO	62.38 (54.19)	64.72 (63.38)	48.53 (49.25)	26.23 (35.39)	19.85 (19.43)	221.7 (221.65)
Inc.Own	107.4 (97.84)	106.2 (106.86)	98.57 (103.81)	96.45 (101.57)	91.38 (89.91)	cTCI/TCI
NET	7.4 (−2.16)	6.2 (6.86)	−1.43 (3.81)	−3.55 (1.57)	−8.62 (−10.09)	55.42(55.41)/44.34(44.33)
NPT	4 (2)	3 (4)	2 (1)	1 (1)	0 (1)	

Notes: Results are based on a TVP-VAR model with a lag length of order two (BIC) and a 100-step ahead generalized forecast error variance decomposition. Values represent positive and negative - in parentheses - connectedness measures, respectively.

consistent with the results of the original test, suggesting the existence of a feedback loop, i.e., market volatility affects sentiment, which feeds back into market volatility. This dynamic interaction evolves over time, creating a periodic effect whereby positive sentiment may drive an increase in trading activity, thereby increasing market volatility. Increased volatility in turn affects market sentiment, either reinforcing the pre-existing sentiment or generating the opposite sentiment. The observed patterns suggest that market participants not only react to price volatility but also influence future price movements through collective sentiment.

The average connectedness tables (Table 6) provide a comprehensive overview of the dynamics over different time periods, rather than focusing on individual events. The Total Connectedness Index (TCI) shows moderate connectivity among the network variables, but there is a clear asymmetry, i.e., positive shocks are more strongly connected than negative shocks. Regardless of the nature of the shocks, the division of variables into net receivers and net transmitters remains consistent. Sentiment, Ethereum (ETH), and dogecoin (DOGE) are the main net receivers, while Ripple (XRP) and Bitcoin (BTC) are mainly net transmitters of positive and negative shocks.

The long-term connectedness, indicated in Table 7, shows a minimal degree of TCI asymmetry, whereas the short-term connectedness, presented in Table 8, reveals more pronounced asymmetry. These results highlight the varying impact of shocks on the network in the long and short term. It is noteworthy that the classification into net recipients and net transmitters in the long-term and short-term TCI deviates slightly from the results in Table 6, with Ethereum (ETH) assuming a net transmitter role in both cases. However, sentiment remains a significant net recipient across all frequencies.

The dynamic connectedness results are instrumental in understanding the evolution of connectedness over time, particularly in response to disruptive events. Fluctuations in connectedness strength often indicate strong co-movement among network variables, hinting at potential contagion dynamics. The analysis of TCI, depicted in Fig. 2, reveals considerable variations over time, with total connectedness being sensitive to external shocks and specific events. Notable peaks in total connectedness were observed at the end of 2018 and mid-2019, and in early 2020, both positive and negative connectedness values remained high, indicating strong co-movement and contagion dynamics.

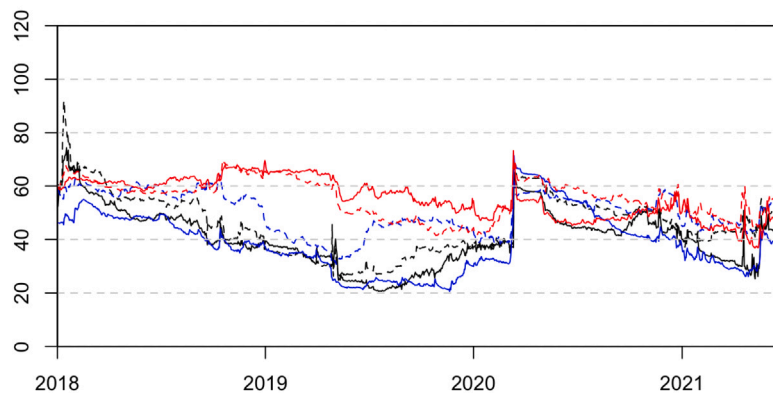
The asymmetry in time-domain connectedness is evident, with long-term negative connectedness values dominating until early 2020. Positive connectedness briefly exceeds negative values during the COVID-19 outbreak, but the difference between positive and negative connectedness remains small throughout 2020. Towards the end of the sample, positive correlations regain their dominance, reflecting the typical risk-averse nature of investors. When positive perceptions of market risk outweigh negative perceptions, positive correlation metrics outweigh negative metrics, and vice versa. The frequency-domain assessment further clarifies the dominance of low-frequency (long-term) dynamics in the initial period, with the exception of early- to mid-2020. By late 2020 and early 2021, low-frequency connectedness again dominates, suggesting that the effects of structural changes in the market are long-lasting. The pattern of asymmetry in the frequency domain is consistent with the pattern of asymmetry in total connectedness, as shown in Fig. 2. The directional connectedness, represented in Fig. 3, reveal shifts between net-transmitting and net-receiving

Table 8

Averaged short-term positive (negative) dynamic connectedness table.

	BTC	ETH	XRP	DOGE	Sentiments	FROM
BTC	45.32 (58.3)	29.29 (25.07)	19.42 (10.3)	2.76 (4.07)	3.22 (2.26)	54.68 (41.7)
ETH	26.61 (24.02)	48.41 (54.3)	20.23 (15.05)	2.7 (2.94)	2.05 (3.69)	51.59 (45.7)
XRP	16.89 (10.79)	18.82 (16.45)	57.49 (66.83)	3.58 (2.38)	3.21 (3.55)	42.51 (33.17)
DOGE	5.68 (6.02)	5.58 (4.41)	8.23 (4.93)	77.22 (83.15)	3.28 (1.49)	22.78 (16.85)
Sentiments	7.22 (3.28)	7.43 (4.97)	10.6 (6.68)	3.67 (3.29)	71.08 (81.77)	28.92 (18.23)
TO	56.41 (44.12)	61.12 (50.9)	58.48 (36.96)	12.71 (12.68)	11.77 (10.99)	200.5 (155.64)
Inc.Own	101.73 (102.42)	109.53 (105.2)	115.97 (103.8)	89.92 (95.83)	82.85 (92.76)	cTCL/TCI
NET	1.73 (2.42)	9.53 (5.2)	15.97 (3.8)	-10.08 (-4.17)	-17.15 (-7.24)	50.12 (38.91)/40.10 (31.13)
NPT	2 (3)	3 (4)	4 (2)	1 (1)	0 (0)	

Notes: Results are based on a TVP-VAR model with a lag length of order two (BIC) and a 100-step ahead generalized forecast error variance decomposition. Values represent positive and negative - in parentheses - connectedness measures, respectively.

**Fig. 2.** Total, long-term and short-term connectedness.

Note: Results are based on a TVP-VAR model with a lag length of order two (BIC) and a 100-step ahead generalized forecast error variance decomposition. Black, red, and blue lines stand for the total, long-term, and short-term connectedness measures, respectively. Solid lines represent negative connectedness measures while dashed lines represent positive connectedness measures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

roles among various cryptocurrencies and sentiment. Sentiment is consistently subject to shocks, suggesting that it is vulnerable to cryptocurrency market dynamics. XRP, primarily a net transmitter, shifts to a net-receiving role around the end of 2018 to early 2019, suggesting that fundamental changes influenced this transition. Bitcoin (BTC) exhibits fluctuations between roles until the end of 2020, with high-frequency dynamics becoming more prevalent after 2021. Ethereum assumes a net transmitting role until mid-2020, dominates by long-term dynamics, and then transitions to a net recipient role, indicative of quick shock absorption. Dogecoin fluctuates between roles until the end of 2020 and becomes a net transmitter after that, with low-frequency dynamics becoming more prevalent after 2021.

The influence of ‘whales’ in the cryptocurrency market is a significant aspect of this study. Whales, typically defined as entities or individuals holding large quantities of cryptocurrency (notably those holding over 1000 Bitcoin³), can have a profound impact on market dynamics. Their trading behaviors, such as large-scale buying or selling, can substantially alter market supply and demand, leading to notable price shifts. Additionally, given their influential position within the cryptocurrency community, their public sentiment, especially when shared via social media platforms, can influence overall market sentiment. The interconnectedness between whale activities and market sentiments is specifically investigated in this analysis. By utilizing a connectedness measure marked by periods of increased whale activity, a correlation between these activities and fluctuations in market connectedness is observed (see Fig. 4).

This suggests that whale actions are closely linked to shifts in market sentiment and activity. Further research, including regression analysis that differentiates between positive (buying) and negative (selling) whale activities, provides deeper insights into how these significant market players influence the broader cryptocurrency landscape.

³ <https://www.ledger.com/academy/glossary/whale>

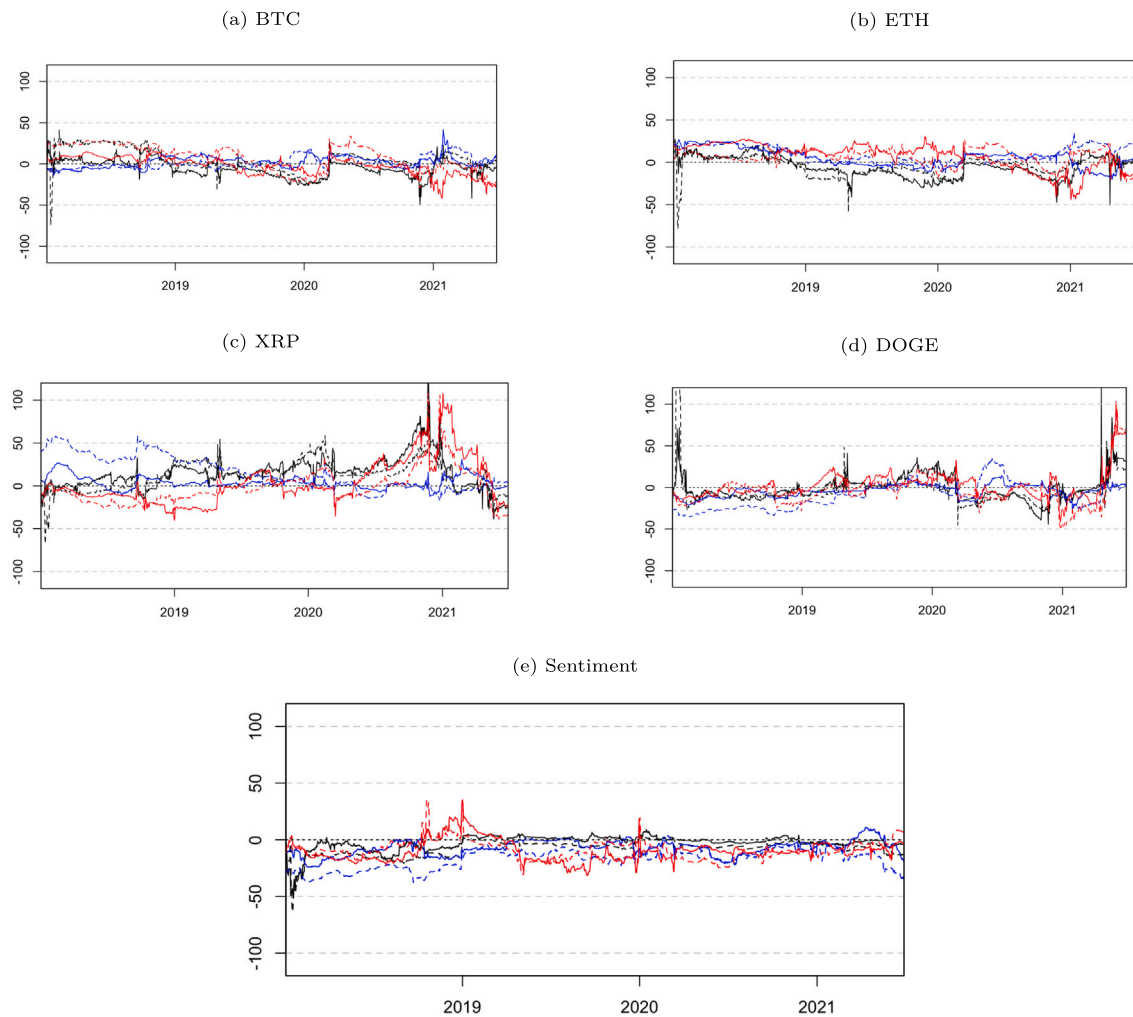


Fig. 3. Net total, long-term and short-term directional connectedness measures.

Note: Results are based on a TVP-VAR model with a lag length of order two (BIC) and a 100 step ahead generalized forecast error variance decomposition. Black, red and blue lines stand for the total, long-term and short-term connectedness measures, respectively. Solid lines represent negative connectedness measures while dashed lines represent positive connectedness measures. Red markers represent whales' activities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

In summary, this study provides a comprehensive examination of sentiment analysis within the cryptocurrency market, focusing on Bitcoin due to its prominent influence. The research utilizes web scraping technology to gather a rich dataset from January 1, 2018, to June 30, 2021, capturing sentiment from various Reddit forums and real-time Twitter comments. This period, marked by significant events and market dynamics, offers a robust basis for analysis. A key finding is the development of a specialized sentiment analysis lexicon, inspired by Loughran and McDonald's methodology, containing over 10,000 terms tailored to the cryptocurrency market. This lexicon enhances the accuracy and depth of sentiment interpretation, achieving an impressive 88.7% accuracy rate compared to existing models like Vader and TextBlob. The model's effectiveness is underscored by its superior performance, establishing a new standard in cryptocurrency sentiment analysis.

This research also explores the impact of sentiment on market volatility, particularly highlighting the bidirectional influence between sentiment and market fluctuations. Granger causality tests reveal that sentiment significantly affects the volatility of key cryptocurrencies such as Bitcoin, Ethereum, Ripple, and Dogecoin. Reverse tests suggest a feedback loop where market changes influence sentiment, which in turn impacts future market trends. The study illustrates that market volatility is sensitive to external shocks, with the Total Connectedness Index (TCI) showing moderate interconnectedness within the market. This interconnectedness highlights the distinctions between cryptocurrencies that primarily receive sentiment-driven shocks and those that transmit them.

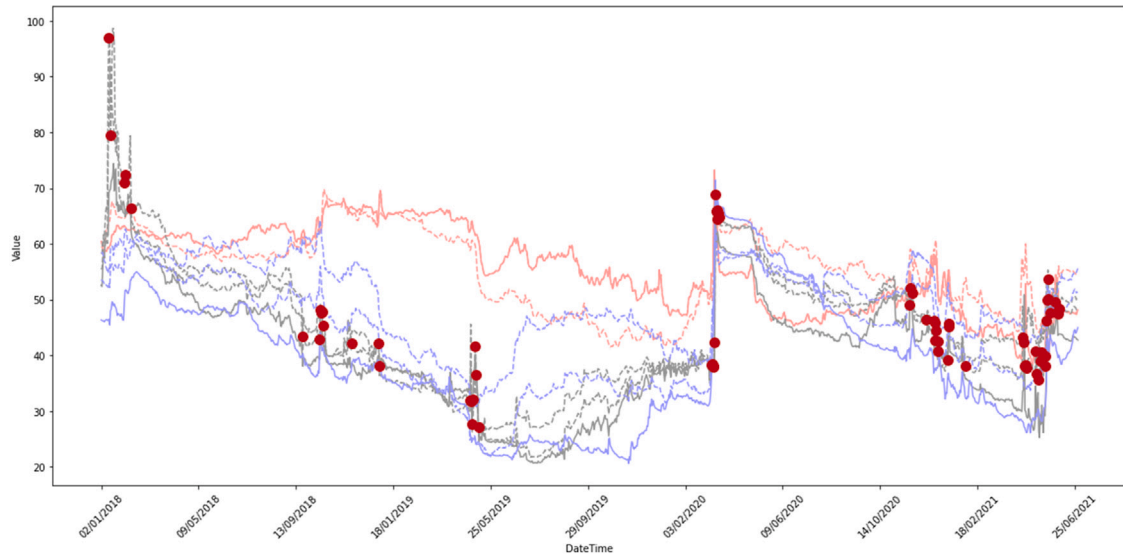


Fig. 4. Connectedness spikes and their associated whales activities.

Note: Results are based on a TVP-VAR model with a lag length of order two (BIC) and a 100-step ahead generalized forecast error variance decomposition. Black, red, and blue lines stand for the total, long-term, and short-term connectedness measures, respectively. Solid lines represent negative connectedness measures while dashed lines represent positive connectedness measures. Red markers represent whales' activities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Overall, the research blends cutting-edge methods, technology, and sector-specific knowledge, setting a new precedent for sentiment analysis within the cryptocurrency market. It provides valuable insights for investors, analysts, and policymakers by illustrating the significant role of investor sentiment in responding to market changes and its potential to foresee or mitigate the effects of market fluctuations. This foundation supports further scholarly inquiry and deepens the understanding of cryptocurrency market dynamics, closing the gap between traditional markets and the unique world of cryptocurrencies.

Appendix

Comprehensive Steps for Sentiment Analysis Dataset Formulation

- (1): 100K A large collection of social media texts undergoes noise labeling techniques. The sentiment of each message is inferred from the emoticons that appear. For example, a message with the 😊 emoji is labeled as “positive”, while a message with the 😞 emoji is categorized as “negative”. In cases where a message contains both positive and negative emoji, the mood is determined by the dominant emoji.
- (2): Manual annotation of a separate set of 10K social media messages. A dedicated team of eight people is specially trained to perform this task, carefully annotating each message.
- (3): At the same time, a state-of-the-art language model, ChatGPT, is deployed to label the same 10K dataset as in (2).
- (4): From the manually annotated 10K set, a subset of 3K messages is kept exclusively for testing purposes. The remaining 7K dataset, together with the 100K nosily labeled dataset in (1), form part of the main training set.
- (5): In order to enrich the training data and further improve prediction performance, the same 3K set retained in (4) is re-annotated by ChatGPT and integrated into the broader training data set.

Data availability

The authors do not have permission to share data.

Table 9
Sample cryptocurrency sentiment lexicon.

Positive	Neutral	Negative
Ability:	Abroad:	Afraid:
able	abroad	Afraid
abilities	abroads	Angry:
ability	Academy:	Angry
Absolute:	academy	Angrier
absolute	academies	Angriest
absolutely	academic	Angrily
absoluteness	Accord:	Annoying:
Accelerate:	accord	Annoy
accelerate	accords	Annoys
accelerates	accorded	Annoyed
accelerated	according	Annoying
accelerating	accordance	Annoyingly
acceleration	accordances	Annoyer
Accept:	Account:	Anxiety:
accept	account	Anxious
accepts	accounts	Anxieties
accepted	accounted	Anxiously
accepting	accounting	Anxiousness
acceptance	accountable	Argue:
Accept Crypto:	accountability	Argue
accepts crypto	accountant	Argues
accepting crypto	accountancy	Argued
accepted crypto	accountants	Arguing
accept crypto	Act:	Argument
Access:	act	Arguments
access	acts	Arrest:
accesses	acted	Arrest
accessed	acting	Arrests
accessing	actor	Arrested
accessibility	actors	Arresting
accessible	action	Arrestable
Accomplish:	action	Arrestor
accomplish	actions	Attack:
accomplishes	Activity:	Attack
accomplished	activity	Attacks
accomplishing	activities	Attacked
accomplishment	Actually:	Attacking
accomplishments	actually	Attacker

References

- Akyildirim, E., Aysan, A. F., Cepni, O., & Darendeli, S. P. C. (2021). Do investor sentiments drive cryptocurrency prices? *Economics Letters*, 206, Article 109980.
- Akyildirim, E., Corbet, S., Lucey, B., Sensoy, A., & Yarovaya, L. (2020). The relationship between implied volatility and cryptocurrency returns. *Finance Research Letters*, 33, Article 101212.
- Allon, G., & Zhang, D. (2017). Managing service systems in the presence of social networks. Available At SSRN 2673137.
- Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129–151.
- Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *The Review of Financial Studies*, 22(1), 151–186.
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307–343.
- Baruník, J., & Křehlík, T. (2018). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics*, 16(2), 271–296.
- Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money*, 54, 177–189.
- Beckers, S. (2019). Do social media trump news? The relative importance of social media and news based sentiment for market timing. *Journal of Portfolio Management*, 45(2), 58–67.
- Biais, B., Bisière, C., Bouvard, M., Casamatta, C., & Menkveld, A. J. (2023). Equilibrium bitcoin pricing. *The Journal of Finance*, 78(2), 967–1014.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
- Bowden, J., & Gemayel, R. (2022). Sentiment and trading decisions in an ambiguous environment: a study on cryptocurrency traders. *Journal of International Financial Markets, Institutions and Money*, 80, Article 101622.
- Bushee, B. J., & Miller, G. S. (2012). Investor relations, firm visibility, and investor following. *The Accounting Review*, 87(3), 867–897.
- Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2013). *Customers as advisors: The role of social media in financial markets: Technical report*, Working paper.
- Chen, A. Y., & Zimmermann, T. (2019). Publication bias and the cross-section of stock returns. *The Review of Asset Pricing Studies*, 10(2), 249–289.
- Cong, L. W., Li, Y., & Wang, N. (2020). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies*, 34(3), 1105–1155.
- Corbet, S., Meegan, A., Larkin, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165, 28–34.
- Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under - and overreactions. *The Journal of Finance*, 53(6), 1839–1885, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00077>.

- Daniel, K. D., Hirshleifer, D., & Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance*, 56(3), 921–965, [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00350](https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00350).
- Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4), 1605–1643, [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2006.00884.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2006.00884.x).
- De Bondt, W. P. M. (1993). Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting*, 9(3), 355–371.
- De Bondt, W. F. M., & Thaler, R. H. (1987). Further evidence on investor overreaction and stock market seasonality. *The Journal of Finance*, 42(3).
- Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1), 57–66.
- Diebold, F. X., & Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1), 119–134.
- Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: The case of REITs. *The Journal of Real Estate Finance and Economics*, 45, 402–434.
- Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics*, 131(3), 693–714.
- El-Gazzar, S. M. (1998). Predisclosure information and institutional ownership: A cross-sectional examination of market revaluations during earnings announcement periods. *The Accounting Review*, 73(1), 119–129, Publisher: American Accounting Association.
- Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements. In *AFA 2009 san francisco meetings paper*.
- Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2008). *The incremental information content of tone change in management discussion and analysis*. NYU Working Paper.
- Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3), 1327–1370.
- Field, L. C., & Lowry, M. (2009). Institutional versus individual investment in IPOs: The importance of firm fundamentals. *Journal of Financial and Quantitative Analysis*, 44(3), 489–516, Publisher: Cambridge University Press.
- Geva, T., Oestreich-Singer, G., Efron, N., & Shimshoni, Y. (2017). Using forum and search data for sales prediction of high-involvement projects. *MIS Quarterly*, 41(1), 65–82.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 44–438.
- Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, 91(1), 153–178.
- Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The Accounting Review*, 89(3), 1083–1113.
- 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.
- Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3), 712–729.
- Jiang, H. (2010). Institutional investors, intangible information, and the book-to-market effect. *Journal of Financial Economics*, 96(1), 98–126.
- Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior and Organization*, 176, 63–90.
- Johnson, R. A., & Greening, D. W. (1999). The effects of corporate governance and institutional ownership types on corporate social performance. *Academy of Management Journal*, 42(5), 564–576, Publisher: Academy of Management.
- Kaniel, R., Liu, S., Saar, G., & Titman, S. (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance*, 67(2), 639–680, [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2012.01727.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2012.01727.x).
- Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), 119–147.
- 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, Article 101188.
- Kyle, A. S., & Wang, F. A. (1997). Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *The Journal of Finance*, 52(5), 2073–2090, [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1997.tb02751.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1997.tb02751.x).
- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196). PMLR.
- Leuz, C., Lins, K. V., & Warnock, F. E. (2009). Do foreigners invest less in poorly governed firms? *The Review of Financial Studies*, 22(8), 3245–3285.
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., Kumar, A., et al. (2022). Holistic evaluation of language models. [arXiv preprint arXiv:2211.09110](https://arxiv.org/abs/2211.09110).
- Liu, B., & McConnell, J. J. (2013). The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? *Journal of Financial Economics*, 110(1), 1–17.
- Liu, Y., & Tsyvinski, A. (2020). Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6), 2689–2727.
- Long, S., Chatziantoniou, I., Gabauer, D., & Lucey, B. (2024). Do social media sentiments drive cryptocurrency intraday price volatility? New evidence from asymmetric TVP-var frequency connectedness measures. *The European Journal of Finance*, 1–20.
- Long, S., Lucey, B., Xie, Y., & Yaroyaya, L. (2023). “I just like the stock”: The role of reddit sentiment in the GameStop share rally. *Financial Review*, 58(1), 19–37.
- 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.
- Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2), 307–326.
- Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. *CoRR*, [arXiv:1705.07874](https://arxiv.org/abs/1705.07874).
- Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research*, 24(1), 146–163.
- Milas, C., Panagiotidis, T., & Dergiades, T. (2021). Does it matter where you search? Twitter versus traditional news media. *Journal of Money, Credit and Banking*, 53(7), 1757–1795.
- Oh, W. Y., Chang, Y. K., & Martynov, A. (2011). The effect of ownership structure on corporate social responsibility: Empirical evidence from Korea. *Journal of Business Ethics*, 104(2), 283–297.
- Pagnotta, E. S. (2021). Decentralizing money: Bitcoin prices and blockchain security. *The Review of Financial Studies*, 35(2), 866–907.
- Pagnotta, E., & Buraschi, A. (2018). An equilibrium valuation of bitcoin and decentralized network assets. Available At SSRN 3142022.
- Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.
- Parrino, R., Sias, R. W., & Starks, L. T. (2003). Voting with their feet: institutional ownership changes around forced CEO turnover. *Journal of Financial Economics*, 68(1), 3–46.
- Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1), 17–29.
- Platanakis, E., & Urquhart, A. (2020). Should investors include bitcoin in their portfolios? a portfolio theory approach. *The British Accounting Review*, 52(4), Article 100837.
- 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, Article 101462.
- Saleh, F. (2019). *Volatility and welfare in a crypto economy: vol. 3235467*, SSRN.
- Smales, L. A. (2022). Investor attention in cryptocurrency markets. *International Review of Financial Analysis*, 79, Article 101972.
- Stiassny, A. (1996). A spectral decomposition for structural VAR models. *Empirical Economics*, 21(4), 535–555.

- Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1422–1432).
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437–1467.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131, Publisher: American Association for the Advancement of Science.
- Wiebe, J., Bruce, R., & O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the association for computational linguistics* (pp. 246–253).
- Wiesen, T. F., Beaumont, P. M., Norrbin, S. C., & Srivastava, A. (2018). Are generalized spillover indices overstating connectedness? *Economics Letters*, 173, 131–134.
- Wu, L., & Brynjolfsson, E. (2009). The future of prediction: how google searches foreshadow housing prices and quantities. *ICIS 2009 Proceedings*, 147.
- Yang, K.-C., & Menczer, F. (2023). Large language models can rate news outlet credibility. arXiv preprint arXiv:2304.00228.
- Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480–1489).
- Yermack, D. (2015). Chapter 2 - is bitcoin a real currency? An economic appraisal. In D. Lee Kuo Chuen (Ed.), *Handbook of digital currency* (pp. 31–43). San Diego: Academic Press.
- Zeng, A., Liu, X., Du, Z., Wang, Z., Lai, H., Ding, M., Yang, Z., Xu, Y., Zheng, W., Xia, X., et al. (2022). Glm-130b: An open bilingual pre-trained model. arXiv preprint arXiv:2210.02414.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

Dr. Suwan(Cheng) Long is an Assistant Professor of Finance at the IESEG School of Management. She earned her Ph.D. in Finance from Trinity College Dublin in 2024 and served as a visiting researcher at the University of Cambridge from 2022 to 2024. Her research interests encompass behavioral finance, fintech (NLP), ESG, corporate finance, cryptocurrencies, and their environmental impacts. Dr. Long has published extensively, with over 10 publications in leading international journals such as the Risk Analysis, European Journal of Finance, International Review of Financial Analysis, Financial Review, etc.

Ying Xie specializes in data analysis, machine learning models, Natural Language Processing, econometric and financial modeling, net zero decarbonization, and circular economy. Her work supports businesses enhancing data driven decision making and improving operational efficiency. Ying has published widely in leading journals, such as International Journal of Production Economics, European Journal of Operational Research, International Review of Financial Analysis, Finance Review, and Journal of Cleaner Production. Ying is the associate editor of Supply Chain Analytics, and special issue editor for European Management Journal, and Maritime Economics and Logistics. Ying has successfully won and completed over 20 projects, totaling more than £20m, including projects from EPSRC, Innovate UK, EU Horizon, Health Education England, ERDF and industries.

Professor Brian Lucey is Professor of International Finance and Commodities in Trinity College Dublin. He is a Research Associate of the Paris School of Business Center of Research for Energy and Climate Change (CRECC), Visiting Professor at Shanghai Lixin University of Accounting and Finance and at Abu Dhabi University. He is Editor in Chief of International Review of Economics and Finance, and of the Journal of Economic Surveys and Senior Editor of Journal of Climate Finance. He has published over 240 papers including many in the area of sustainable finance, climate finance, applications of NLP and text-as-data, and on corporate finance

Professor Andrew Urquhart is Professor of Finance & Financial Technology at the University of Birmingham, where is also Head of the Department of Finance. . His main research interests are Fintech, Blockchain, Corporate Finance and Financial Markets. Andrew has published widely with over 80 publications in leading international journals such as the Nature, Journal of Corporate Finance, European Journal of Operational Research, Journal of Financial Markets, while garnering over 9000 cites. He has also received over £500,000 in research income and he is also an associate editor at British Accounting Review, Economics Letters, European Journal of Finance, Financial Review, Global Finance Journal, International Review of Financial Analysis, and Research in International Business and Finance.



DSPACE

<https://dspace.org/>

From whales to waves: social media sentiment, volatility, and whales in cryptocurrency markets

Long, Suwan (Cheng); Xie, Ying; Zhou, Zhengyuan; Lucey, Brian; Urquhart, Andrew

2026-12-31

Elsevier

<https://doi.org/10.1016/j.bar.2025.101682>

Long SC, Xie Y, Zhou Z, et al., (2025) From whales to waves: social media sentiment, volatility, and whales in cryptocurrency markets. The British Accounting Review, Available online 16 May 2025, Article number 101682