BULLETIN
OF
ECONOMIC THEORY AND ANALYSIS

Journal homepage: https://dergipark.org.tr/tr/pub/beta

Relationship Between Twitter Sentiment Analysis and Bitcoin Prices: Econometric Analysis of Long and Short Term Dynamics

Çağrı ULU https://orcid.org/0000-0001-5338-2987

Cansu ULU https://orcid.org/0000-0002-3676-218X

To cite this article: Ulu, Ç. & Ulu, C. (2024). Relationship Between Twitter Sentiment Analysis and Bitcoin Prices: Econometric Analysis of Long and Short Term Dynamics. *Bulletin of Economic Theory and Analysis*, *9*(2), 605-627.

Received: 05 Feb 2024

Accepted: 10 Jun 2024

Published online: 30 Jun 2024

Bulletin of EconomicTheoryandAnalysis

Volume 9, Issue 2, pp. 605-626, 2024

https://dergipark.org.tr/tr/pub/beta

OriginalArticle / Araştırma Makalesi

Received / Alınma: 05.02.2024 Accepted / Kabul: 10.06.2024

Relationship Between Twitter Sentiment Analysis and Bitcoin Prices: Econometric Analysis of Long and Short Term Dynamics

Çağrı ULU[a]

Cansu ULU[b]

[a] Öğr. Gör. Dr., İzmir Kavram Meslek Yüksekokulu, Dış Ticaret Bölümü, İzmir, TÜRKİYE
https://orcid.org/0000-0001-5338-2987
[b] Öğr. Gör. Dr., Rüştü Ünsal Polis Meslek Yüksekokulu, İzmir, TÜRKİYE
https://orcid.org/0000-0002-3676-218X

ABSTRACT

The significance of social media in influencing cryptocurrency pricing has grown considerably in recent years. This study aims to explore the correlations between social media sentiments and Bitcoin pricing, both in the short and long terms, while also investigating the direction of these relationships. Sentiment analysis was conducted using the TextBlob model, which uncovers the underlying meaning in text through analysis. The study tested the hypothesis that there exists a relationship between sentiment analysis scores and Bitcoin prices over both short and long periods. Ensuring stationarity was crucial for time series analysis, involving the use of structural break and traditional unit root tests. Daily data from June 2021 to June 2022 was examined, with December 2021 serving as the focal point due to a peak in Bitcoin prices. The study focused on Bitcoin price data and sentiment analysis scores. Results revealed Twitter data as the dependent variable, showing no long-term relationship with Bitcoin prices. However, a significant and positive relationship was observed in the short term. This research contributes valuable insights into the intricate dynamics between social media sentiments and cryptocurrency pricing.

Keywords
Bitcoin,
Twitter,
Sentiment Analysis,
ARDL Bounds Test,
Text mining

JELClassification
G0, C1, A1, E60

İLETİŞİM Çağrı ULU cagri.ulu@kavram.edu.trl İzmir Kavram Meslek Yüksekokulu, Dış Ticaret Bölümü, İzmir, TÜRKİYE.

Twitter Duyarlılık Analizi ile Bitcoin Fiyatları Arasındaki İlişki: Uzun ve Kısa Dönem Dinamiklerin Ekonometrik Analizi

ÖZ

Sosyal medyanın kripto para birimi fiyatlandırmasını etkilemedeki önemi son yıllarda önemli ölçüde artmıştır. Bu çalışma, sosyal medya duyarlılığı ile Bitcoin fiyatlandırması arasındaki ilişkileri hem kısa hem de uzun vadede araştırmaktadır. Aynı zamanda bu ilişkilerin yönünü tespit etmeyi amaçlamaktadır. Duyarlılık analizi, metnin altında yatan anlamı ortaya çıkaran TextBlob modeli kullanılarak gerçekleştirilmiştir. Çalışma, duyarlılık analizi puanları ile Bitcoin fiyatları arasında hem kısa hem de uzun vadede bir ilişki olduğu hipotezini test etmiştir. Durağanlığın sağlanması, yapısal kırılma ve geleneksel birim kök testlerinin kullanımını içeren zaman serisi analizine imkan sağlamıştır. Çalışma kapsamındaHaziran 2021'den Haziran 2022'ye kadar olan günlük veriler incelenmiştir ve Bitcoin fiyatlarındaki zirve nedeniyle Aralık 2021 odak noktası kabul edilmiştir. Çalışma Bitcoin fiyat verilerine ve duyarlılık analizi puanlarına odaklanmaktadır. Sonuçlar, Bitcoin fiyatlarıyla uzun vadeli bir ilişki göstermeyen Twitter verilerinin bağımlı değişken olduğunu ortaya çıkarmıştır. Ancak kısa vadede anlamlı ve pozitif bir ilişki gözlenmiştir. Bu araştırma, sosyal medya duyarlılığı ile kripto para birimi fiyatlandırması arasındaki karmaşık dinamiklere dair literatüre değerli bilgiler sağlamaktadır.

AnahtarKelimeler
Bitcoin,
Twitter,
Sentiment Analysis,
ARDL Bounds Test,
Text mining

JEL Kodu
G0, C1, A1, E60

1. Introduction

Understanding the dynamics between social media sentiment and financial markets has become increasingly crucial in the age of digital information. This study navigates this intersection by delving into the intricate relationship between Twitter sentiment analysis (TSA) and Bitcoin prices (BTC). By employing advanced data mining techniques, including sentiment analysis and econometric modeling, we aim to uncover valuable insights into both short-term and long-term relationships. This paper not only provides a rigorous analysis but also contributes to the evolving discourse on the impact of social media sentiments on cryptocurrency markets.

Neoclasical economics suggests that, individuals, both investors and consumers, are presumed to strive for profit maximization based on the theory of expected utility. In this context, challenges arise when attempting to incorporate social factors into the equation. However, this theoretical stance proves insufficient in accurately describing actions in practice (Polat, 2021). Therefore, Kahneman and Tversky have questioned whether individuals under risk adhere to the expected utility theory, highlighting the inadequacies of this perspective in explaining real-world behavior. Their inquiry into the rationality of decision-making under uncertainty has spurred a shift towards behavioral economics, acknowledging the limitations of neoclassical assumptions and underscoring the importance of incorporating psychological and social considerations in economic analyses.

According to Kahneman and Tversky's Prospect Theory, uncertainties and risks encountered in decision-making mechanisms are influential factors. Therefore, in the analysis of economic behaviors, concepts such as social norms, information, and misinformation should be taken into consideration (Kahneman & Tversky, 1979). The theoretical framework proposed by Kahneman and Tversky highlights that individuals do not always make decisions based solely on rational calculations of expected utility.

In contemporary times, behavioral methods are gaining prominence over classical economic paradigms. Analyzing individuals' preferences under the assumption of "ceteris paribus" does not necessarily yield optimal results. In the financial market, when comparing two or more data sets, relying solely on time series may prove insufficient. Factors influencing behavior must also be considered and examined. This study focuses on econometric analyses of Bitcoin pricing and Twitter Sentiment Analysis, with particular emphasis on evaluating the impact of social media manipulation on Bitcoin pricing, which holds significant importance for investors. The integration of behavioral factors and sentiment analysis contributes to a more comprehensive understanding of the dynamics influencing financial markets, transcending traditional economic models.

Sentiment analysis, a subset of text mining, involves the extraction of valuable insights from textual data through data mining techniques. Data mining, as described by Baykal (2006), utilizes algorithms to extract meaningful information from both structured and unstructured data. In the realm of text mining, natural language processing algorithms are employed to analyze text or data transformable into text, treating it as unstructured data to unveil meaningful patterns.

Various sub-analytic methods within text mining, including text classification, clustering, concept and topic detection, summarization, sentiment analysis, and emotion analysis, leverage techniques like frequency analysis, syllable analysis, tagging, pattern extraction, and data visualization to extract meaningful information. This study focuses on sentiment analysis, employing it as a valuable tool to delve into the relationship between Twitter TSA and BTC. To support our findings, a word cloud—a form of data visualization—was utilized. Words in the cloud were visualized based on their frequency of use, providing a clear representation of the sentiment landscape. The study's time frame spans from June 1, 2021, to May 31, 2022.

For the econometric analysis, data from both sentiment analysis and Bitcoin/USD prices sourced from investing.com were inputted into the E-views 10 package program. Stationarity of the series was ensured through Augmented Dickey-Fuller (ADF) and Phillips Perron (PP) tests, with additional validation using the Fourier Augmented Dickey-Fuller (FADF) test to check for structural breaks. The results indicated that TSA was stationary at I(0), while BTC was stationary at I(1), leading to the use of the Autoregressive Distributed Lag Bound Test (ARDL) for further analysis.

To address the valuable feedback on the theoretical background, we recognize the need for a more detailed and comprehensive exploration of the theoretical underpinnings. The theoretical foundation of the study is now expounded upon, providing a thorough explanation of sentiment analysis and its relevance to cryptocurrency pricing. The intricate relationship between social media sentiments and the cryptocurrency market is elucidated, offering a robust background that serves as a solid foundation for the subsequent analysis.

Within this theoretical framework, our hypothesis gains prominence. We posit that fluctuations in Twitter sentiment, as analyzed through sophisticated sentiment analysis techniques, have a discernible impact on Bitcoin prices. This hypothesis forms the crux of our investigation, guiding the subsequent data mining and econometric analyses. By examining Twitter sentiment over the selected time period and correlating it with Bitcoin price movements, our study seeks to not only uncover patterns but also contribute valuable insights into the short-term and long-term dynamics shaping cryptocurrency markets. Through this focused hypothesis, our research aims to advance the understanding of the intricate relationship between social media sentiments and cryptocurrency pricing, adding depth to the evolving discourse in this domain.

In the subsequent sections, we provide a comprehensive overview of the relevant literature, elucidate data mining and sentiment analysis as fundamental research methods, and detail the creation and discussion of the econometric model. The study concludes with the results and discussion section, offering a thorough exploration of the insights gained through our analysis.

2. Literature Review

In recent years, the burgeoning field of data mining has gained significant prominence within social sciences and finance, propelled by the analytical capabilities that computer technologies afford to researchers. This analytical method, rooted in big data, has facilitated the establishment of substantial relationships between variables. Within the domain of finance and economics, a multitude of studies has explored the interplay between Twitter sentiment analysis and BTC prices. Extensive literature reveals diverse investigations into causality relationships, examining not only the association between Twitter sentiment analysis and BTC prices but also exploring the nuanced relationships with various other variables.

Interdisciplinary approaches within social sciences have delved into language processing models' accuracy, investigating the nature of relationships between Twitter sentiment analysis and other variables. Finance-centric studies have endeavored to establish connections between sentiment analysis and market movements. These investigations span a spectrum of data sources, including web pages, financial news, internet comments, and Twitter. In alignment with this research landscape, our study focuses on scrutinizing the relationship between sentiment values derived from Twitter and BTC prices.

The literature pertinent to our research is summurized in Table 1, providing a condensed overview of key studies in the field.

Table 1

Literature Review

Author and Year	Variables	Dates	Method	Findings
Soomro, Rajput, Soomro (2022)	- Google Trends Sentiment Analysis Index - BTC price - USD	- 2013-2018	-ARDL	- The Sensitivity Analysis Index can predict BTC returns, trading volume, and volatility associated with BTC returns. -The ARDL results indicate that the USD has a negative impact on BTC prices in both the short and long term. -The asymmetrical effects of BTC sensitivity, as measured by the Sensitivity Analysis Index, are evident in both BTC prices and fiat currencies.

Katsafados, Nikoloutsopoulos, Leledakis (2023)	- BEL-20, - DAX-30, - CAC40, - FTSE-MIB, - IBEX-35, - FTSE - S&P-500 -VADER Sentiment	- January 2021 - June 2021	- Panel VAR - Panel ARDL	- An increase in positive emotions can be associated with a short-term increase in stock prices. On the other hand, an increase in negative emotions in English-speaking countries is inversely related to stock prices and has a long-term impact. It has been observed that positive emotions are associated with high returns and low volatility in the short term. Negative emotions, on the other hand, have been found to be associated with low returns in the short term.
Chen (2022)	-BTC price - BTC mining data - Google Trend	-07.02.2016 - 07.02.2021	- ARDL	- The price of BTC exhibits a more dramatic response to changes in attitudes towards mining activities. - This relationship has become even stronger after 2019, and the correlation has become closer.
Gözbaşı, Şahin, Altınöz (2021)	-BTC price -S&P 500	- August 2010 – April 2021 montly data	-ARDL -Granger Causality	- BTC prices have a limited response to volatility in the commodity markets (only responding to oil prices in the short term), that gold prices do not have a statistically significant effect on BTC prices, and that an increase in crude oil prices has a negative effect on BTC prices in the short term but no effect in the long term, as revealed in the study.
Chalkiadakis, Peters, Ames (2021)	-BTC price -ETH price	- January 2018- January 2021	- ARDL - MIDAS -VADER -LSTM -BERT	- Advantages of the modeled infinite lag ARDL-MIDAS-Transformer time series regression model class in the study were observed in terms of incorporating sophisticated long-memory signal features and their interpretability capabilities.
Kjærland, Meland, Oust, Øyen (2018)	- BTC price - S&P 500 - Google Trends	- 18.09.2011 - 05.02.2017 weekly data	- ARDL - OLS	- Relationship has been identified between the interest in BTC and its price fluctuations. There is a significant and positive correlation between Google searches and BTC prices.
Gaies, Nakhli, Sahut, Guesmi (2021)	- BMI Sentiment Analysis Index - BTC price	- August 2011- July 2020 montly data	- ARDL - NARDL	- Investors rely on sentiment indexes to predict BTC prices. - The sentiment index helps portfolio managers to determine short and long-term investment strategies.

Sattarov, Jeon, Oh, Lee (2020)	-VADER TSA Index -BTC price	-12 March 2018 – 12 May 2018 daily data	- Random Forest Model -VADER	- A partial relationship was found between the fluctuations in BTC price and the fluctuations in sentiment classes. A strong relationship was observed between the percentage changes in BTC and Twitter sensitivity.
Shahvari (2022)	- VIX Sentiment Analysis Index - BTC price - Gold price	- August 2020-August 2022 daily data	- ARDL	- It has been found that gold prices have a positive impact on investor sentiment index. Similarities have been identified between gold and BTC. Therefore, it has been suggested that the sentiment index could be an important indicator for investors.
Li, Wang (2017)	- USD Indexes - BTC Indexes - Google Trends - TSA Index -Mining data	- 01.01.2011 - 12.12.2014	- ARDL	- The BTC price is less sensitive to technological variables and is more influenced by economic factors.
Galeshchuk, Vasylchyshyn, Krysovatyy (2018)	- BTC price - TSA Index	- January 201 – September 2017 daily data	- Random Walk -ARIMA - Convolutional Neural Network	- It has been observed that models using sentiment analysis data in addition to Bitcoin price fluctuations perform better in BTC predictions compared to models that only rely on BTC price fluctuations.
Wong (2021)	- BTC price - TSA Index	- January 2016-November 2019	-LSTM -Naïve Bayes model	- The relationship between BTC prices and TDA was better explained by LSTM's higher explanatory capability.

The collective findings from the literature review indicate a compelling relationship between Twitter sentiment analysis and Bitcoin prices. Numerous studies underscore the predictive power of sentiment analysis metrics, such as the Sensitivity Analysis Index and VADER Sentiment, forecasting BTC returns, trading volume, and volatility. Economic factors, including the USD and gold prices, are identified as influential, showcasing the intricate dynamics of cryptocurrency markets. Notably, the studies reveal distinct short-term and long-term effects of sentiment on both BTC prices and fiat currencies. Interdisciplinary approaches employ advanced models like ARDL, MIDAS, LSTM, and BERT, enhancing interpretability and capturing intricate relationships. Practical implications for investors emerge, with sentiment indexes guiding predictions and aiding in formulating investment strategies. Correlations with external variables and advancements in model performance, particularly when combining sentiment analysis with traditional BTC price data, further enrich our understanding of this complex relationship. This comprehensive literature review sets the stage for our study, providing insights into the diverse methodologies and findings of previous research, and guiding the subsequent exploration of the relationship between Twitter sentiment and BTC prices in our unique context.

3. Data Mining and Twitter Sentiment Analysis

Sentiment analysis encompasses three distinct model types: lexicon-based models, machine learning models, and hybrid models. This study opts for the TextBlob model, a lexicon-based approach for sentiment analysis. Operating within the Python programming language, TextBlob leverages the Natural Language Toolkit (NLTK) for natural language processing tasks, including sentiment analysis (NLTK, 2023). The sentiment analysis yields a numerical value ranging from -1 to +1, where proximity to -1 denotes a negative sentiment, and proximity to +1 indicates a positive sentiment.

Data collection constitutes the initial phase of sentiment analysis. Retrieving the most popular 50 English tweets with the Bitcoin hashtag between June 1, 2021, and May 31, 2022, was achieved through the APIFY platform. Given constraints with the Twitter developer account, the 'Twitter URL Scraper' module within APIFY was employed to extract tweets from the internet. Daily, 50 tweets were consolidated into a single text for analysis.

The subsequent step involves data structuring. Following the acquisition of data for Twitter Sentiment Analysis (TSA), a series of preprocessing steps were applied:

- Conversion of uppercase letters to lowercase
- Removal of non-English words
- Elimination of unnecessary words listed in the stopword collection
- Exclusion of numeric values
- Removal of special characters, emoticons, emojis, and single-word entities
- Handling 'url' extensions indicative of internet connections

- Omission of user names
- Extraction of sentiment-associated words.

Sentiment analysis results were derived through custom code implementation and requisite Python libraries. Complementing the result assessment, a word cloud visualizing word frequency in the cleaned texts was generated using the wordcloud and matplotlib libraries.

Figure 1. TSA Word Clouds

The TSA Word Cloud for the period from June 1, 2021, to May 31, 2022 (Figure 1) illustrates the aggregate of Twitter content as a singular text. Notably, frequent occurrences of terms such as "buy" and "new" align with themes of BTC innovation and purchasing activities. On November 10, 2021, a date coinciding with peak BTC prices, the word cloud prominently features "one," "new," and "high." The term "new" may signify the influx of new investors or the introduction of novel market offerings, while "high" emphasizes the elevated BTC price levels. These insights serve as crucial indicators for investors reflecting on market dynamics.

4. Analysis Method

As part of the econometric analysis in this study, we first conducted stationary tests on the series. We tested stationarity using unit root tests. Specifically, we employed the widely used

Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root tests, which do not allow for structural breaks, on the two variables. Subsequently, to strengthen the evidence of stationarity, we applied the Fourier Augmented Dickey-Fuller (FADF) unit root test, which allows for structural breaks. Based on the stationary results, we established an ARDL model and estimated the results. We selected the time interval from June 1st, 2021 to May 31st, 2022, which corresponds to the date on which BTC attained its highest value against the dollar, as the center point in this study.

4.1. Dataset and Econometric Method

Time series analysis has been the focus of numerous studies, and the direction of research on series parameters varies. In this study, we utilized BTC price (USD) and sentiment analysis values of tweets posted using the Bitcoin hashtag as variables. The table below presents the variables and their sources used in this study.

Table 2

Variables and Sources

Variables	Explanations	Sources
BTC	Bitcoin/USD	Investing
TSA	Sensitivity Analysis Results of 50 Most Liked Daily Tweets with Bitcoin Hashtag	Twitter, APIFY

The descriptive statistics of the variables analyzed are presented in the table below.

Table 3

Descriptive Statistics

	lnBTC	TSA
Mean	10.69500	0.113246
Median	10.67993	0.112522
Maximum	11.14173	0.275645
Minimum	10.28173	-0.033787
Std. Dev.	0.198507	0.048066
Skewness	0.161267	-0.072996
Kurtosis	2.548311	3.240433
Jarque-Bera	4.672112	1.200009
Probability	0.096708	0.548809
Observations	364	364

Upon examining the descriptive statistics table, the maximum and minimum values were not interpreted due to the logarithmic nature of the series. However, an analysis of the standard deviations reveals that the volatility in Bitcoin is higher. Furthermore, the Jarque-Bera probability values indicate that both series follow a normal distribution at the 5% significance level.

The correlation matrix among variables is presented in the table below, following the descriptive statistics.

Table 4

Correlation Matrix between TSA and BTC

	TSA	BTC
TSA	1.000000	

BTC	0.095608	1.000000
	0.0685	-----

As shown in the table, there is a significant positive correlation at the 10% significance level between BTC and TSA. The low coefficient can be interpreted as indicating a weak relationship between the two variables.

The trend charts of the selected variables during the chosen period are presented in the figures below, prior to conducting the unit root tests.

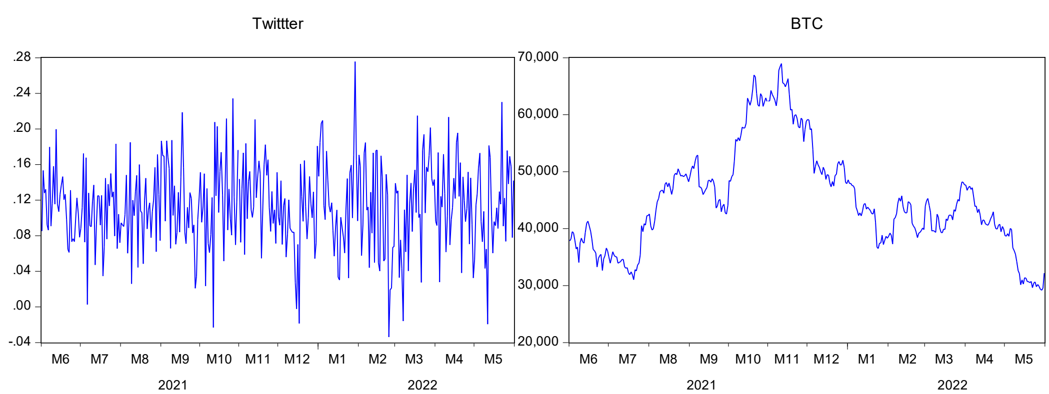

Figure 2. TSA and BTC Charts

The figures reveal that the TSA variable has exhibited a positive trend over time. Additionally, it can be observed that BTC reached a peak of 68,000 USD towards the end of 2021.

4.2. Unit Root Test Results with and without Structural Breakage

Stability is of utmost importance in time series analysis, and the direction of the study is shaped by the results of stationarity tests. In this study, first, the unit root tests of ADF developed by Dickey and Fuller (1979) and PP developed by Phillips and Perron (1988), which are widely used traditional methods that do not allow for structural breaks, were employed. The test results for the BTC variable are shown below.

Table 5

BTC Unit Root Test Results (ADF,PP)

	ADF			PP		
	Level					
		Intercept	Trend and Intercept		Intercept	Trend and Intercept
	T-Statistics	-1.164164	-1.344020	T-Statistics	-1.279077	-1.415580
	%1	-3.448161	-3.983541	%1	-3.448161	-3.983541
	%5	-2.869285	-3.422252	%5	-2.869285	-3.422252
	%10	-2.570963	-3.133975	%10	-2.570963	-3.133975
	Prob.	0.6908	0.8751	Prob.	0.6403	0.8551
lnBTC	1st diff.					
		Intercept	Trend and Intercept		Intercept	Trend and Intercept
	T-Statistics	-17.46629	-17.52155	T-Statistics	-17.43800	-17.52321
	%1	-3.448211	-3.983612	%1	-3.448211	-3.983612
	%5	-2.869307	-3.422286	%5	-2.869307	-3.422286
	%10	-2.570975	-3.133995	%10	-2.570975	-3.133995
	Prob.	0.0000***	0.0000***	Prob.	0.0000***	0.0000***

*Note. *, **, and *** denote significance at the levels of 10%, 5%, and 1%, respectively.*

When traditional unit root test results are examined for BTC, it is understood that the series is not stationary at the level. However, when the first difference is taken, stationary is achieved at the 1% significance level in both the ADF and PP tests. The results of the same unit root test applied to TSA are shown in the table below.

Table 6

TSA Unit Root Test Results (ADF,PP)

ADF				PP		
Level						
TSA		Intercept	Trend and Intercept		Intercept	Trend and Intercept
	T-Statistics	-17.44182	-18.00761	T-Statistics	-18.02647	-18.00761
	%1	-3.448161	-3.983541	%1	-3.448161	-3.983541
	%5	-2.869285	-3.422252	%5	-2.869285	-3.422252
	%10	-2.570963	-3.133975	%10	-2.570963	-3.133975
	Prob.	0.0000***	0.0000***	Prob.	0.0000***	0.0000***

*Note. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.*

When examining the unit root test results applied to TSA, it is observed that the series is stationary in level values. However, in order for the unit root tests to be reliable, the FADF test developed by Enders and Lee (2012), which is a modern approach allowing for structural breaks, was decided to be applied..

As is known, for the FADF test to be valid, certain conditions must be met. Firstly, a new model is created by taking the difference values of the series. The following formulation is applied to find the appropriate frequency and lag number on the difference values.

$$\Delta lnBTC_t = \beta_0 + \beta_1 lnBTC_{t-1} + \gamma_1 sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t$$

and;

$$\Delta TSA_t = \beta_0 + \beta_1 TSA_{t-1} + \gamma_1 sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t$$

Here, the "k" values are used to determine the "total sum of squares" and to find the appropriate frequency range. Using this formulation installed in E-views 10 program, suitable frequencies at different ranges were found for the two series. The suitable frequency range for BTC was found to be "1". For TSA, the suitable frequency range was found to be 5. Another essential assumption of the FADF is to calculate the appropriate lag length after the frequency range is found. The following formulation is added to the above formulation to determine the lag length.

$dBTC(-1\ to\ -12)$ and $dTSA(-1\ to\ -12)$

In our study, no lag length could be calculated for any value. As the series did not meet the assumptions of FADF, the ADF unit root test results are considered valid. Accordingly, BTC is stationary at the level of I(1), while TSA is stationary at the level of I(0).

Different levels of stationarity in time series are important for econometric modeling. Therefore, as the series exhibit stationarity at different levels, the applied ARDL model was considered appropriate for the study.

4.3. ARDL Model

As is well known, the ARDL boundary test developed by Pesaran et al. (2001) is a widely used analysis model in time series when there is stationary at different levels. The purpose of constructing the model is to determine the presence or absence of a co-integration relationship between the series. If there is co-integration, the short- and long-term coefficients of the series are estimated. In summary, the ARDL model is a versatile econometric tool for analyzing the dynamic interactions between variables, capturing both short-run fluctuations and long-run relationships, and it is particularly useful for time series data analysis. In this study, the ARDL model has been estimated. The appropriate ARDL model is formulated as follows:

$$\Delta Y_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \sum_{i=0}^{m} \alpha_{2i} \Delta X_{t-i} + \alpha_3 Y_{t-1} + \alpha_4 X_{t-1} + \varepsilon_t$$

When the data used in the study is included in the model, the ARDL(4,1) model is as follows:

$$\begin{aligned}\Delta TSA_t = \alpha_0 &+ \sum_{i=1}^{m} \alpha_i \Delta TSA_{t-1} + \sum_{i=1}^{m} \alpha_{2i} \Delta TSA_{t-2} + \sum_{i=1}^{m} \alpha_{3i} \Delta TSA_{t-3} \\ &+ \sum_{i=1}^{m} \alpha_{4i} \Delta TSA_{t-4} + \sum_{i=0}^{m} \alpha_{5i} \Delta lnBTC_{t-i} + \alpha_6 TSA_{t-1} + \alpha_7 lnBTC_{t-1} + \varepsilon_t\end{aligned}$$

In this equation, ΔTSA_t represents the first difference of the TSA variable at time *t*, $\Delta lnBTC_{t-1}$ represents the first difference of the BTC variable lagged by one period, α_0 through α_{4i} are coefficients associated with the first differences of TSA, α_{5i} is the coefficient associated with the first difference of BTC, α_6 is the coefficient associated with the lagged TSA variable, α_7

is the coefficient associated with the lagged BTC variable, and ε_t represents the error term at time t. According this formula hypotesis is created below:

$H_o : \alpha_6 = \alpha_7 = 0$

$H_1 : \alpha_6 \neq \alpha_7 \neq 0$

According to the results of the ARDL model constructed using the EViews program, it is evident that there is cointegration among the series. Both the long-run and short-run statistics support this conclusion. H_o has been rejected.

In the ARDL model, the following formulation is used to examine the long-run relationship between the variables:

$$TSA_t = \alpha_0 + \sum_{i=1}^{m} \alpha_{1i} TSA_{t-i} + \sum_{i=0}^{m} \alpha_{2i} lnBTC_{t-i} + \varepsilon_t$$

To analyze the short-run and error correction form relationship between the variables, the following formulation is employed. This formulation allows us to capture the immediate interactions and dynamic adjustments among the variables over a shorter time horizon. The short-run relationship is represented by the following equation:

$$\Delta TSA_t = \alpha_0 + \sum_{i=1}^{m} \alpha_{1i} \Delta TSA_{t-1} + \sum_{i=0}^{m} \alpha_{2i} \Delta lnBTC_{t-2} + \alpha_1 EC_{t-1} + \varepsilon_t$$

The model we constructed satisfies all the requirements of the ARDL bounds test. The Akaike information criterion was selected as the information criterion, and Newey and West's (1987) Heteroskedasticity and Autocorrelation Consistent Covariance (HAC) was selected to remove the problems of changing variances and autocorrelation in the covariance matrix. The appropriate lag length for the ARDL model was determined to be 4, with the Akaike information criterion used as the selection criterion. According to the model constructed in the E-views 10 program, the suitable model was estimated to be ARDL(4,1). The F-statistic value and criteria for the ARDL(4,1) model are shown in the table below.

Table 7

ARDL(4.1) F Statistics and Critical Values

Model	k	M	F-Statistics	Critical Values	Min. Limit	Max Limit

ARDL(4,1)	1	4	17.68262*	%10	3.02	3.51
				%5	3.62	4.16
				%1	4.94	5.58

*Note. M represents the maximum lag order, k represents the number of explanatory variables, and * denotes the significance level of 1%. The critical values used for the lower and upper bounds were obtained from Table CI(ii) in the study conducted by Pesaran et al. (2001, p. 300).*

The F-statistic value of 17.68 is greater than the critical values, indicating that the model is stationary and there is cointegration between the independent and dependent variables at a significance level of 1%.

After the F-statistic value is found to be significant, indicating the presence of cointegration between the independent and dependent variables, the short- and long-run coefficients can be interpreted. However, to interpret these coefficients, the assumptions of the ARDL model must be met. Specifically, there should be no autocorrelation problem for short- and long-run relationships, no model specification error, and no heteroskedasticity issue. In the autocorrelation test, the probability value was found to be 0.94, which is greater than 0.01, indicating the absence of an autocorrelation problem. The result of the normality test applied to the model is shown in the table below.

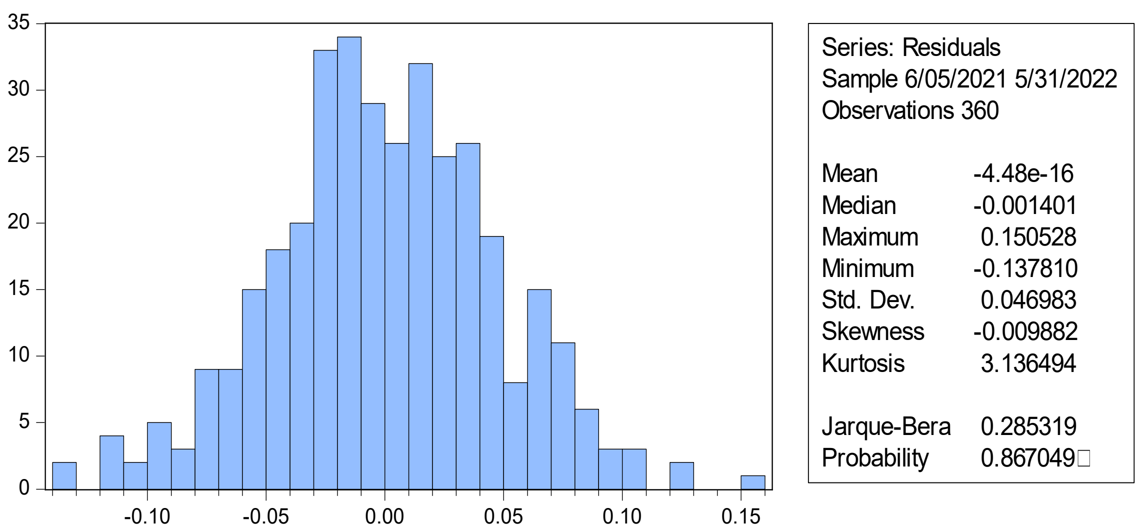

Figure 3. Normality Test of ARDL(4,1) Model

Since the probability value (0.86) in the normality assumption is greater than 0.01, the normality assumption is satisfied.

To test for heteroscedasticity, the test statistic was 2.95, and the probability value was 0.81. Since the probability value is greater than 0.01, it can be concluded that there is no heteroscedasticity issue, and therefore, the model has constant variance.

As seen, three assumptions of ARDL test have been tested and there was no problem in terms of model building. However, to increase the reliability, the CUSUM test developed by Brown et al. (1975) was also applied to determine if there is any structural break in the model. The results are shown in the following figure.

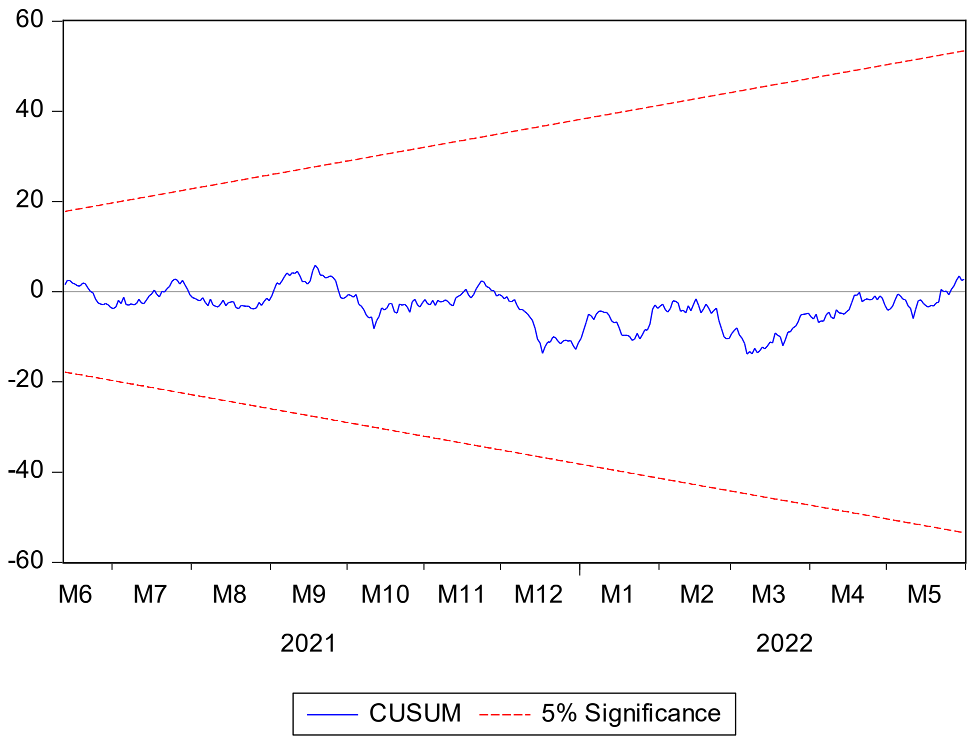

Figure 4. CUSUM Chart

According to the CUSUM test, it can be seen that the coefficients moved within the critical values. Therefore, there is no structural break. Thus, there was no need for a dummy variable.

After fulfilling the ARDL assumptions, the long and short-term coefficients of the model can now be interpreted. The tables below display the long and short-term coefficients of the BTC and TSA series.

Table 8

ARDL Long Term Coefficients

Variables	Coefficient	Standart Error	t-Statistics	Prob.
BTC	0.018747	0.018609	1.007403	0.3144
C	-0.087289	0.199093	-0.438434	0.6613

When examining the probability values in the long-run, the coefficient of BTC is not significant at the 1% level. However, it is observed that the sign of the coefficient is positive. This means that the effect of BTC prices on TSA is positive. The formula for the established relationship in the long run is as follows:

$$TSA = EC + (0.0187 * BTC - 0.0873)$$

Table 9

ARDL Short Term Coefficients

Variables	Coefficient	Standart Error	t-Statistics	Prob.
ΔTSA(-1)	-0.274967	0.084666	-3.247655	0.0013***
ΔTSA (-2)	-0.217511	0.072183	-3.013332	0.0028***
ΔTSA(-3)	-0.117846	0.052696	-2.236326	0.0260**
ΔBTC	0.155024	0.080709	1.920765	0.0556*
ECT_{t-1}	-0.679491	0.093030	-7.304001	0.0000***

*Note. *, **, *** represent the significance level of 10%, 5%, and 1% respectively, and ECTt-1 refers to "Error Correction Term".*

The efficacy of the error correction term is paramount in the model's performance. To ensure its functionality, it necessitates a negative coefficient and a statistically significant probability value at the 0.01 level. In the current model, the coefficient stands at -0.67 with a probability value of 0.00, affirming the model's accuracy. Interpretatively, the coefficient of the error correction term suggests that a deviation existing in the long term between the variables will converge by 67% in the subsequent period, signifying imminent closure of the observed difference.

The examination of probability values associated with lagged terms for Twitter Sentiment Analysis (TSA) reveals noteworthy patterns. The first two periods exhibit significance at the 1% level, while the third lagged term achieves significance at the 5% level. Despite this significance, all three lagged terms manifest a negative relationship between variables in the short term. Specifically, a change in Twitter sentiment in one lagged term prompts a short-term decrease of 0.27%, followed by a 0.21% decrease for two lagged terms and a 0.11% decrease for three lagged terms.

Descriptive test outcomes shed light on the model's explanatory power, encapsulated in the adjusted R-squared value. In the established model, 47% of the variance in Bitcoin (BTC) price is explicable. Furthermore, the absence of autocorrelation issues is evident in the Durbin-Watson statistic, registering at 1.99, surpassing the 0.01 threshold, indicating a lack of serial correlation.

5. Conclusion and Discussions

In the realm of cryptocurrency research, this study contributes noteworthy insights to the existing literature by delving into the intricate relationship between Twitter Sentiment Analysis (TSA) and Bitcoin (BTC) prices. Through robust econometric analysis, the research elucidates the nuances of this relationship in both the long and short term. The findings, particularly the unexpected absence of significant coefficients in the long term and the discernible positive relationship in the short term, challenge prevailing assumptions and enrich the discourse on the dynamics between social media sentiment and cryptocurrency market movements.

Moreover, the study advances our understanding by undertaking a novel approach of reciprocal analysis, swapping the roles of dependent and independent variables. This methodological innovation reveals that while Twitter discussions may not significantly impact BTC prices when the cryptocurrency is the dependent variable, the reverse scenario indicates a tangible influence of Twitter sentiment on short-term price fluctuations. This nuanced exploration provides a nuanced perspective on the multifaceted interplay between social media discourse and cryptocurrency valuations.

Furthermore, the study addresses a critical gap in the literature by scrutinizing the potential manipulative effects of Twitter on BTC price volatility. By systematically analyzing the influence of Twitter content on cryptocurrency market dynamics, the research contributes empirical evidence to the ongoing discussions surrounding the susceptibility of digital assets to external manipulation via social media platforms. The findings, indicating the absence of manipulative effects in the studied period, challenge prevalent concerns and underscore the need for a nuanced understanding of the factors driving BTC price volatility.

The study's findings conclusively demonstrate that social media manipulation has no significant impact on BTC pricing in both the short and long term. Consequently, investors are advised to prioritize enhancing their financial literacy rather than relying on tracking social media traders when making investment decisions. From these perspectives, the study makes a substantial contribution to the existing literature. It emphasizes the importance of a discerning and informed approach to investment, suggesting that investors would benefit more from improving their financial literacy skills rather than basing decisions on social media trends. This insight not only adds to the scholarly discourse but also provides practical guidance for investors navigating the dynamic landscape of cryptocurrency markets.

In sum, this study adds substantial value to the literature by offering novel empirical insights, challenging existing assumptions, introducing innovative methodological approaches, and addressing pertinent gaps in the discourse on cryptocurrency valuation dynamics in the context of social media influence. The nuanced findings not only contribute to the academic understanding of the field but also provide practical implications for investors and market analysts navigating the volatile landscape of cryptocurrency markets.

References

Baykal, A. (2006). Veri madenciliği uygulama alanları. *D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi, (7),* 95-107.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Royal Statistical Society, 37*(2), 149-192.

Chalkiadakis, I. G., Peters, W., & Ames, M. (2021). Hybrid ARDL-MIDAS-transformer time-series regressions for multi-topic crypto market sentiment driven by price and technology factors. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3908066 (Accessed on: 18.04.2022).

Chen, N. (2022). Bitcoin price and miner behavior: An application of ARDL model. *Journal of Financial Studies & Research, 2022,* 1-16.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427-431.

Enders, W., & Lee, J. (2012). The flexible Fourier form and Dickey–Fuller type unit root tests. *Economics Letters, 117*(1), 196-199.

Engle, R. F., & Granger, C. W. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica, 55*(2), 251-276.

Gaies, B., Nakhli, M. S., Sahut, J. M., & Guesmi, K. (2021). Is Bitcoin rooted in confidence? – Unraveling the determinants of globalized digital currencies. *Technological Forecasting & Social Change, 172,* 1-11.

Galeshchuk, S., Vasylchyshyn, O., & Krysovatyy, A. (2018). Bitcoin response to Twitter sentiments. *ICTERI Workshops,* Kiev, Ukraine, pp.160-168.

Gözbaşi, O., Altinöz, B., & Şahin, E. E. (2021). Is Bitcoin a safe haven? A study on the factors that affect Bitcoin prices. *International Journal of Economics and Financial Issues, Econjournals, 11*(4), 35-40.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-292.

Katsafados, A. G., Nikoloutsopoulos, S., & Leledakis, G. N. (2023). Twitter sentiment and stock market: A COVID-19 analysis. *Journal of Economic Studies*. https://doi.org/10.1108/JES-09-2022-0486.

Kjærland, F., Meland, M., Oust, A., & Øyen, V. (2018). How can Bitcoin price fluctuations be explained?. *International Journal of Economics and Financial Issues, 8*(3), 323-332.

Li, X., & Wang, C. A. (2021). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems, 95,* 49-60.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703–708.

NLTK. (2023). Documentation. https://www.nltk.org/ (Accessed on: 19.04.2023).

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16*(3), 289-326.

Phillips, C. B. P., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75*(2), 335-346.

Polat, A. Y. (2021). Davranışsal finans teorisi ve kripto paralar. Efe Akademi.

Rajput, S. K. O., Soomro, I. A., & Soomro, N. A. (2022). Bitcoin sentiment index, Bitcoin performance and US dollar exchange rate. *Journal of Behavioral Finance, 23*(2), 150-165.

Sattarov, O., Jeon, H. S., Oh, R., & Lee, J. D. (2020). Forecasting Bitcoin price fluctuation by Twitter sentiment analysis. *2020 International Conference on Information Science and Communications Technologies (ICISCT),* Taşkent, Özbekistan, 1-4.

Shahvari, N. (2022). The relation between gold price movement and Bitcoin investment sentiment. *International Journal of Management, Accounting and Economics, 9*(9), 566-582.

Wong, E. L. X. (2021). Prediction of Bitcoin prices using Twitter data and natural language processing. *Duke University Libraries Duke Space*. https://dukespace.lib.duke.edu/dspace/handle/10161/24081 (Accessed on: 18.04.2023).