



Influencing cryptocurrency: analyzing celebrity sentiments on X (formerly Twitter) and their impact on bitcoin prices

Takeshi Inuduka¹ · Akihito Yokose² · Shunsuke Managi¹

Received: 9 January 2024 / Accepted: 8 February 2024 / Published online: 19 April 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

This paper explores the dynamic relationship between Bitcoin prices and the sentiments expressed by celebrities on X (formerly Twitter), employing a Vector Autoregression (VAR) model. The aim of this study is to understand the impact of a diverse range of sentiments (positive, negative, valence, arousal, dominance) extracted from the posts of 30 influential celebrities in the Bitcoin community on future Bitcoin prices and volumes. The analysis reveals that negative sentiments have a statistically significant impact on Bitcoin price fluctuations. Furthermore, while the shocks to price caused by sentiments converge within approximately 2 weeks, it was found that in the short term, spanning just 2 days, positive sentiments and valence sentiments positively influence price fluctuations, whereas negative sentiments exert a stronger negative effect. This finding underscores the direct influence of celebrity sentiments on the short-term emotions and actions of Bitcoin market participants, supporting psychological researches that indicate a strong influence of negative information on individual cognition and behavior. The significance of this study lies in its broad analysis of Bitcoin price fluctuations using the sentiments of various influencers on social media, not limited to globally recognized figures, and shows one possible way to respond to an immature market. This study offers valuable insights for investors and market analysts in refining investment strategies and risk management by considering market sentiment fluctuations.

Keywords Bitcoin · Cryptocurrency · VAR · Sentiment analysis · Social media

✉ Takeshi Inuduka
takeshi@onegrit.com

Akihito Yokose
yokose@onegrit.com

Shunsuke Managi
managi.s@gmail.com

¹ Kyushu University, Fukuoka City, Fukuoka 819-0395, Japan

² University of Electro-Communications, Chofu, Tokyo 182-8585, Japan

JEL Classification C32 · G41 · G12

1 Introduction

Bitcoin, a pioneer in the realm of cryptocurrencies, was first conceptualized by Nakamoto (2008) and came into practical use in January 2009. Within its ecosystem, each transaction is transparently recorded in a ledger maintained by a decentralized network of nodes, eliminating the need for intermediaries such as central banks or governing institutions. Originally developed as a digital currency intended to replace traditional money, Bitcoin is now predominantly recognized and utilized as an innovative investment asset, as indicated by Baur et al. (2018). This recognition is further substantiated by several studies arguing that incorporating Bitcoin into traditional asset portfolios could enhance overall performance and offer diversification benefits. For instance, Platanakis and Urquhart (2020) conclude that including Bitcoin in a traditional portfolio of stocks and bonds can significantly increase risk-adjusted returns, while Akhtaruzzaman et al. (2020) demonstrate Bitcoin's capability to function as a hedge against risks in industrial portfolios and bonds. As of today, according to CoinMarketCap,¹ Bitcoin's market capitalization has soared to \$723 billion, with daily trading volumes exceeding \$14 billion, demonstrating its firm establishment as a formidable financial asset in the global market. Based on data from CompaniesMarketCap,² this market capitalization is comparable to the world's 10th largest corporate value in the stock market, ranking alongside companies like Meta Platforms, Berkshire Hathaway, and Tesla.

Despite its astonishing market capitalization, studies on the informational efficiency of the Bitcoin market suggest that it remains less mature compared to traditional stock markets. For instance, Vidal-Tomás and Ibañez (2018) and Sensoy (2019) argue that the informational efficiency of the Bitcoin market has been increasing over time, but Al-Yahyaee et al. (2018) contend that the Bitcoin market is inefficient when compared to gold, stock, and currency markets. On another note, a study by Ante (2023) tracking Elon Musk's statements found that his posts could increase Bitcoin's price by up to 16.9%. Yet, in the stock market, it is unlikely for an individual's remarks to have such a significant impact on a brand with such a high market capitalization as Bitcoin.

The emerging Bitcoin market exhibits several key differences from traditional stock markets. Firstly, as shown by Chaim and Laurini (2018) and Baur and Dimpfl (2021), the Bitcoin market is characterized by high price volatility, representing a significant risk for investors, yet also offering the potential for high returns. Secondly, the Bitcoin market operates 24/7, enabling investors worldwide to engage in trading at any time, unlike traditional stock markets which are limited to specific trading hours in certain regions. Thirdly, as Vidal-Tomás and Ibañez (2018) emphasized, Bitcoin is underpinned by decentralized ledger technology, which doesn't require central institutional management or intervention, making it less susceptible

¹ <https://coinmarketcap.com/currencies/bitcoin/>. Accessed: 2023-11-27.

² <https://companiesmarketcap.com/>. Accessed: 2023-11-27.

to the influences of central banks and governments compared to traditional markets. Lastly, the Bitcoin market, due to its novelty and innovative nature, attracts unique market participants and investment behaviors not typically seen in traditional stock markets. In this context, Kim et al. (2020), argue that individual investors play an extremely important role in the Bitcoin market, a setting where information asymmetry is more likely to occur. Considering these factors, if individual investors play a significant role in the nascent and immature Bitcoin market, it is plausible that many decisions are made based on statements by celebrities, hence positing the hypothesis that sentiments on social media could be a potent predictor of the Bitcoin market. Our study aims to generalize this hypothesis, not by focusing solely on well-known figures like Elon Musk, but by utilizing statements from multiple celebrities.

In the field of sentiment analysis, there has been a growing trend in recent years to utilize this approach in the stock market, particularly in conjunction with 'alternative data' that differs from traditional data such as financial reports and economic indicators. For instance, Baker and Wurgler (2006), empirically demonstrated that investor sentiment, measured using a top-down approach, significantly impacts individual companies and the stock market as a whole. Tetlock (2007), showed that pessimistic columns in the Wall Street Journal (WSJ) could predict downward pressure on the Dow Jones, followed by a market price return to fundamentals. The advancement of natural language processing (NLP) tools, especially those of machine learning, has enabled the extraction of diverse features from text. For example, Chen et al. (2022) elucidated the interplay between market trends and pandemic narratives by extracting various features from financial news, such as narrative intensity, textual sentiments and tones, and virality. Since the late 2000s, the rapid proliferation of social media platforms as tools for information sharing, communication, and collaboration has led to a surge in sentiment analysis using data from social networking services (SNS). Bollen et al. (2011), suggested the possibility that the collective mood reflected on X (formerly Twitter)³ could impact the stock market. Zhang et al. (2011), discovered a significant negative correlation between the proportion of emotional posts and major indices like the Dow Jones, NASDAQ, and S&P500. However, when it comes to the emerging Bitcoin market, while there are some available studies such as Mai et al. (2018) and Baig et al. (2019), detailed analyses addressing the effects of various sentiments extracted from SNS posts are still insufficient. This paper endeavors to bridge this gap by conducting a comprehensive time series analysis of Bitcoin, focusing on how varying sentiments expressed on social media platforms influence its market dynamics and price fluctuations.

This research employs the Vector Autoregression (VAR) model to analyze the dynamic interplay between Bitcoin prices and the sentiments expressed by celebrities. Initially introduced by Sims (1980) as a tool for empirical analysis of macroeconomic dynamics, the VAR model is founded on the philosophy of 'letting the data speak for themselves'. This approach allows us to understand the magnitude and direction of influence and interactions without pre-assuming any specific trends, purely based on the data. The VAR model is particularly adept at capturing

³ In July 2023, Twitter was renamed to "X". This paper will use the name "X" for clarity and consistency.

the dynamic relationships among multiple time series data using the lagged values (lags) of each variable. This makes it suitable for modeling the potential mutual influences between Bitcoin prices and sentiment over time. Furthermore, the VAR model is capable of handling the endogenous relationships between variables, effectively analyzing internal dynamics such as how Bitcoin prices may influence sentiment, which in turn, feeds back into the prices. Our research also leverages tools like the Granger causality test, which helps in verifying whether one time series significantly influences another, and the impulse response function, which tracks over time how a shock in one variable impacts others, thereby providing a comprehensive analysis of the relationship between Bitcoin prices and sentiment.

In this study, we conduct a detailed analysis using the VAR model to examine the impact of sentiment data from celebrities on X, specifically positive and negative sentiments, on Bitcoin price fluctuations. The research aims to elucidate the unique characteristics of the emerging Bitcoin market, providing insightful information for market participants and researchers. This paper is structured as follows: in Sect. 2, we describe data and variables, while Sect. 3 elucidates the relevant methodology. We present results in Sect. 4 and conclude the paper with the final remarks in Sect. 5.

2 Data and variables

This section describes the data and variables. Following the methodology of Tetlock (2007), which involved including lagged volume as a variable in regression to capture the liquidity effect, this study utilizes data on Bitcoin volume. Specifically, the data encompasses Bitcoin prices and trading volumes, as well as X posts by celebrities in the Bitcoin community. It should be noted that the X posts were collected using the official API of the “Academic Research product track,” a program designed for researchers. However, due to the discontinuation of API access in 2023, the collection of posts was limited to a span of 36 months, from July 1, 2019, to June 30, 2022. The gathered data were subjected to initial processing, resulting in the preparation of seven types of daily time-series data.

2.1 Log return on Bitcoin prices (P_t)

Bitcoin can be traded 24/7 on many exchanges. Daily historical data was obtained using the API provided by the U.S. company Coinbase.⁴ This historical data includes the opening, closing, high, low prices, and the trading volume of Bitcoin in U.S. dollars. From this data, the closing price at time step t was used as the daily Bitcoin price, denoted as $close_t$, and the logarithmic return on Bitcoin price, P_t , was defined as follows:

⁴ <https://www.coinbase.com>. Accessed: 2023-11-27.

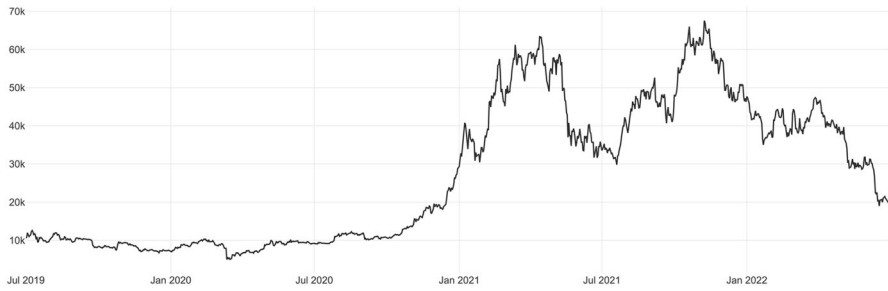


Fig. 1 The trend of $close_t$: Bitcoin prices (01/07/2019–30/06/2022)

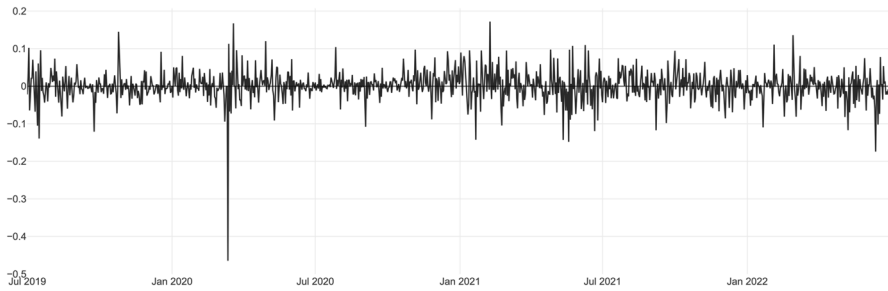


Fig. 2 The trend of P_t : Log return on Bitcoin prices (01/07/2019–30/06/2022)

Table 1 Descriptive statistics for variables $close_t$ and P_t

Variables	Mean	Std. Dev	Min	Max
$close_t$	27,144.4	18,446.3	4,970.8	67,566.8
P_t	0.0006	0.0395	-0.4647	0.1718

$$P_t = \ln\left(\frac{close_t}{close_{t-1}}\right) \tag{1}$$

As Banerjee et al. (1993) have noted, using logarithmic returns resolves the issue of non-stationarity, enabling the application of statistical methods that assume stationarity. Furthermore, logarithmic returns are helpful in transforming non-stationary data into stationary data, which is essential for econometric analysis. The dataset covers the period from July 1, 2019, to June 30, 2022. Figures 1 and 2 illustrate the trends of $close_t$ and P_t , respectively. During this period, the percentage of days when the price exceeded the previous day was 51.37%, and the days it was lower were 48.54%. Other characteristics are presented in Table 1.

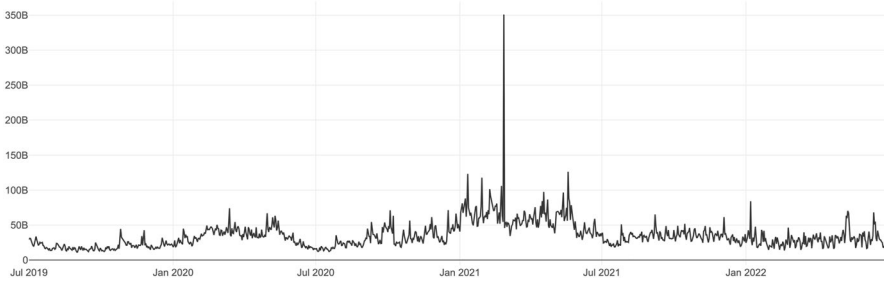


Fig. 3 The trend of $volume_t$: Bitcoin volumes (01/07/2019–30/06/2022)

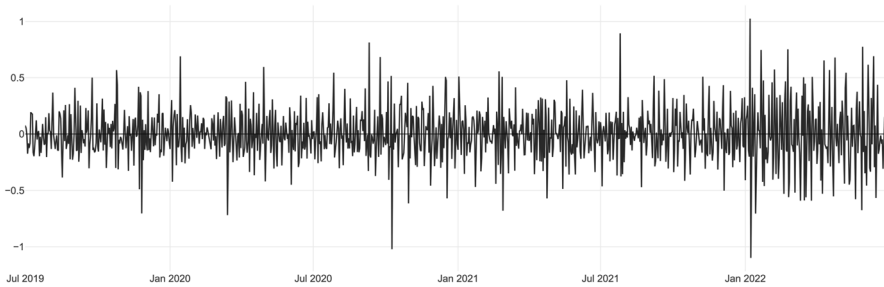


Fig. 4 The trend of V_t : Log return on Bitcoin volumes (01/07/2019–30/06/2022)

Table 2 Descriptive statistics for variables $volume_t$ and V_t

Variables	Daily mean	Daily std. dev	Daily min	Daily max
$volume_t$	34,836,887,655	19,205,971,765	11,445,355,859	350,967,941,479
V_t	– 0.0000	0.2282	– 1.0986	1.0250

2.2 Log return on Bitcoin volumes (V_t)

The acquired historical data encompass the daily trading volumes from July 1, 2019, to June 30, 2022. These volumes are defined as daily Bitcoin volumes $volume_t$ at time step t and are illustrated in Fig. 3. During the period, it was observed that 47.35% of the volumes exceeded the previous day's volumes, while 52.55% fell below.

As evident from the figure, multiple instances of data points significantly deviating from the average were observed in the volume data. To enhance the stability of the model, outlier detection and preprocessing were conducted as part of the statistical treatment. Specifically, data points with an absolute Z-score exceeding 3 were identified as outliers and replaced with values corresponding to a Z-score of 3. This preprocessing resulted in the replacement of 10 data points (0.84% of the total). The

Table 3 Selected X accounts of 30 celebrities from the Bitcoin community

#	X account	#	X account
1	@adam3us	16	@dergigi
2	@jackmallers	17	@DocumentingBTC
3	@BitcoinIsSaving	18	@nvk
4	@ODELL	19	@LynAldenContact
5	@saylor	20	@BitcoinMagazine
6	@jack	21	@starkness
7	@MartyBent	22	@real_vijay
8	@stephanlivera	23	@Excellion
9	@nic_carter	24	@TuurDemeester
10	@lopp	25	@TheGuySwann
11	@bitstein	26	@BTCsessions
12	@gladstein	27	@hodlonaut
13	@PrestonPysh	28	@pwuille
14	@jimmysong	29	@parkeralewis
15	@saifedean	30	@_benkaufman

preprocessed daily volumes at time step t are defined as *adjusted volume_t*. The log returns, denoted as V_t , were defined as follows.

$$V_t = \ln\left(\frac{\text{adjusted volume}_t}{\text{adjusted volume}_{t-1}}\right) \tag{2}$$

The trend of V_t is presented in Fig. 4. The characteristics of both *volume_t* and V_t are summarized in Table 2.

2.3 Positive/Negative post rate by Bitcoin celebrities (S^{pos}_t/S^{neg}_t)

For this study, we selected the X accounts of 30 celebrities based on rankings from the Bitcoin community by Hive.one,⁵ as displayed in Table 3. Hive.one ranks individuals primarily based on their X follower graph, facilitating the discovery of reputable accounts within specific communities. All posts from these selected accounts, from July 1, 2019, to June 30, 2022, were collected using the official X API.

For the analysis of positive and negative sentiments, this study employed the VADER (Valence Aware Dictionary and sEntiment Reasoner) model, developed by Hutto and Gilbert (2014). VADER is a rule-based model designed for the sentiment analysis of social media text and constructs a gold-standard sentiment lexicon with over 7,500 lexical features, including words, emoticons, slang, and acronyms. The VADER lexicon is particularly attuned to the context of microblogging platforms such as X. Therefore, VADER has been used in various studies in the field of data science using social media sentiment, including Valdez et al. (2020) and Bouktif et al. (2020). The compound score, a key metric used in VADER for social media

⁵ <https://hive.one/c/Bitcoin>. Accessed: 2022-5-31.

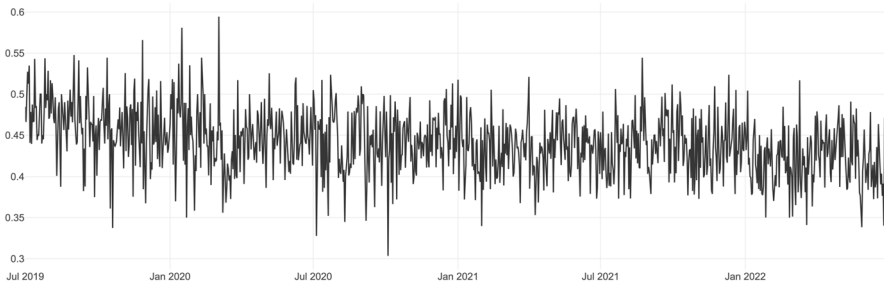


Fig. 5 The trend of S_t^{pos} ; Positive post rate by Bitcoin celebrities (01/07/2019–30/06/2022)

text sentiment analysis, is calculated by normalizing the sum of valence scores of each token within the text, as shown in the following formula:

$$\text{compound score} = \frac{\sum \text{valence score of token}}{\sqrt{(\sum \text{valence score of token})^2 + \alpha}} \quad (3)$$

In this formula, α represents a constant for normalization. The compound score indicates the overall sentiment strength of a given text, ranging from -1 (most extreme negative) to +1 (most extreme positive). In this study, we employed the following thresholds as described in the documentation of vaderSentiment,⁶ a Python implementation of VADER:

- Positive sentiment: compound score ≥ 0.05
- Neutral sentiment: compound score > -0.05 and < 0.05
- Negative sentiment: compound score ≤ -0.05

Sentiment analysis was performed on the collected posts from the 30 celebrities, categorizing them into positive, neutral, and negative based on the above thresholds. Daily counts of positive (pos_t), negative (neg_t), and neutral (neu_t) posts were aggregated. Time series S_t^{pos} and S_t^{neg} were then calculated using the following formula:

$$S_t^{pos} = \frac{pos_t}{pos_t + neg_t + neu_t} \quad (4)$$

$$S_t^{neg} = \frac{neg_t}{pos_t + neg_t + neu_t} \quad (5)$$

The trends for S_t^{pos} and S_t^{neg} are presented in Figs. 5 and 6, respectively. Descriptive statistics for pos_t , neg_t , neu_t are shown in Table 4, and for S_t^{pos} and S_t^{neg} in Table 5.

⁶ <https://github.com/cjhutto/vaderSentiment>. Accessed: 2023–11–27.

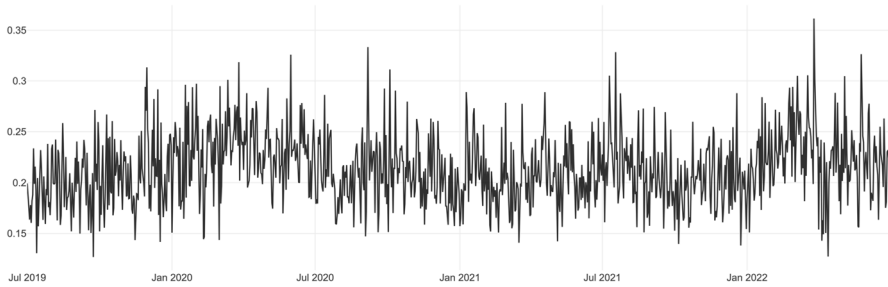


Fig. 6 The trend of S_t^{neg} ; Negative post rate by Bitcoin celebrities (01/07/2019–30/06/2022)

Table 4 Descriptive statistics for variables pos_t , neg_t , and neu_t

Variables	Total count	Daily mean	Daily std. dev	Daily min	Daily max
pos_t	189,832	173.20	53.91	61	468
neg_t	92,034	83.97	28.13	21	196
neu_t	149,849	136.72	48.33	35	458

Table 5 Descriptive statistics for variables S_t^{pos} and S_t^{neg}

Variables	Daily mean	Daily std. dev	Daily min	Daily max
S_t^{pos}	0.4411	0.0386	0.3035	0.5944
S_t^{neg}	0.2135	0.0342	0.1268	0.3615

2.4 Valence/Arousal/Dominance post rate by Bitcoin celebrities (S_t^{val} , S_t^{aro} , S_t^{dom})

In the field of psychological research, emotions are traditionally conceptualized through three fundamental elements: valence (the pleasantness), arousal (the intensity of emotion), and dominance (the degree of control). Albert Mehrabian initially introduced the concept of a three-dimensional approach to emotions in his work in the 1960s, focusing on the dimensions of pleasure-displeasure, arousal-nonarousal, and dominance-submissiveness. This framework was further developed and refined in collaboration with James A. Russell. Their collaborative work in the 1970s, such as Mehrabian and Russell (1974), laid the foundation for the VAD (Valence, Arousal, and Dominance) model.

To conduct a multifaceted analysis that goes beyond the positive–negative dimension, the three-dimensional aspects of VAD in psychology are utilized as variables to examine the dynamic relationship with Bitcoin prices and volumes. Sentiment analysis for valence, arousal, and dominance employs the NRC VAD Lexicon and

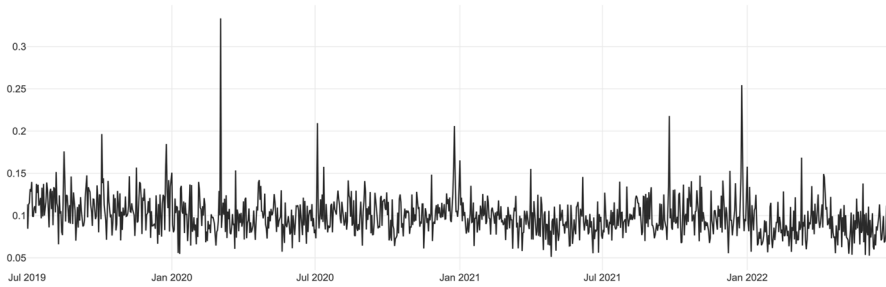


Fig. 7 The trend of S^{val}_t : Valence post rate by Bitcoin celebrities (01/07/2019–30/06/2022)

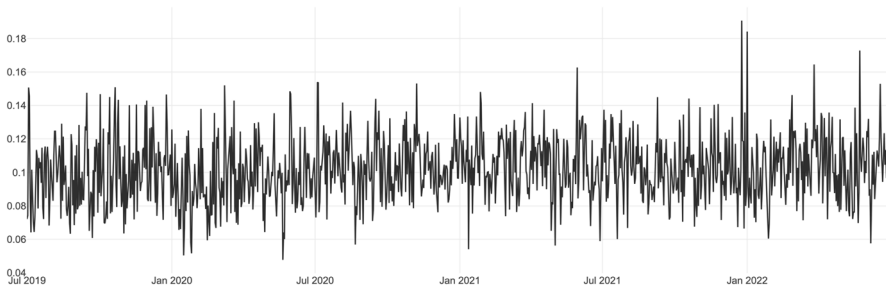


Fig. 8 The trend of S^{aro}_t : Arousal post rate by Bitcoin celebrities (01/07/2019–30/06/2022)

the Python package EmotionDynamics,⁷ summing the scores for each word to determine the overall average score for the text. The NRC VAD Lexicon, developed by Mohammad (2018), contains over 20,000 common English words, each scored for valence (ranging from 0: extremely unpleasant to 1: extremely pleasant), arousal (from 0: extremely sleepy/sluggish to 1: extremely activated/excited), and dominance (from 0: extremely powerful to 1: extremely weak). A sentiment is considered active if the scores calculated using the NRC VAD Lexicon exceed one standard deviation above the mean. We aggregate daily counts of active valence posts (val_t), arousal posts (aro_t), and dominance posts (dom_t), and defines the total daily posts as $all\ post_t$. Time series S^{val}_t , S^{aro}_t , S^{dom}_t are then established as follows:

$$S^{val}_t = \frac{val_t}{all\ post_t} \quad (6)$$

$$S^{aro}_t = \frac{aro_t}{all\ post_t} \quad (7)$$

⁷ <https://github.com/Priya22/EmotionDynamics>.

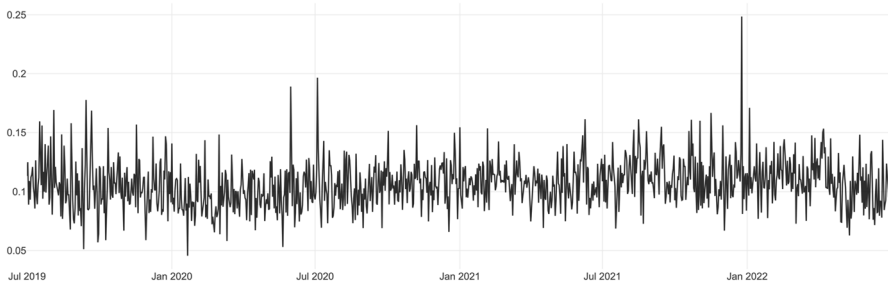


Fig. 9 The trend of S_t^{dom} : Dominance post rate by Bitcoin celebrities (01/07/2019–30/06/2022)

Table 6 Descriptive statistics for variables val_t , aro_t , and dom_t

Variables	Total count	Daily mean	Daily std. dev	Daily min	Daily max
val_t	42,419	38.70	14.12	10	143
aro_t	44,007	40.15	15.10	8	145
dom_t	46,160	42.12	15.80	9	181

Table 7 Descriptive statistics for variables S_t^{val} , S_t^{aro} , and S_t^{dom}

Variables	Daily mean	Daily std. dev	Daily min	Daily max
S_t^{val}	0.0993	0.0231	0.0512	0.3333
S_t^{aro}	0.1017	0.0195	0.0477	0.1908
S_t^{dom}	0.1067	0.0200	0.0456	0.2486

$$S_t^{dom} = \frac{dom_t}{all\ post_t} \tag{8}$$

The trends of S_t^{val} , S_t^{aro} , and S_t^{dom} are presented in Figs. 7, 8, and 9, respectively. The descriptive statistics for val_t , aro_t , and dom_t are provided in Table 6, while the descriptive statistics for S_t^{val} , S_t^{aro} , and S_t^{dom} are displayed in Table 7.

3 Analysis methods

In the analysis of this study, a three-variable Vector Autoregression (VAR) model is employed to investigate the potential interactions among selected variables. Specifically, to examine the impact of positive sentiment associated with celebrities, the variables P_t , V_t , and S_t^{pos} are utilized. Similarly, to explore the influence of negative sentiment, the variables P_t , V_t , and S_t^{neg} are used. Additionally, to assess the effects of VAD sentiment, three sets of variables are employed: P_t , V_t , and S_t^{val} ; P_t , V_t , and S_t^{aro} ; and P_t , V_t , and S_t^{dom} .

3.1 Data validation

Before estimating the VAR model, the stationarity of all time series data was tested using the Augmented Dickey–Fuller (ADF) test to avoid the possibility of spurious regression. Time series with unit roots are considered non-stationary and exhibit behaviors akin to a Random Walk, where past values significantly influence future values. Due to this characteristic, series with unit roots are challenging to predict. The importance of testing for stationarity, especially in the context of macroeconomic data, is emphasized by Mushtaq (2011). The “Dickey” and “Fuller” in the ADF test are referenced from the work of Dickey and Fuller (1979). They detailed the properties of time series with unit roots in autoregressive models and the distribution of associated statistics, offering a statistical criterion to ascertain whether a time series has a unit root. The ADF test extends this approach, incorporating a broader range of lags for a more comprehensive analysis.

3.2 VAR model

The general equation of a VAR model with p lags can be represented as follows:

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + u_t \tag{9}$$

Here, Y_t is a vector of k endogenous variables at time t . Each A_i (where $i=1,2,\dots,p$) is a $k \times k$ coefficient matrix. These matrices capture the influence that past values of the vector Y have on its current value. The term C is a $k \times 1$ vector of constants (intercepts). The u_t represents the error terms at time t , which are assumed to be white noise with a zero mean and a constant covariance matrix. The first model of this study, a three-variable VAR model of P_t , V_t , and S^{pos}_t , is expressed as:

$$P_t = C_1 + \sum_{i=1}^p \alpha_{1i} P_{t-i} + \sum_{i=1}^p \beta_{1i} V_{t-i} + \sum_{i=1}^p \gamma_{1i} S^{pos}_{t-i} + u_{1t} \tag{10}$$

$$V_t = C_2 + \sum_{i=1}^p \alpha_{2i} P_{t-i} + \sum_{i=1}^p \beta_{2i} V_{t-i} + \sum_{i=1}^p \gamma_{2i} S^{pos}_{t-i} + u_{2t} \tag{11}$$

$$S^{pos}_t = C_3 + \sum_{i=1}^p \alpha_{3i} P_{t-i} + \sum_{i=1}^p \beta_{3i} V_{t-i} + \sum_{i=1}^p \gamma_{3i} S^{pos}_{t-i} + u_{3t} \tag{12}$$

The second model, a three-variable VAR model of P_t , V_t , and S^{neg}_t , is as follows:

$$P_t = C_4 + \sum_{i=1}^p \alpha_{4i} P_{t-i} + \sum_{i=1}^p \beta_{4i} V_{t-i} + \sum_{i=1}^p \gamma_{4i} S^{neg}_{t-i} + u_{4t} \tag{13}$$

$$V_t = C_5 + \sum_{i=1}^p \alpha_{5i} P_{t-i} + \sum_{i=1}^p \beta_{5i} V_{t-i} + \sum_{i=1}^p \gamma_{5i} S^{neg}_{t-i} + u_{5t} \tag{14}$$

$$S^{neg}_t = C_6 + \sum_{i=1}^p \alpha_{6i} P_{t-i} + \sum_{i=1}^p \beta_{6i} V_{t-i} + \sum_{i=1}^p \gamma_{6i} S^{neg}_{t-i} + u_{6t} \tag{15}$$

The third model, a three-variable Vector Autoregression (VAR) model involving P_t , V_t , and S^{val}_t , is expressed by the following formula:

$$P_t = C_7 + \sum_{i=1}^p \alpha_{7i} P_{t-i} + \sum_{i=1}^p \beta_{7i} V_{t-i} + \sum_{i=1}^p \gamma_{7i} S^{val}_{t-i} + u_{7t} \tag{16}$$

$$V_t = C_8 + \sum_{i=1}^p \alpha_{8i} P_{t-i} + \sum_{i=1}^p \beta_{8i} V_{t-i} + \sum_{i=1}^p \gamma_{8i} S^{val}_{t-i} + u_{8t} \tag{17}$$

$$S^{val}_t = C_9 + \sum_{i=1}^p \alpha_{9i} P_{t-i} + \sum_{i=1}^p \beta_{9i} V_{t-i} + \sum_{i=1}^p \gamma_{9i} S^{val}_{t-i} + u_{9t} \tag{18}$$

Similarly, the formulas for the three-variable VAR models incorporating P_t , V_t , and S^{aro}_t , as well as P_t , V_t , and S^{dom}_t , can be represented, but these are omitted here.

The Akaike Information Criterion (AIC) is then employed to determine the appropriate lag length for the VAR models. AIC, proposed by Akaike (1974), is a statistical measure defined as follows:

$$AIC = \ln\left(\frac{SSR}{T}\right) + (K + 1)\frac{2}{T} \tag{19}$$

Here, T represents the sample size, SSR represents the sum of squared residuals, and K represents the number of explanatory variables.

Subsequently, the Granger causality test is applied to the estimated VAR model to evaluate the predictive capability of one endogenous variable on another. Established by Granger (1969), the Granger causality test is extensively utilized in econometrics and other fields to examine lead-lag relationships between time series variables.

Impulse Response Functions (IRFs) are utilized to track the impact of a single shock to one variable on the current and future values of endogenous variables within a VAR system. This approach enables a dynamic examination of how shocks to the system propagate over time and affect the levels of variables.

3.3 Evaluation methods

Conducting hypothesis testing on the estimated residuals, which serve as the disturbance terms, is critical for assessing the appropriateness, reliability, and statistical validity of VAR models. The standard assumptions regarding the disturbance term u_t in VAR models are expressed in the following forms:

$$E(u_t) = 0, t = 1, 2, \dots, T \tag{20}$$

$$Var(u_t) = E(u_t u_t') = \sum_u = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \dots & \sigma_{kk} \end{bmatrix} \tag{21}$$

$$t = 1, 2, \dots, T \tag{22}$$

$$\text{Cov}(u_t, u_s) = E(u_t u_s) = 0, t \neq s \quad (23)$$

Equation (21) represents the multivariate version of homoscedasticity, signifying that the covariance matrix is uniform across time changes. Equation (23) indicates that the disturbance term u_t exhibits auto-uncorrelatedness. Furthermore, the assumption that the disturbance term u_t follows a normal distribution underpins methodologies such as the Maximum Likelihood Estimation. Consequently, we conducted three tests related to the assumptions of disturbance terms, focusing on: testing for auto-uncorrelation, homoscedastic variance, and normal distribution of the disturbance terms.

For the hypothesis testing of auto-uncorrelation in the disturbance term, methods such as the Ljung–Box test are applied. The Ljung–Box test, proposed by Ljung and Box (1978), is a Portmanteau test that aggregates the squared autocorrelations of residuals across multiple lags. By calculating its test statistic, the Ljung–Box test adheres to a chi-squared distribution under the null hypothesis. A statistically significant outcome of this test suggests the presence of autocorrelation in the residuals, indicating the potential for a VAR model to have not fully captured the fundamental dynamics of the time series data. Conversely, an insignificant result from the Ljung–Box test, confirming the absence of such autocorrelation, implies that the residuals are essentially random, thereby reinforcing the reliability of the VAR model.

For the hypothesis testing of homoscedasticity in disturbance terms, techniques such as the ARCH-LM test are implemented. The ARCH-LM test, developed by Engle (1982), is a widely-recognized approach for detecting Autoregressive Conditional Heteroscedasticity (ARCH) effects in regression residuals over a specific number of lags. It entails fitting an autoregressive model to the squared residuals from a VAR model, followed by using the Lagrange Multiplier method to evaluate the presence of time-varying volatility in these residuals.

For testing the hypothesis of normal distribution of disturbance terms, methods such as the Jarque–Bera test are employed. The Jarque–Bera test, developed by Jarque and Bera (1980), employs the Lagrange multiplier procedure, which is simple to compute and asymptotically distributed as a χ^2 .

4 Results

The analysis results using a three-variable VAR model with P_t , V_t , and S^{pos}_t are presented as follows. The unit root verification results via the ADF test are as shown in Appendix Table 14, confirming the absence of unit roots in all time series. From the AIC values displayed in Appendix Table 15, the appropriate lag order for the VAR model was determined to be 14. The coefficients, standard errors, and p-values for each variable in the VAR model with 14 lags are presented in Appendix Table 16, 17, and 18. Notably, the standard errors included in these tables utilize Newey–West standard errors, which are robust to heteroskedasticity and autocorrelation, in accordance with Newey and West (1987). Finally, the results of Granger

Table 8 Granger causality test results (P_t , V_t , and S_t^{POS})

Variables	p-value
$P_t \rightarrow V_t$	0.203
$P_t \rightarrow S_t^{POS}$	0.550
$V_t \rightarrow S_t^{POS}$	0.611
$V_t \rightarrow P_t$	0.795
$S_t^{POS} \rightarrow P_t$	0.098
$S_t^{POS} \rightarrow V_t$	0.489

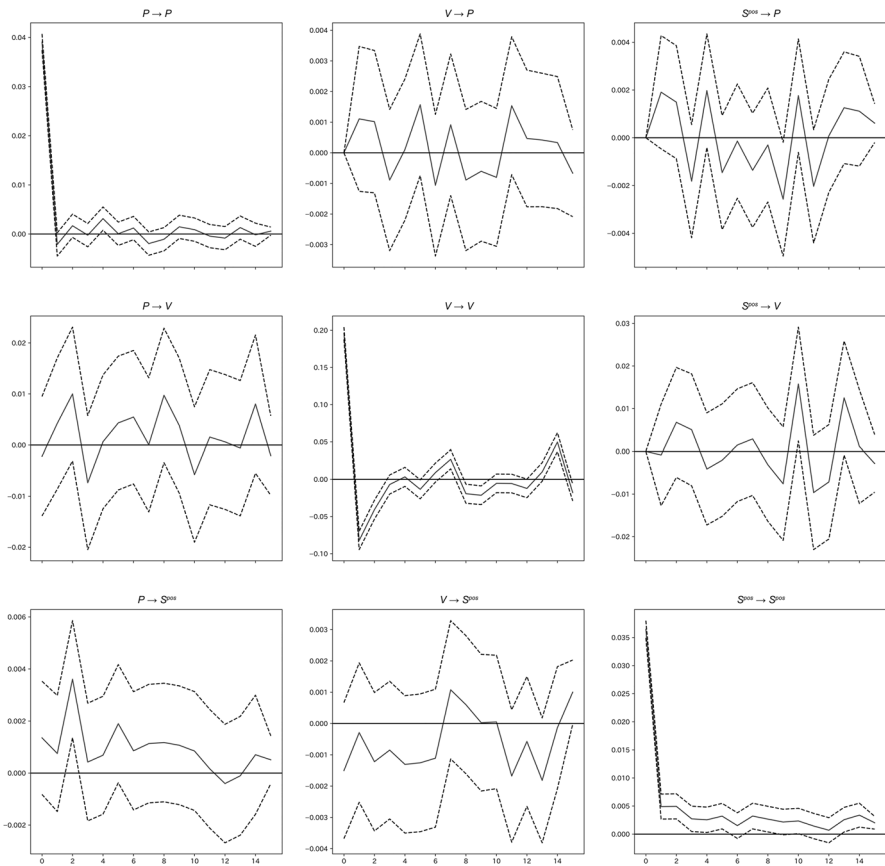


Fig. 10 Impulse response results (P_t , V_t , and S_t^{POS})

causality among pairs of variables are presented in Table 8, and the effects via impulse response are illustrated in Fig. 10. The results in Table 8 reveal that none of the variables had a statistically significant impact on the others. Figure 10 and Table 16 suggest that immediately following a shock, positive sentiments influence price fluctuations positively for the first two days. The impact of the shock gradually

Table 9 Granger causality test results (P_t , V_t , and S^{neg}_t)

Variables	p-value
$P_t \rightarrow V_t$	0.207
$P_t \rightarrow S^{neg}_t$	0.228
$V_t \rightarrow S^{neg}_t$	0.002 ^{***}
$V_t \rightarrow P_t$	0.707
$S^{neg}_t \rightarrow P_t$	0.015 ^{**}
$S^{neg}_t \rightarrow V_t$	0.066

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

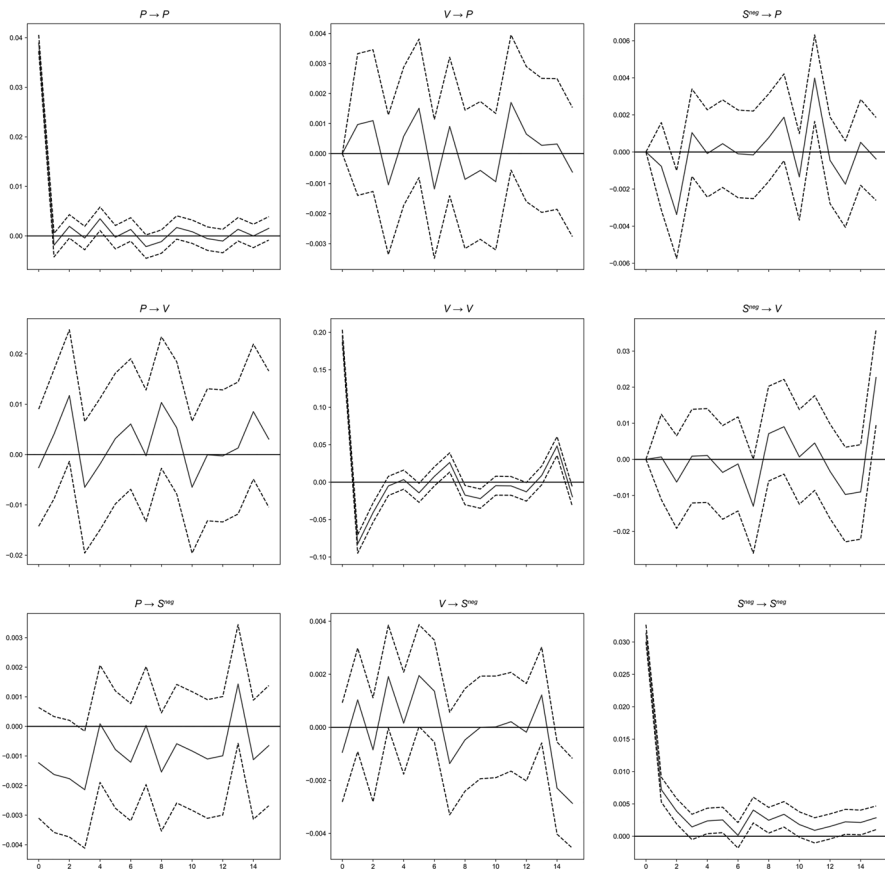


Fig. 11 Impulse response results (P_t , V_t , S^{neg}_t)

diminishes, but a significant shock reappears after the 9th day, and by the 14th day, it nearly converges.

Similarly, the analysis results for the three-variable VAR model with P_t , V_t , and S^{neg}_t are presented. The ADF test results in Appendix Table 14 confirm the absence

Table 10 Granger causality test results (P_t , V_t , and S^{val}_t)

Variables	p-value
$P_t \rightarrow V_t$	0.319
$P_t \rightarrow S^{val}_t$	0.114
$V_t \rightarrow S^{val}_t$	0.411
$V_t \rightarrow P_t$	0.706
$S^{val}_t \rightarrow P_t$	0.004**
$S^{val}_t \rightarrow V_t$	0.853

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

of unit roots in all time series. The appropriate lag order for the VAR model, determined from the AIC values shown in Appendix Table 20, was 15. The coefficients, standard errors, and p-values for each variable in the VAR model with 15 lags are presented in Appendix Tables 21, 22 and 23, along with the results of Granger causality among the variables in Table 9, and the effects via impulse response in Fig. 11. The results in Table 9 reveal that negative sentiments from celebrities have a statistically significant impact on future price fluctuations, and volume changes significantly affect future negative sentiments. It is noteworthy to add that when the same VAR analysis was conducted with the period split into two 18-month intervals, the former characteristic was more pronounced in the latter half with higher price volatility, while the latter characteristic was evident in the first half. Figures 11 and Table 21 show that negative sentiments negatively affect price fluctuations up to the second day post-occurrence, with a particularly significant negative shock observed on the second day. The shock diminishes after the third day but, akin to positive sentiments, a significant shock reemerges after the 9th day and nearly converges by the 15th day.

Next, the analysis results of a three-variable VAR model using VAD sentiment are presented. The analysis of P_t , V_t , and valence sentiment S^{val}_t includes ADF test results in Appendix Table 24, AIC values in Appendix Table 25, statistical values for each variable in the VAR model in Appendix Tables 26, 27, 28, Granger causality results in Table 10, and impulse response in Fig. 12. For the analysis of P_t , V_t , and arousal sentiment S^{aro}_t , ADF test results are in Appendix Table 29, AIC values in Appendix Table 30, statistical values in Appendix Tables 31, 32, 33, Granger causality results in Table 11, and impulse response in Fig. 13. The analysis of P_t , V_t , and dominance sentiment S^{dom}_t includes ADF test results in Appendix Table 32, AIC values in Appendix Table 33, statistical values in Appendix Table 34–36, Granger causality results in Table 12, and impulse response in Fig. 14. The only statistically significant Granger causality observed was from S^{val}_t to P_t . Examining the impulse response of S^{val}_t to P_t , it is noted that in the immediate aftermath of the shock, the first two days show a positive impact on price fluctuations. This characteristic is consistent with the impulse response from positive sentiment S^{pos}_t to P_t . Additionally, a significant spike is observed after the 9th day, a feature common to both the impulse responses from positive sentiment S^{pos}_t and negative sentiment S^{neg}_t to P_t .

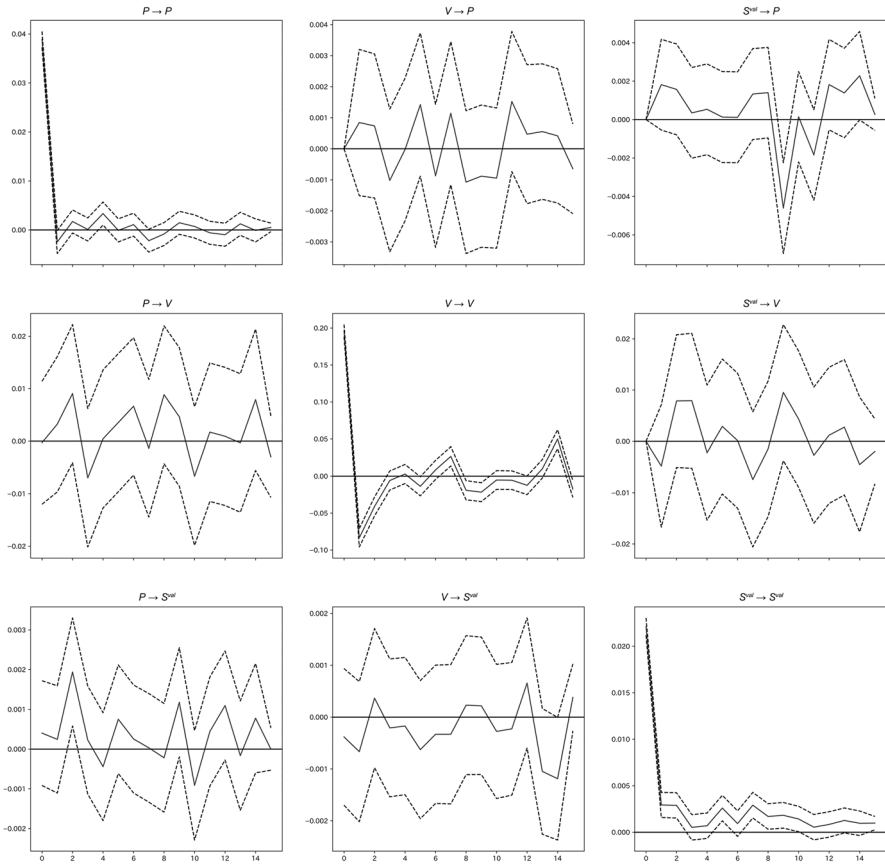


Fig. 12 Impulse response results (P_t , V_t , and S^{aro}_t)

Table 11 Granger causality test results (P_t , V_t , and S^{aro}_t)

Variables	p-value
$P_t \rightarrow V_t$	0.448
$P_t \rightarrow S^{aro}_t$	0.957
$V_t \rightarrow S^{aro}_t$	0.746
$V_t \rightarrow P_t$	0.526
$S^{aro}_t \rightarrow P_t$	0.267
$S^{aro}_t \rightarrow V_t$	0.288

Furthermore, to compare the magnitude of the shock caused by one unit of sentiment on future price fluctuations in the first two days across the five VAR models, the impulse response values were cumulated, as shown in Table 13. It was demonstrated that the impact of negative sentiment is approximately 1.2 times greater than that of positive sentiment. The magnitude of the impact over two days caused by

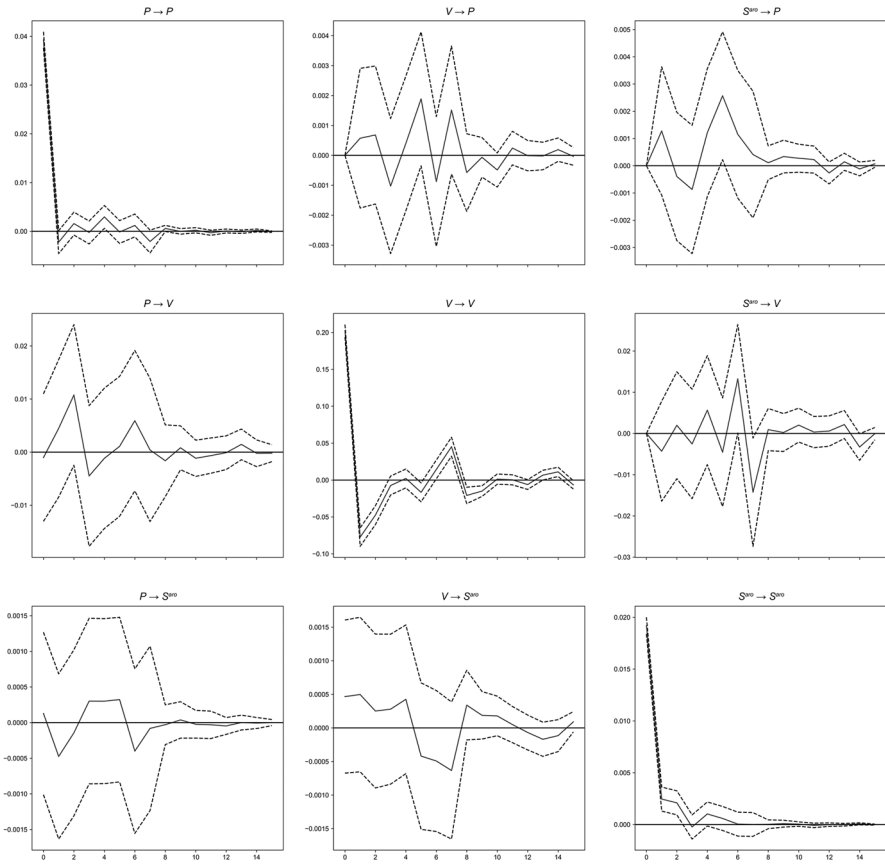


Fig. 13 Impulse response results (P_t , V_t , and S^{aro}_t)

Table 12 Granger causality test results (P_t , V_t , and S^{dom}_t)

Variables	p-value
$P_t \rightarrow V_t$	0.210
$P_t \rightarrow S^{dom}_t$	0.314
$V_t \rightarrow S^{dom}_t$	0.726
$V_t \rightarrow P_t$	0.743
$S^{dom}_t \rightarrow P_t$	0.079
$S^{dom}_t \rightarrow V_t$	0.886

valence sentiment was almost the same as that of positive sentiment. The impacts over two days of arousal sentiment and dominance sentiment were smaller compared to other sentiments.

The results of hypothesis tests for the residuals of the VAR model are all presented in Tables 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, and 53 of

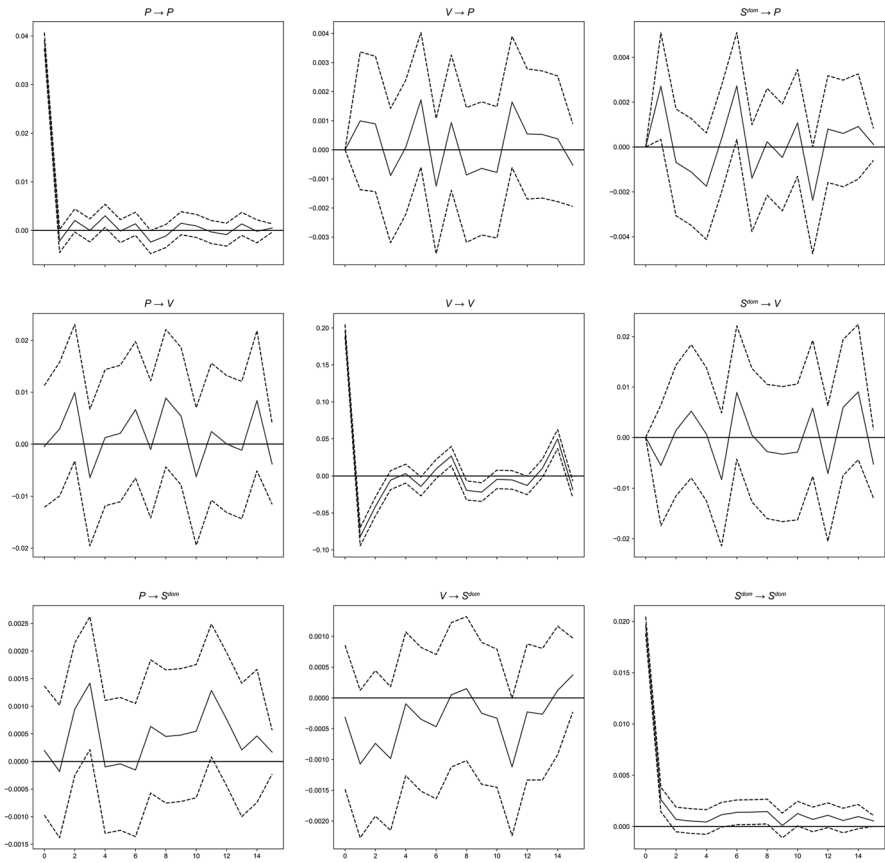


Fig. 14 Impulse response results (P_t , V_t , and S^{dom}_t)

Table 13 Impulse response values for the first 2 days after shock

Lag	$S^{pos}_t \rightarrow P_t$	$S^{neg}_t \rightarrow P_t$	$S^{val}_t \rightarrow P_t$	$S^{aro}_t \rightarrow P_t$	$S^{dom}_t \rightarrow P_t$
1	0.0019	- 0.0008	0.0018	0.0013	0.0027
2	0.0015	- 0.0034	0.0016	- 0.0004	- 0.0007
Amount of change after 2 days	0.0034	- 0.0041	0.0034	0.0009	0.0020

Appendix 2. While the Ljung–Box test results were favorable, there were instances in the results of the ARCH-LM and Jarque–Bera tests where the null hypothesis was rejected.

5 Conclusions

This study analyzed the impact of X sentiments (positive, negative, valence, arousal, dominance) of 30 influential celebrities in the Bitcoin community on future Bitcoin prices and volumes using a three-variable VAR model. The Granger causality tests revealed that negative sentiments significantly influence future Bitcoin price fluctuations. Analysis of impulse responses indicated that positive and valence sentiments lead to a positive impact on price fluctuations for up to two days following the sentiment's occurrence, whereas negative sentiments have a negative impact for the same duration. These results suggest that the sentiments expressed by influential celebrities directly affect the short-term emotions and actions of Bitcoin market participants. Furthermore, it was observed that the impact of all sentiments on price fluctuations tends to converge within approximately two weeks, indicating that celebrities' sentiments do not have a lasting effect on future Bitcoin prices. This outcome aligns with the "Sentiment Theory", suggesting that while celebrities' statements on social media may not contain intrinsic information about future Bitcoin prices, they could include information about market sentiment. Another intriguing finding from these results is that the impact of negative sentiments on immediate price fluctuations was greater than that of positive sentiments. This aligns with psychological research, like that of Baumeister et al. (2001), which shows that negative information has a stronger effect on individual cognition and behavior than positive information, supporting the findings of this study. Additionally, these results are consistent with the loss aversion concept of Prospect Theory proposed by Kahneman and Tversky (1979), indicating that market participants are overly sensitive to potential losses and risks, potentially leading to significant price volatility. However, it is important to note, as highlighted by Loughran and McDonald (2016), the ambiguous use of positive words in English, particularly in financial documents. While negative words are rarely negated to form positive expressions, positive words are frequently used to construct negative statements. Therefore, sentiments classified as positive may carry ambiguity, and positive sentiments may not necessarily have a positive impact on the market.

The significance of this study lies in its analysis of Bitcoin price fluctuations using X sentiments of not just globally recognized figures like Elon Musk but also those influencers who garner widespread support from market participants. This research holds importance for investors and market analysts in refining investment strategies and risk management, and making more effective investment decisions based on market sentiment fluctuations. Compared to similar research by Kraaijeveld and De Smedt (2020), this study introduces novelty by employing five diverse emotions for a multifaceted analysis. Additionally, while their research had to remove bot-generated posts from numerous unspecified accounts, this study benefits from using fixed accounts, thus eliminating the need to consider bots, significantly reducing the analytical effort, which could be advantageous for investors.

However, this study has several limitations. Firstly, there is the validity of the selected celebrity accounts. The representativeness of these accounts as opinion leaders in the Bitcoin market is not sufficiently verified, necessitating further analysis with data from different periods and various accounts. Secondly, there is a lack of consideration for the characteristics of X. X has features like “likes,” “reposts,” and commenting, which allow opinions to form and spread over time. One possible explanation for this could be that statements with substantial impact are spread through reposts on X, leading to a viral effect that amplifies and subsequently results in noticeable market reactions. However, detailed analysis on how quickly and extensively X’s viral effects influence users has been largely unexplored, remaining largely speculative. This remains an important area for future research. Thirdly, the study does not consider the immediate impact of X sentiment. While it follows the framework of VAR analysis that considers only the impact up to the previous day, excluding the impact on the same day, the potential immediate effects of X sentiment have not been addressed. Future empirical research will also have to deal with this aspect. Lastly, the study does not account for the intensity of sentiment. Developing methods to incorporate sentiment intensity into the analysis could lead to more accurate predictions of market trends. Future research is expected to address these issues and construct a more comprehensive and accurate predictive model for Bitcoin market trends.

As of December 2023, the United States has not yet approved a spot Bitcoin ETF (Exchange-Traded Fund), but its eventual realization could be a structural change for the Bitcoin market. The introduction of a spot ETF is expected to broadly open Bitcoin investment to both retail and institutional investors, potentially encouraging an influx of new capital. Moreover, it could signify the wider acceptance of Bitcoin as a conventional financial asset, marking an important step towards the maturation of the market. This study, conducted against the backdrop of an immature Bitcoin market, has demonstrated one potential approach to this nascent field. The approval of a spot Bitcoin ETF not only presents the possibility for the Bitcoin market to mature but also paves the way for future analysis.

Appendix 1. Additional tables for the analysis

See Tables [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#), [22](#), [23](#), [24](#), [25](#), [26](#), [27](#), [28](#), [29](#), [30](#), [31](#), [32](#), [33](#), [34](#), [35](#), [36](#), [37](#), [38](#).

Table 14 Augmented Dickey-Fuller test results (P_t , V_t , and S^{pos}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{pos}_t	0.000***

Note: *, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 15 AIC (P_t , V_t , and S^{pos}_t)

lag	AIC
0	- 15.95
1	- 16.04
2	- 16.14
3	- 16.16
4	- 16.16
5	- 16.21
6	- 16.23
7	- 16.24
8	- 16.23
9	- 16.23
10	- 16.22
11	- 16.21
12	- 16.22
13	- 16.25
14	- 16.26
15	- 16.25
16	- 16.24
17	- 16.24
18	- 16.23
19	- 16.22
20	- 16.22
21	- 16.21
22	- 16.21
23	- 16.20
24	- 16.19

Table 16 Estimation results for α , β , and γ in the P_t equation of P_t , V_t , and S^{pos}_t

Variables	Coefficient	Std err	p-value
Intercept	-0.0040	0.0301	0.895
P_{t-1} (lag 1)	-0.0554	0.0367	0.130
P_{t-2} (lag 2)	0.0376	0.0303	0.214
P_{t-3} (lag 3)	-0.0073	0.0299	0.808
P_{t-4} (lag 4)	0.0727	0.0468	0.121
P_{t-5} (lag 5)	0.0170	0.0252	0.499
P_{t-6} (lag 6)	0.0179	0.0295	0.543
P_{t-7} (lag 7)	-0.0471	0.0508	0.354
P_{t-8} (lag 8)	-0.0378	0.0276	0.171
P_{t-9} (lag 9)	0.0357	0.0263	0.174
P_{t-10} (lag 10)	0.0242	0.0276	0.381
P_{t-11} (lag 11)	0.0024	0.0337	0.944
P_{t-12} (lag 12)	-0.0196	0.0314	0.532
P_{t-13} (lag 13)	0.0302	0.0303	0.318
P_{t-14} (lag 14)	-0.0014	0.0339	0.967
V_{t-1} (lag 1)	0.0060	0.0056	0.280
V_{t-2} (lag 2)	0.0084	0.0058	0.148
V_{t-3} (lag 3)	0.0002	0.0063	0.974
V_{t-4} (lag 4)	0.0029	0.0057	0.609
V_{t-5} (lag 5)	0.0091	0.0056	0.102
V_{t-6} (lag 6)	-0.0003	0.0063	0.956
V_{t-7} (lag 7)	0.0061	0.0064	0.334
V_{t-8} (lag 8)	-0.0028	0.0064	0.662
V_{t-9} (lag 9)	-0.0055	0.0062	0.381
V_{t-10} (lag 10)	-0.0045	0.0060	0.455
V_{t-11} (lag 11)	0.0025	0.0065	0.695
V_{t-12} (lag 12)	0.0039	0.0068	0.571
V_{t-13} (lag 13)	0.0063	0.0058	0.278
V_{t-14} (lag 14)	0.0060	0.0064	0.351
S^{pos}_{t-1} (lag 1)	0.0523	0.0328	0.111
S^{pos}_{t-2} (lag 2)	0.0369	0.0271	0.173
S^{pos}_{t-3} (lag 3)	-0.0625	0.0318	0.050*
S^{pos}_{t-4} (lag 4)	0.0473	0.0372	0.204
S^{pos}_{t-5} (lag 5)	-0.0434	0.0383	0.257
S^{pos}_{t-6} (lag 6)	-0.0144	0.0306	0.637
S^{pos}_{t-7} (lag 7)	-0.0323	0.0269	0.230
S^{pos}_{t-8} (lag 8)	-0.0078	0.0307	0.800
S^{pos}_{t-9} (lag 9)	-0.0620	0.0553	0.262
S^{pos}_{t-10} (lag 10)	0.0580	0.0315	0.065
S^{pos}_{t-11} (lag 11)	-0.0471	0.0322	0.144
S^{pos}_{t-12} (lag 12)	0.0049	0.0297	0.869
S^{pos}_{t-13} (lag 13)	0.0564	0.0272	0.038*
S^{pos}_{t-14} (lag 14)	0.0239	0.0287	0.405

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 17 Estimation results for α , β , and γ in the V_t equation of P_t , V_t , and S^{pos}_t

Variables	Coefficient	Std err	p-value
Intercept	-0.1828	0.1354	0.177
P_{t-1} (lag 1)	0.0858	0.1924	0.656
P_{t-2} (lag 2)	0.2790	0.1688	0.098
P_{t-3} (lag 3)	-0.0509	0.1491	0.733
P_{t-4} (lag 4)	0.0231	0.1232	0.851
P_{t-5} (lag 5)	0.1099	0.1408	0.435
P_{t-6} (lag 6)	0.1882	0.1244	0.130
P_{t-7} (lag 7)	0.1338	0.1452	0.357
P_{t-8} (lag 8)	0.3034	0.1586	0.056
P_{t-9} (lag 9)	0.2527	0.1243	0.042*
P_{t-10} (lag 10)	0.04589	0.1496	0.759
P_{t-11} (lag 11)	0.1337	0.1236	0.280
P_{t-12} (lag 12)	0.0382	0.1481	0.797
P_{t-13} (lag 13)	0.0655	0.1362	0.631
P_{t-14} (lag 14)	0.3186	0.1376	0.021*
V_{t-1} (lag 1)	-0.4192	0.0350	0.000***
V_{t-2} (lag 2)	-0.3831	0.0353	0.000***
V_{t-3} (lag 3)	-0.2839	0.0317	0.000***
V_{t-4} (lag 4)	-0.1966	0.0400	0.000***
V_{t-5} (lag 5)	-0.2137	0.0321	0.000***
V_{t-6} (lag 6)	-0.1167	0.0353	0.001***
V_{t-7} (lag 7)	0.0340	0.0367	0.354
V_{t-8} (lag 8)	-0.0553	0.0374	0.139
V_{t-9} (lag 9)	-0.1198	0.0322	0.000***
V_{t-10} (lag 10)	-0.1376	0.0356	0.000***
V_{t-11} (lag 11)	-0.1692	0.0311	0.000***
V_{t-12} (lag 12)	-0.2022	0.0307	0.000***
V_{t-13} (lag 13)	-0.1523	0.0376	0.000***
V_{t-14} (lag 14)	0.0830	0.0323	0.010*
S^{pos}_{t-1} (lag 1)	-0.0229	0.1398	0.870
S^{pos}_{t-2} (lag 2)	0.1750	0.1686	0.299
S^{pos}_{t-3} (lag 3)	0.1693	0.1612	0.294
S^{pos}_{t-4} (lag 4)	-0.0395	0.1682	0.814
S^{pos}_{t-5} (lag 5)	-0.0215	0.1985	0.914
S^{pos}_{t-6} (lag 6)	0.0083	0.1837	0.964
S^{pos}_{t-7} (lag 7)	0.0861	0.1681	0.609
S^{pos}_{t-8} (lag 8)	-0.0601	0.1650	0.716
S^{pos}_{t-9} (lag 9)	-0.2548	0.1638	0.120
S^{pos}_{t-10} (lag 10)	0.3147	0.1504	0.036*
S^{pos}_{t-11} (lag 11)	-0.1468	0.1725	0.395
S^{pos}_{t-12} (lag 12)	-0.1993	0.1794	0.267
S^{pos}_{t-13} (lag 13)	0.3101	0.1505	0.039*
S^{pos}_{t-14} (lag 14)	0.0935	0.1467	0.524

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 18 Estimation results for α , β , and γ in the S^{pos}_t equation of P_t , V_t , and S^{pos}_t

Variables	Coefficient	Std err	p-value
Intercept	0.1598	0.0264	0.000***
P_{t-1} (lag 1)	0.0146	0.0279	0.602
P_{t-2} (lag 2)	0.0862	0.0267	0.001**
P_{t-3} (lag 3)	− 0.0007	0.0283	0.981
P_{t-4} (lag 4)	0.0011	0.0261	0.966
P_{t-5} (lag 5)	0.0391	0.0225	0.083
P_{t-6} (lag 6)	0.0051	0.0257	0.843
P_{t-7} (lag 7)	0.0117	0.0246	0.633
P_{t-8} (lag 8)	0.0188	0.0242	0.437
P_{t-9} (lag 9)	0.0111	0.0269	0.680
P_{t-10} (lag 10)	0.0082	0.0335	0.807
P_{t-11} (lag 11)	− 0.0110	0.0260	0.670
P_{t-12} (lag 12)	− 0.0221	0.0222	0.321
P_{t-13} (lag 13)	− 0.0071	0.0259	0.783
P_{t-14} (lag 14)	0.0135	0.0272	0.620
V_{t-1} (lag 1)	− 0.0005	0.0050	0.927
V_{t-2} (lag 2)	− 0.0054	0.0065	0.407
V_{t-3} (lag 3)	− 0.0060	0.0064	0.351
V_{t-4} (lag 4)	− 0.0090	0.0059	0.124
V_{t-5} (lag 5)	− 0.0091	0.0067	0.175
V_{t-6} (lag 6)	− 0.0098	0.0066	0.140
V_{t-7} (lag 7)	0.0009	0.0061	0.884
V_{t-8} (lag 8)	0.0027	0.0062	0.658
V_{t-9} (lag 9)	0.0015	0.0060	0.803
V_{t-10} (lag 10)	0.0025	0.0066	0.704
V_{t-11} (lag 11)	− 0.0061	0.0066	0.361
V_{t-12} (lag 12)	− 0.0034	0.0056	0.538
V_{t-13} (lag 13)	− 0.0107	0.0069	0.122
V_{t-14} (lag 14)	− 0.0065	0.0062	0.296
S^{pos}_{t-1} (lag 1)	0.1345	0.0309	0.000***
S^{pos}_{t-2} (lag 2)	0.1167	0.0310	0.000***
S^{pos}_{t-3} (lag 3)	0.0356	0.0334	0.287
S^{pos}_{t-4} (lag 4)	0.0373	0.0338	0.270
S^{pos}_{t-5} (lag 5)	0.0648	0.0311	0.038*
S^{pos}_{t-6} (lag 6)	0.0006	0.0312	0.984
S^{pos}_{t-7} (lag 7)	0.0615	0.0304	0.043*
S^{pos}_{t-8} (lag 8)	0.0415	0.0322	0.198
S^{pos}_{t-9} (lag 9)	0.0159	0.0312	0.609
S^{pos}_{t-10} (lag 10)	0.0253	0.0309	0.413
S^{pos}_{t-11} (lag 11)	0.0062	0.0307	0.841
S^{pos}_{t-12} (lag 12)	− 0.0189	0.0299	0.526
S^{pos}_{t-13} (lag 13)	0.0543	0.0296	0.066
S^{pos}_{t-14} (lag 14)	0.0611	0.0271	0.024*

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 19 Augmented Dickey-Fuller test results (P_t , V_t , and S^{neg}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{neg}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 20 AIC (P_t , V_t , and S^{neg}_t)

lag	AIC
0	- 16.18
1	- 16.32
2	- 16.41
3	- 16.42
4	- 16.44
5	- 16.48
6	- 16.51
7	- 16.54
8	- 16.53
9	- 16.53
10	- 16.52
11	- 16.52
12	- 16.53
13	- 16.56
14	- 16.56
15	- 16.57
16	- 16.56
17	- 16.55
18	- 16.54
19	- 16.53
20	- 16.53
21	- 16.52
22	- 16.51
23	- 16.50
24	- 16.49

Table 21 Estimation results for α , β , and γ in the P_t equation of P_t , V_t , and S^{neg}_t

Variables	Coefficient	Std err	p-value
Intercept	−0.0043	0.0151	0.777
P_{t-1} (lag 1)	−0.0483	0.0345	0.162
P_{t-2} (lag 2)	0.0429	0.0319	0.178
P_{t-3} (lag 3)	−0.0124	0.0306	0.686
P_{t-4} (lag 4)	0.0810	0.0491	0.099
P_{t-5} (lag 5)	0.0021	0.0250	0.933
P_{t-6} (lag 6)	0.0271	0.0281	0.335
P_{t-7} (lag 7)	−0.0555	0.0514	0.280
P_{t-8} (lag 8)	−0.0450	0.0272	0.098
P_{t-9} (lag 9)	0.0471	0.0264	0.074
P_{t-10} (lag 10)	0.0187	0.0263	0.476
P_{t-11} (lag 11)	−0.0002	0.0332	0.996
P_{t-12} (lag 12)	−0.0185	0.0308	0.547
P_{t-13} (lag 13)	0.0189	0.0302	0.532
P_{t-14} (lag 14)	0.0058	0.0343	0.865
P_{t-15} (lag 15)	0.0295	0.0270	0.274
V_{t-1} (lag 1)	0.0048	0.0056	0.388
V_{t-2} (lag 2)	0.0075	0.0057	0.182
V_{t-3} (lag 3)	−0.0004	0.0063	0.955
V_{t-4} (lag 4)	0.0036	0.0057	0.531
V_{t-5} (lag 5)	0.0106	0.0053	0.045*
V_{t-6} (lag 6)	−0.0011	0.0063	0.865
V_{t-7} (lag 7)	0.0076	0.0061	0.214
V_{t-8} (lag 8)	−0.0016	0.0064	0.800
V_{t-9} (lag 9)	−0.0041	0.0059	0.482
V_{t-10} (lag 10)	−0.0050	0.0061	0.408
V_{t-11} (lag 11)	0.0053	0.0069	0.441
V_{t-12} (lag 12)	0.0043	0.0068	0.527
V_{t-13} (lag 13)	0.0066	0.0060	0.270
V_{t-14} (lag 14)	0.0042	0.0067	0.528
V_{t-15} (lag 15)	−0.0010	0.0057	0.861
S^{neg}_{t-1} (lag 1)	−0.0246	0.0381	0.518
S^{neg}_{t-2} (lag 2)	−0.1032	0.0341	0.002**
S^{neg}_{t-3} (lag 3)	0.0566	0.0365	0.120
S^{neg}_{t-4} (lag 4)	0.0055	0.0387	0.886
S^{neg}_{t-5} (lag 5)	0.0112	0.0376	0.766
S^{neg}_{t-6} (lag 6)	0.0115	0.0365	0.753
S^{neg}_{t-7} (lag 7)	−0.0046	0.0350	0.896
S^{neg}_{t-8} (lag 8)	0.0256	0.0408	0.531
S^{neg}_{t-9} (lag 9)	0.0635	0.0336	0.059
S^{neg}_{t-10} (lag 10)	−0.0607	0.0498	0.223
S^{neg}_{t-11} (lag 11)	0.1399	0.0543	0.010**
S^{neg}_{t-12} (lag 12)	−0.0342	0.0358	0.339
S^{neg}_{t-13} (lag 13)	−0.0857	0.0327	0.009**

Table 21 (continued)

Variables	Coefficient	Std err	p-value
S_{t-14}^{neg} (lag 14)	0.0390	0.0347	0.261
S_{t-15}^{neg} (lag 15)	−0.0174	0.0423	0.681

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 22 Estimation results for α , β , and γ in the V_t equation of P_t , V_t , and S_t^{neg}

Variables	Coefficient	Std err	p-value
Intercept	0.0250	0.0638	0.695
P_{t-1} (lag 1)	0.0768	0.1958	0.695
P_{t-2} (lag 2)	0.3174	0.1674	0.058
P_{t-3} (lag 3)	−0.0135	0.1535	0.930
P_{t-4} (lag 4)	−0.0079	0.1318	0.952
P_{t-5} (lag 5)	0.0694	0.1395	0.619
P_{t-6} (lag 6)	0.1677	0.1235	0.175
P_{t-7} (lag 7)	0.1038	0.1512	0.492
P_{t-8} (lag 8)	0.2927	0.1592	0.066
P_{t-9} (lag 9)	0.2827	0.1280	0.027*
P_{t-10} (lag 10)	0.0333	0.1475	0.821
P_{t-11} (lag 11)	0.1162	0.1250	0.352
P_{t-12} (lag 12)	0.0417	0.1413	0.768
P_{t-13} (lag 13)	0.0643	0.1391	0.644
P_{t-14} (lag 14)	0.3176	0.1387	0.022*
P_{t-15} (lag 15)	0.1653	0.1591	0.299
V_{t-1} (lag 1)	−0.4229	0.0376	0.000***
V_{t-2} (lag 2)	−0.3892	0.0353	0.000***
V_{t-3} (lag 3)	−0.2799	0.0342	0.000***
V_{t-4} (lag 4)	−0.1957	0.0396	0.000***
V_{t-5} (lag 5)	−0.2144	0.0324	0.000***
V_{t-6} (lag 6)	−0.1249	0.0340	0.000***
V_{t-7} (lag 7)	0.0206	0.0374	0.581
V_{t-8} (lag 8)	−0.0503	0.0368	0.171
V_{t-9} (lag 9)	−0.1224	0.0337	0.000***
V_{t-10} (lag 10)	−0.1334	0.0343	0.000***
V_{t-11} (lag 11)	−0.1598	0.0296	0.000***
V_{t-12} (lag 12)	−0.1971	0.0291	0.000***
V_{t-13} (lag 13)	−0.1527	0.0371	0.000***
V_{t-14} (lag 14)	0.0664	0.0370	0.073
V_{t-15} (lag 15)	−0.0220	0.0352	0.532
S_{t-1}^{neg} (lag 1)	0.0207	0.1815	0.909
S_{t-2}^{neg} (lag 2)	−0.1953	0.2009	0.331
S_{t-3}^{neg} (lag 3)	0.0092	0.1870	0.961
S_{t-4}^{neg} (lag 4)	0.0250	0.2017	0.901
S_{t-5}^{neg} (lag 5)	−0.1544	0.1771	0.383
S_{t-6}^{neg} (lag 6)	−0.0581	0.1877	0.757
S_{t-7}^{neg} (lag 7)	−0.4491	0.1812	0.013*
S_{t-8}^{neg} (lag 8)	0.1226	0.1923	0.524

Table 22 (continued)

Variables	Coefficient	Std err	p-value
S^{neg}_{t-9} (lag 9)	0.2799	0.2201	0.203
S^{neg}_{t-10} (lag 10)	0.0705	0.1704	0.679
S^{neg}_{t-11} (lag 11)	0.2061	0.2014	0.306
S^{neg}_{t-12} (lag 12)	−0.0341	0.1908	0.858
S^{neg}_{t-13} (lag 13)	−0.3256	0.1883	0.084
S^{neg}_{t-14} (lag 14)	−0.2336	0.1750	0.182
S^{neg}_{t-15} (lag 15)	0.5940	0.1969	0.003**

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 23 Estimation results for α , β , and γ in the S^{neg}_t equation of P_p , V_p , and S^{neg}_t

Variables	Coefficient	Std err	p-value
Intercept	0.0682	0.0130	0.000***
P_{t-1} (lag 1)	−0.0341	0.0213	0.109
P_{t-2} (lag 2)	−0.0361	0.0207	0.082
P_{t-3} (lag 3)	−0.0427	0.0233	0.067
P_{t-4} (lag 4)	0.0198	0.0243	0.415
P_{t-5} (lag 5)	−0.0105	0.0250	0.675
P_{t-6} (lag 6)	−0.0228	0.0214	0.286
P_{t-7} (lag 7)	0.0163	0.0210	0.438
P_{t-8} (lag 8)	−0.0332	0.0213	0.120
P_{t-9} (lag 9)	−0.0025	0.0257	0.922
P_{t-10} (lag 10)	−0.0029	0.0221	0.896
P_{t-11} (lag 11)	−0.0224	0.0230	0.330
P_{t-12} (lag 12)	−0.0097	0.0257	0.705
P_{t-13} (lag 13)	0.0491	0.0211	0.020*
P_{t-14} (lag 14)	−0.0320	0.0286	0.263
P_{t-15} (lag 15)	−0.0035	0.0247	0.887
V_{t-1} (lag 1)	0.0064	0.0047	0.174
V_{t-2} (lag 2)	−0.0024	0.0056	0.674
V_{t-3} (lag 3)	0.0111	0.0051	0.028*
V_{t-4} (lag 4)	0.0037	0.0056	0.505
V_{t-5} (lag 5)	0.0128	0.0052	0.013*
V_{t-6} (lag 6)	0.0116	0.0058	0.048*
V_{t-7} (lag 7)	−0.0013	0.0054	0.806
V_{t-8} (lag 8)	−0.0001	0.0052	0.988
V_{t-9} (lag 9)	0.0015	0.0053	0.772
V_{t-10} (lag 10)	−0.0012	0.0060	0.841
V_{t-11} (lag 11)	0.0011	0.0056	0.838
V_{t-12} (lag 12)	−0.0023	0.0060	0.698
V_{t-13} (lag 13)	0.0059	0.0062	0.344
V_{t-14} (lag 14)	−0.0084	0.0061	0.167
V_{t-15} (lag 15)	−0.0139	0.0045	0.002**
S^{neg}_{t-1} (lag 1)	0.2297	0.0308	0.000***

Table 23 (continued)

Variables	Coefficient	Std err	p-value
S^{neg}_{t-2} (lag 2)	0.0694	0.0348	0.046*
S^{neg}_{t-3} (lag 3)	-0.0014	0.0341	0.967
S^{neg}_{t-4} (lag 4)	0.0521	0.0340	0.125
S^{neg}_{t-5} (lag 5)	0.0471	0.0292	0.106
S^{neg}_{t-6} (lag 6)	-0.0317	0.0303	0.296
S^{neg}_{t-7} (lag 7)	0.1214	0.0347	0.000***
S^{neg}_{t-8} (lag 8)	0.0233	0.0328	0.477
S^{neg}_{t-9} (lag 9)	0.0537	0.0313	0.086
S^{neg}_{t-10} (lag 10)	0.0085	0.0343	0.805
S^{neg}_{t-11} (lag 11)	-0.0099	0.0284	0.728
S^{neg}_{t-12} (lag 12)	0.0190	0.0318	0.551
S^{neg}_{t-13} (lag 13)	0.0436	0.0299	0.145
S^{neg}_{t-14} (lag 14)	0.0086	0.0346	0.803
S^{neg}_{t-15} (lag 15)	0.0490	0.0361	0.175

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 24 Augmented Dickey-Fuller test results (P_t , V_t , and S^{val}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{val}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 25 AIC (P_t , V_t , and S^{val}_t)

lag	AIC
0	-17.25
1	-17.32
2	-17.39
3	-17.41
4	-17.41
5	-17.46
6	-17.48
7	-17.49
8	-17.49
9	-17.48
10	-17.47
11	-17.46
12	-17.48
13	-17.50
14	-17.50
15	-17.49
16	-17.49
17	-17.48
18	-17.47
19	-17.46
20	-17.45
21	-17.45
22	-17.44
23	-17.43
24	-17.42

Table 26 Estimation results for α , β , and γ in the P_t equation of P_t , V_t , and S^{val}_t

Variables	Coefficient	Std err	p-value
Intercept	-0.0187	0.0145	0.197
P_{t-1} (lag 1)	-0.0639	0.0342	0.061
P_{t-2} (lag 2)	0.0389	0.0308	0.207
P_{t-3} (lag 3)	0.0017	0.0316	0.956
P_{t-4} (lag 4)	0.0802	0.0514	0.119
P_{t-5} (lag 5)	0.0103	0.0242	0.671
P_{t-6} (lag 6)	0.0191	0.0282	0.498
P_{t-7} (lag 7)	-0.0590	0.0515	0.252
P_{t-8} (lag 8)	-0.0381	0.0284	0.180
P_{t-9} (lag 9)	0.0357	0.0271	0.187
P_{t-10} (lag 10)	0.0179	0.0285	0.531
P_{t-11} (lag 11)	0.0066	0.0356	0.854
P_{t-12} (lag 12)	-0.0195	0.0315	0.535
P_{t-13} (lag 13)	0.0155	0.0306	0.614
P_{t-14} (lag 14)	-0.0054	0.0348	0.876
V_{t-1} (lag 1)	0.0045	0.0057	0.435
V_{t-2} (lag 2)	0.0063	0.0057	0.267
V_{t-3} (lag 3)	-0.0014	0.0061	0.816
V_{t-4} (lag 4)	0.0002	0.0058	0.971
V_{t-5} (lag 5)	0.0072	0.0054	0.188
V_{t-6} (lag 6)	-0.0008	0.0064	0.902
V_{t-7} (lag 7)	0.0074	0.0061	0.224
V_{t-8} (lag 8)	-0.0020	0.0062	0.752
V_{t-9} (lag 9)	-0.0053	0.0061	0.388
V_{t-10} (lag 10)	-0.0066	0.0056	0.243
V_{t-11} (lag 11)	0.0033	0.0066	0.619
V_{t-12} (lag 12)	0.0037	0.0068	0.583
V_{t-13} (lag 13)	0.0064	0.0059	0.276
V_{t-14} (lag 14)	0.0059	0.0064	0.357
S^{val}_{t-1} (lag 1)	0.0821	0.0583	0.159
S^{val}_{t-2} (lag 2)	0.0664	0.0481	0.167
S^{val}_{t-3} (lag 3)	-0.0026	0.0515	0.960
S^{val}_{t-4} (lag 4)	0.0078	0.0433	0.857
S^{val}_{t-5} (lag 5)	-0.0060	0.0608	0.921
S^{val}_{t-6} (lag 6)	-0.0116	0.0556	0.834
S^{val}_{t-7} (lag 7)	0.0438	0.0511	0.392
S^{val}_{t-8} (lag 8)	0.0517	0.0503	0.304
S^{val}_{t-9} (lag 9)	-0.2280	0.1838	0.215
S^{val}_{t-10} (lag 10)	-0.0039	0.0675	0.954
S^{val}_{t-11} (lag 11)	-0.0657	0.0418	0.115
S^{val}_{t-12} (lag 12)	0.0787	0.0545	0.148
S^{val}_{t-13} (lag 13)	0.0835	0.0570	0.143
S^{val}_{t-14} (lag 14)	0.0979	0.0505	0.053

Table 27 Estimation results for α , β , and γ in the V_t equation of P_t , V_t , and S^{val}_t

Variables	Coefficient	Std err	p-value
Intercept	−0.1008	0.0605	0.096
P_{t-1} (lag 1)	0.0823	0.1873	0.660
P_{t-2} (lag 2)	0.2694	0.1754	0.125
P_{t-3} (lag 3)	−0.0316	0.1541	0.838
P_{t-4} (lag 4)	0.0181	0.1253	0.885
P_{t-5} (lag 5)	0.0758	0.1401	0.588
P_{t-6} (lag 6)	0.2094	0.1254	0.095
P_{t-7} (lag 7)	0.1042	0.1425	0.465
P_{t-8} (lag 8)	0.2747	0.1680	0.102
P_{t-9} (lag 9)	0.2811	0.1282	0.028*
P_{t-10} (lag 10)	0.0375	0.1479	0.800
P_{t-11} (lag 11)	0.0869	0.1186	0.463
P_{t-12} (lag 12)	0.0586	0.1544	0.704
P_{t-13} (lag 13)	0.0788	0.1447	0.586
P_{t-14} (lag 14)	0.2720	0.1401	0.052
V_{t-1} (lag 1)	−0.4249	0.0356	0.000***
V_{t-2} (lag 2)	−0.3894	0.0342	0.000***
V_{t-3} (lag 3)	−0.2830	0.0317	0.000***
V_{t-4} (lag 4)	−0.1989	0.0386	0.000***
V_{t-5} (lag 5)	−0.2172	0.0323	0.000***
V_{t-6} (lag 6)	−0.1229	0.0348	0.000***
V_{t-7} (lag 7)	0.0258	0.0368	0.483
V_{t-8} (lag 8)	−0.0600	0.0372	0.107
V_{t-9} (lag 9)	−0.1201	0.0332	0.000***
V_{t-10} (lag 10)	−0.1377	0.0350	0.000***
V_{t-11} (lag 11)	−0.1674	0.0302	0.000***
V_{t-12} (lag 12)	−0.2006	0.0281	0.000***
V_{t-13} (lag 13)	−0.1521	0.0373	0.000***
V_{t-14} (lag 14)	0.0808	0.0332	0.015*
S^{val}_{t-1} (lag 1)	−0.2170	0.2919	0.457
S^{val}_{t-2} (lag 2)	0.2858	0.2526	0.258
S^{val}_{t-3} (lag 3)	0.3874	0.2846	0.173
S^{val}_{t-4} (lag 4)	0.0289	0.2501	0.908
S^{val}_{t-5} (lag 5)	0.2264	0.2537	0.372
S^{val}_{t-6} (lag 6)	0.1075	0.3335	0.747
S^{val}_{t-7} (lag 7)	−0.2933	0.2800	0.295
S^{val}_{t-8} (lag 8)	−0.1032	0.2598	0.691
S^{val}_{t-9} (lag 9)	0.2456	0.2700	0.363
S^{val}_{t-10} (lag 10)	0.1162	0.3127	0.710
S^{val}_{t-11} (lag 11)	0.0222	0.2381	0.926
S^{val}_{t-12} (lag 12)	0.0637	0.2515	0.800
S^{val}_{t-13} (lag 13)	0.1736	0.3418	0.612
S^{val}_{t-14} (lag 14)	−0.0385	0.3132	0.902

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 28 Estimation results for α , β , and γ in the S^{val}_t equation of P_t , V_t , and S^{val}_t

Variables	Coefficient	Std err	p-value
Intercept	0.0421	0.0065	0.000***
P_{t-1} (lag 1)	0.0049	0.0173	0.777
P_{t-2} (lag 2)	0.0485	0.0178	0.006**
P_{t-3} (lag 3)	0.0021	0.0173	0.904
P_{t-4} (lag 4)	- 0.0205	0.0168	0.221
P_{t-5} (lag 5)	0.0182	0.0146	0.211
P_{t-6} (lag 6)	0.0025	0.0171	0.884
P_{t-7} (lag 7)	- 0.0086	0.0139	0.534
P_{t-8} (lag 8)	- 0.0079	0.0159	0.619
P_{t-9} (lag 9)	0.0290	0.0194	0.134
P_{t-10} (lag 10)	- 0.0285	0.0205	0.164
P_{t-11} (lag 11)	0.0038	0.0159	0.810
P_{t-12} (lag 12)	0.0307	0.0170	0.071
P_{t-13} (lag 13)	- 0.0082	0.0145	0.569
P_{t-14} (lag 14)	0.0147	0.0153	0.336
V_{t-1} (lag 1)	- 0.0031	0.0032	0.330
V_{t-2} (lag 2)	0.0012	0.0031	0.709
V_{t-3} (lag 3)	- 0.0013	0.0036	0.715
V_{t-4} (lag 4)	- 0.0015	0.0031	0.614
V_{t-5} (lag 5)	- 0.0032	0.0031	0.294
V_{t-6} (lag 6)	- 0.0028	0.0036	0.444
V_{t-7} (lag 7)	- 0.0033	0.0034	0.329
V_{t-8} (lag 8)	0.0006	0.0042	0.886
V_{t-9} (lag 9)	0.0003	0.0045	0.947
V_{t-10} (lag 10)	- 0.0012	0.0037	0.737
V_{t-11} (lag 11)	- 0.0006	0.0032	0.839
V_{t-12} (lag 12)	0.0039	0.0034	0.253
V_{t-13} (lag 13)	- 0.0047	0.0029	0.101
V_{t-14} (lag 14)	- 0.0070	0.0031	0.022*
S^{val}_{t-1} (lag 1)	0.1325	0.0349	0.000***
S^{val}_{t-2} (lag 2)	0.1119	0.0289	0.000***
S^{val}_{t-3} (lag 3)	- 0.0115	0.0269	0.670
S^{val}_{t-4} (lag 4)	0.0118	0.0242	0.625
S^{val}_{t-5} (lag 5)	0.1105	0.0259	0.000***
S^{val}_{t-6} (lag 6)	0.0064	0.0280	0.819
S^{val}_{t-7} (lag 7)	0.0967	0.0342	0.005**
S^{val}_{t-8} (lag 8)	0.0398	0.0314	0.205
S^{val}_{t-9} (lag 9)	0.0349	0.0315	0.268
S^{val}_{t-10} (lag 10)	0.0188	0.0276	0.496
S^{val}_{t-11} (lag 11)	0.0004	0.0282	0.989
S^{val}_{t-12} (lag 12)	- 0.0017	0.0319	0.958
S^{val}_{t-13} (lag 13)	0.0263	0.0297	0.375
S^{val}_{t-14} (lag 14)	- 0.0033	0.0293	0.911

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 29 Augmented Dickey-Fuller test results (P_t , V_t , and S^{aro}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{aro}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 30 AIC (P_t , V_t , and S^{aro}_t)

lag	AIC
0	-17.31
1	-17.37
2	-17.45
3	-17.46
4	-17.47
5	-17.51
6	-17.52
7	-17.53
8	-17.52
9	-17.51
10	-17.50
11	-17.49
12	-17.50
13	-17.52
14	-17.52
15	-17.51
16	-17.51
17	-17.50
18	-17.49
19	-17.48
20	-17.47
21	-17.47
22	-17.46
23	-17.46
24	-17.45

Table 31 Estimation results for α , β , and γ in the P_t equation of P_t , V_t , and S^{aro}_t

Variables	Coefficient	Std err	p-value
Intercept	-0.0235	0.0136	0.084
P_{t-1} (lag 1)	-0.0568	0.0371	0.126
P_{t-2} (lag 2)	0.0376	0.0321	0.241
P_{t-3} (lag 3)	-0.0031	0.0329	0.926
P_{t-4} (lag 4)	0.0711	0.0433	0.101
P_{t-5} (lag 5)	0.0062	0.0266	0.817
P_{t-6} (lag 6)	0.0248	0.0267	0.353
P_{t-7} (lag 7)	-0.0506	0.0481	0.292
V_{t-1} (lag 1)	0.0027	0.0053	0.612
V_{t-2} (lag 2)	0.0045	0.0055	0.419
V_{t-3} (lag 3)	-0.0026	0.0064	0.690
V_{t-4} (lag 4)	0.0016	0.0056	0.772
V_{t-5} (lag 5)	0.0091	0.0055	0.100
V_{t-6} (lag 6)	-0.0004	0.0056	0.938
V_{t-7} (lag 7)	0.0092	0.0054	0.088
S^{aro}_{t-1} (lag 1)	0.0665	0.0659	0.313
S^{aro}_{t-2} (lag 2)	-0.0247	0.0555	0.656
S^{aro}_{t-3} (lag 3)	-0.0524	0.0535	0.327
S^{aro}_{t-4} (lag 4)	0.0712	0.0806	0.377
S^{aro}_{t-5} (lag 5)	0.1275	0.0736	0.083
S^{aro}_{t-6} (lag 6)	0.0417	0.0591	0.481
S^{aro}_{t-7} (lag 7)	0.0058	0.0518	0.911

Table 32 Estimation results for α , β , and γ in the V_t equation of P_t , V_t , and S^{aro}_t

Variables	Coefficient	Std err	p-value
Intercept	0.0311	0.0680	0.648
P_{t-1} (lag 1)	0.1048	0.2101	0.618
P_{t-2} (lag 2)	0.3110	0.1828	0.089
P_{t-3} (lag 3)	0.0404	0.1509	0.789
P_{t-4} (lag 4)	0.0464	0.1388	0.738
P_{t-5} (lag 5)	0.0635	0.1397	0.650
P_{t-6} (lag 6)	0.1752	0.1265	0.166
P_{t-7} (lag 7)	0.1340	0.1555	0.389
V_{t-1} (lag 1)	-0.3820	0.0336	0.000***
V_{t-2} (lag 2)	-0.3814	0.0323	0.000***
V_{t-3} (lag 3)	-0.2731	0.0349	0.000***
V_{t-4} (lag 4)	-0.1980	0.0363	0.000***
V_{t-5} (lag 5)	-0.2306	0.0327	0.000***
V_{t-6} (lag 6)	-0.1044	0.0321	0.001**
V_{t-7} (lag 7)	0.1172	0.0329	0.000***
S^{aro}_{t-1} (lag 1)	-0.2246	0.3032	0.459
S^{aro}_{t-2} (lag 2)	0.0397	0.2889	0.891
S^{aro}_{t-3} (lag 3)	-0.1775	0.3047	0.560
S^{aro}_{t-4} (lag 4)	0.2464	0.3104	0.427
S^{aro}_{t-5} (lag 5)	-0.1858	0.3407	0.585
S^{aro}_{t-6} (lag 6)	0.6099	0.4296	0.156
S^{aro}_{t-7} (lag 7)	-0.6188	0.3221	0.055

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 33 Estimation results for α , β , and γ in the S^{aro}_t equation of P_t , V_t , and S^{aro}_t

Variables	Coefficient	Std err	p-value
Intercept	0.0772	0.0066	0.000***
P_{t-1} (lag 1)	- 0.0125	0.0131	0.343
P_{t-2} (lag 2)	- 0.0033	0.0134	0.804
P_{t-3} (lag 3)	0.0090	0.0148	0.542
P_{t-4} (lag 4)	0.0066	0.0149	0.658
P_{t-5} (lag 5)	0.0071	0.0145	0.626
P_{t-6} (lag 6)	- 0.0113	0.0156	0.468
P_{t-7} (lag 7)	- 0.0025	0.0146	0.864
V_{t-1} (lag 1)	0.0022	0.0027	0.420
V_{t-2} (lag 2)	0.0016	0.0029	0.591
V_{t-3} (lag 3)	0.0022	0.0030	0.457
V_{t-4} (lag 4)	0.0030	0.0027	0.256
V_{t-5} (lag 5)	- 0.0009	0.0030	0.757
V_{t-6} (lag 6)	- 0.0017	0.0029	0.563
V_{t-7} (lag 7)	- 0.0034	0.0026	0.193
S^{aro}_{t-1} (lag 1)	0.1281	0.0304	0.000***
S^{aro}_{t-2} (lag 2)	0.0940	0.0308	0.002**
S^{aro}_{t-3} (lag 3)	- 0.0380	0.0318	0.231
S^{aro}_{t-4} (lag 4)	0.0487	0.0289	0.092
S^{aro}_{t-5} (lag 5)	0.0233	0.0315	0.461
S^{aro}_{t-6} (lag 6)	- 0.0138	0.0297	0.643
S^{aro}_{t-7} (lag 7)	- 0.0007	0.0332	0.982

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 34 Augmented Dickey-Fuller test results (P_t , V_t , and S^{dom}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{dom}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 35 AIC (P_t , V_t , and S^{dom}_t)

lag	AIC
0	-17.25
1	-17.32
2	-17.39
3	-17.41
4	-17.41
5	-17.46
6	-17.48
7	-17.49
8	-17.49
9	-17.48
10	-17.47
11	-17.46
12	-17.48
13	-17.50
14	-17.50
15	-17.49
16	-17.49
17	-17.48
18	-17.47
19	-17.46
20	-17.45
21	-17.45
22	-17.44
23	-17.43
24	-17.42

Table 36 Estimation results for α , β , and γ in the P_t equation of P_t , V_t , and S^{dom}_t

Variables	Coefficient	Std err	p-value
Intercept	- 0.0069	0.0159	0.665
P_{t-1} (lag 1)	- 0.0566	0.0340	0.096
P_{t-2} (lag 2)	0.0497	0.0301	0.099
P_{t-3} (lag 3)	0.0010	0.0299	0.974
P_{t-4} (lag 4)	0.0696	0.0435	0.110
P_{t-5} (lag 5)	0.0086	0.0252	0.733
P_{t-6} (lag 6)	0.0298	0.0288	0.300
P_{t-7} (lag 7)	- 0.0570	0.0525	0.277
P_{t-8} (lag 8)	- 0.0518	0.0284	0.068
P_{t-9} (lag 9)	0.0316	0.0260	0.223
P_{t-10} (lag 10)	0.0283	0.0269	0.294
P_{t-11} (lag 11)	0.0000	0.0331	1.000
P_{t-12} (lag 12)	- 0.0226	0.0316	0.474
P_{t-13} (lag 13)	0.0244	0.0299	0.415
P_{t-14} (lag 14)	0.0025	0.0346	0.941
V_{t-1} (lag 1)	0.0053	0.0056	0.342
V_{t-2} (lag 2)	0.0077	0.0058	0.183
V_{t-3} (lag 3)	0.0000	0.0062	0.997
V_{t-4} (lag 4)	0.0018	0.0056	0.751
V_{t-5} (lag 5)	0.0087	0.0054	0.111
V_{t-6} (lag 6)	- 0.0021	0.0061	0.737
V_{t-7} (lag 7)	0.0059	0.0062	0.339
V_{t-8} (lag 8)	- 0.0026	0.0063	0.680
V_{t-9} (lag 9)	- 0.0041	0.0061	0.506
V_{t-10} (lag 10)	- 0.0042	0.0058	0.473
V_{t-11} (lag 11)	0.0053	0.0065	0.415
V_{t-12} (lag 12)	0.0058	0.0070	0.403
V_{t-13} (lag 13)	0.0073	0.0060	0.223
V_{t-14} (lag 14)	0.0060	0.0064	0.349
S^{dom}_{t-1} (lag 1)	0.1381	0.0818	0.091
S^{dom}_{t-2} (lag 2)	- 0.0443	0.0588	0.451
S^{dom}_{t-3} (lag 3)	- 0.0627	0.0637	0.325
S^{dom}_{t-4} (lag 4)	- 0.0865	0.0578	0.134
S^{dom}_{t-5} (lag 5)	0.0174	0.0543	0.749
S^{dom}_{t-6} (lag 6)	0.1449	0.0521	0.005**
S^{dom}_{t-7} (lag 7)	- 0.0880	0.0589	0.135
S^{dom}_{t-8} (lag 8)	0.0174	0.0510	0.732
S^{dom}_{t-9} (lag 9)	- 0.0141	0.0503	0.779
S^{dom}_{t-10} (lag 10)	0.0510	0.0778	0.512
S^{dom}_{t-11} (lag 11)	- 0.1398	0.0857	0.103
S^{dom}_{t-12} (lag 12)	0.0474	0.0700	0.499
S^{dom}_{t-13} (lag 13)	0.0484	0.0668	0.469
S^{dom}_{t-14} (lag 14)	0.0411	0.0589	0.486

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 37 Estimation results for α , β , and γ in the V_t equation of P_t , V_t , and S^{dom}_t

Variables	Coefficient	Std err	p-value
Intercept	- 0.0088	0.0692	0.898
P_{t-1} (lag 1)	0.0713	0.1954	0.715
P_{t-2} (lag 2)	0.2838	0.1678	0.091
P_{t-3} (lag 3)	- 0.0142	0.1535	0.926
P_{t-4} (lag 4)	0.0757	0.1247	0.544
P_{t-5} (lag 5)	0.0841	0.1380	0.542
P_{t-6} (lag 6)	0.1897	0.1284	0.140
P_{t-7} (lag 7)	0.1185	0.1429	0.407
P_{t-8} (lag 8)	0.2883	0.1672	0.085
P_{t-9} (lag 9)	0.2768	0.1306	0.034*
P_{t-10} (lag 10)	0.0454	0.1493	0.761
P_{t-11} (lag 11)	0.1483	0.1209	0.220
P_{t-12} (lag 12)	0.0970	0.1528	0.526
P_{t-13} (lag 13)	0.0495	0.1414	0.726
P_{t-14} (lag 14)	0.2905	0.1418	0.040*
V_{t-1} (lag 1)	- 0.4193	0.0341	0.000***
V_{t-2} (lag 2)	- 0.3880	0.0347	0.000***
V_{t-3} (lag 3)	- 0.2820	0.0324	0.000***
V_{t-4} (lag 4)	- 0.1970	0.0393	0.000***
V_{t-5} (lag 5)	- 0.2165	0.0320	0.000***
V_{t-6} (lag 6)	- 0.1179	0.0349	0.001***
V_{t-7} (lag 7)	0.0316	0.0364	0.386
V_{t-8} (lag 8)	- 0.0570	0.0378	0.132
V_{t-9} (lag 9)	- 0.1176	0.0327	0.000***
V_{t-10} (lag 10)	- 0.1371	0.0348	0.000***
V_{t-11} (lag 11)	- 0.1667	0.0306	0.000***
V_{t-12} (lag 12)	- 0.2006	0.0304	0.000***
V_{t-13} (lag 13)	- 0.1530	0.0388	0.000***
V_{t-14} (lag 14)	0.0841	0.0321	0.009**
S^{dom}_{t-1} (lag 1)	- 0.2794	0.3238	0.388
S^{dom}_{t-2} (lag 2)	- 0.0167	0.2751	0.952
S^{dom}_{t-3} (lag 3)	0.1637	0.3675	0.656
S^{dom}_{t-4} (lag 4)	0.0971	0.3032	0.749
S^{dom}_{t-5} (lag 5)	- 0.3391	0.2565	0.186
S^{dom}_{t-6} (lag 6)	0.3889	0.2977	0.192
S^{dom}_{t-7} (lag 7)	0.0351	0.3474	0.919
S^{dom}_{t-8} (lag 8)	- 0.0222	0.3061	0.942
S^{dom}_{t-9} (lag 9)	- 0.1432	0.2607	0.583
S^{dom}_{t-10} (lag 10)	- 0.3006	0.3399	0.376
S^{dom}_{t-11} (lag 11)	0.2383	0.2852	0.404
S^{dom}_{t-12} (lag 12)	- 0.3256	0.3070	0.289
S^{dom}_{t-13} (lag 13)	0.2168	0.4748	0.648
S^{dom}_{t-14} (lag 14)	0.3592	0.3012	0.233

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Table 38 Estimation results for α , β , and γ in the S^{dom}_t equation of P_t , V_t , and S^{dom}_t

Variables	Coefficient	Std err	p-value
Intercept	0.0533	0.0086	0.000***
P_{t-1} (lag 1)	- 0.0054	0.0126	0.669
P_{t-2} (lag 2)	0.0248	0.0149	0.095
P_{t-3} (lag 3)	0.0364	0.0123	0.003**
P_{t-4} (lag 4)	- 0.0060	0.0146	0.679
P_{t-5} (lag 5)	- 0.0023	0.0141	0.872
P_{t-6} (lag 6)	- 0.0058	0.0131	0.660
P_{t-7} (lag 7)	0.0146	0.0126	0.247
P_{t-8} (lag 8)	0.0085	0.0153	0.578
P_{t-9} (lag 9)	0.0100	0.0140	0.472
P_{t-10} (lag 10)	0.0153	0.0133	0.250
P_{t-11} (lag 11)	0.0321	0.0125	0.010*
P_{t-12} (lag 12)	0.0147	0.0142	0.298
P_{t-13} (lag 13)	- 0.0020	0.0129	0.879
P_{t-14} (lag 14)	0.0084	0.0129	0.514
V_{t-1} (lag 1)	- 0.0053	0.0027	0.049*
V_{t-2} (lag 2)	- 0.0052	0.0030	0.081
V_{t-3} (lag 3)	- 0.0078	0.0033	0.019*
V_{t-4} (lag 4)	- 0.00425	0.0027	0.103
V_{t-5} (lag 5)	- 0.0050	0.0035	0.149
V_{t-6} (lag 6)	- 0.0049	0.0036	0.173
V_{t-7} (lag 7)	- 0.0022	0.0033	0.501
V_{t-8} (lag 8)	- 0.0001	0.0034	0.970
V_{t-9} (lag 9)	- 0.0009	0.0036	0.801
V_{t-10} (lag 10)	- 0.0016	0.0034	0.627
V_{t-11} (lag 11)	- 0.0064	0.0033	0.053
V_{t-12} (lag 12)	- 0.0039	0.0034	0.255
V_{t-13} (lag 13)	- 0.0045	0.0031	0.144
V_{t-14} (lag 14)	- 0.0032	0.0028	0.249
S^{dom}_{t-1} (lag 1)	0.1344	0.0327	0.000***
S^{dom}_{t-2} (lag 2)	0.0163	0.0303	0.590
S^{dom}_{t-3} (lag 3)	0.0153	0.0370	0.679
S^{dom}_{t-4} (lag 4)	0.0113	0.0317	0.721
S^{dom}_{t-5} (lag 5)	0.0568	0.0272	0.036*
S^{dom}_{t-6} (lag 6)	0.0570	0.0268	0.033*
S^{dom}_{t-7} (lag 7)	0.0554	0.0311	0.075
S^{dom}_{t-8} (lag 8)	0.0448	0.0299	0.134
S^{dom}_{t-9} (lag 9)	- 0.0201	0.0283	0.478
S^{dom}_{t-10} (lag 10)	0.0566	0.0354	0.109
S^{dom}_{t-11} (lag 11)	0.0058	0.0322	0.857
S^{dom}_{t-12} (lag 12)	0.0302	0.0274	0.270
S^{dom}_{t-13} (lag 13)	- 0.0020	0.0298	0.945
S^{dom}_{t-14} (lag 14)	0.0383	0.0284	0.177

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively. These tables use Newey-West standard errors

Appendix 2. Additional tables for the results of the hypothesis tests

See Tables 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53.

Table 39 Ljung-Box test results (P_t , V_t , and S_t^{pos})

Variables	p-value
P_t	1.000
V_t	0.815
S_t^{pos}	0.997

Table 40 ARCH-LM test results (P_t , V_t , and S_t^{pos})

Variables	p-value
P_t	0.650
V_t	0.000***
S_t^{pos}	0.057

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 41 Jarque–Bera test results (P_t , V_t , and S_t^{pos})

Variables	p-value
P_t	0.000***
V_t	0.000***
S_t^{pos}	0.001**

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 42 Ljung-Box test results (P_t , V_t , and S_t^{neg})

Variables	p-value
P_t	1.000
V_t	0.889
S_t^{neg}	1.000

Table 43 ARCH-LM test results (P_t , V_t , and S_t^{neg})

Variables	p-value
P_t	0.650
V_t	0.000***
S_t^{neg}	0.542

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 44 Jarque–Bera test results (P_t , V_t , and S^{neg}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{neg}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 45 Ljung–Box test results (P_t , V_t , and S^{val}_t)

Variables	p-value
P_t	1.000
V_t	0.892
S^{val}_t	1.000

Table 46 ARCH–LM test results (P_t , V_t , and S^{val}_t)

Variables	p-value
P_t	0.162
V_t	0.000***
S^{val}_t	1.000

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 47 Jarque–Bera test results (P_t , V_t , and S^{val}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{val}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 48 Ljung–Box test results (P_t , V_t , and S^{aro}_t)

Variables	p-value
P_t	1.000
V_t	0.373
S^{aro}_t	1.000

Table 49 ARCH–LM test results (P_t , V_t , and S^{aro}_t)

Variables	p-value
P_t	0.268
V_t	0.000***
S^{aro}_t	0.002**

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 50 Jarque–Bera test results (P_t , V_t , and S^{aro}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{aro}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 51 Ljung-Box test results (P_t , V_t , and S^{dom}_t)

Variables	p-value
P_t	1.000
V_t	0.812
S^{dom}_t	1.000

Table 52 ARCH-LM test results (P_t , V_t , and S^{dom}_t)

Variables	p-value
P_t	0.622
V_t	0.000***
S^{dom}_t	0.748

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Table 53 Jarque–Bera test results (P_t , V_t , and S^{dom}_t)

Variables	p-value
P_t	0.000***
V_t	0.000***
S^{dom}_t	0.000***

*, **, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively

Author contributions T.I. wrote the main body of the manuscript. A.Y. wrote part of the manuscript and he prepared all figures and tables. S.M., the supervising professor, was responsible for the supervision and leadership of this study. All authors reviewed the manuscript. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Data availability We are unable to publish the users' post data due to the terms of use for the Academic Research product track of X (formerly Twitter).

Declarations

Conflict of interest The authors declare no competing interests.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Akhtaruzzaman, M., Sensoy, A., & Corbet, S. (2020). The influence of Bitcoin on portfolio diversification and design. *Finance Research Letters*, 37, 101344.
- Al-Yahyaee, K. H., Mensi, W., & Yoon, S. M. (2018). Efficiency, multifractality, and the long-memory property of the Bitcoin market: A comparative analysis with stock, currency, and gold markets. *Finance Research Letters*, 27, 228–234.
- Ante, L. (2023). How Elon Musk's twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change*, 186, 122112.
- Baig, A., Blau, B. M., & Sabah, N. (2019). Price clustering and sentiment in bitcoin. *Finance Research Letters*, 29, 111–116.
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645–1680.
- Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford University Press.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.
- Baur, D. G., & Dimpfl, T. (2021). The volatility of Bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics*, 61(5), 2663–2683.
- 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.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- Bouktif, S., Fiaz, A., & Awad, M. (2020). Augmented textual features-based stock market prediction. *IEEE Access*, 8, 40269–40282.
- Chaim, P., & Laurini, M. P. (2018). Volatility and return jumps in bitcoin. *Economics Letters*, 173, 158–163.
- Chen, Y., Bredin, D., Poti, V., & Matkovskyy, R. (2022). COVID risk narratives: A computational linguistic approach to the econometric identification of narrative risk during a pandemic. *Digital Finance*, 4(1), 17–61.
- 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(366a), 427–431.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50, 987–1007.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37, 424–438.
- Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the International AAAI Conference on web and social media* (Vol. 8, No. 1, pp. 216–225).
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255–259.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292.
- Kim, H. J., Hong, J. S., Hwang, H. C., Kim, S. M., & Han, D. H. (2020). Comparison of psychological status and investment style between bitcoin investors and share investors. *Frontiers in Psychology*, 11, 502295.
- Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65, 101188.
- Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230.

- Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, R. H. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, 35(1), 19–52.
- Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.
- Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In: *Proceedings of the 56th Annual Meeting of the Association for computational linguistics* (volume 1: Long papers) (pp. 174–184).
- Mushtaq, R. (2011). Augmented dickey fuller test. *Econometrics: Mathematical Methods & Programming eJournal*. <https://doi.org/10.2139/ssrn.1911068>.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, 2008, 21260.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, 55(3), 703–708.
- Platanakis, E., & Urquhart, A. (2020). Should investors include bitcoin in their portfolios? A portfolio theory approach. *The British Accounting Review*, 52(4), 100837.
- Sensoy, A. (2019). The inefficiency of Bitcoin revisited: A high-frequency analysis with alternative currencies. *Finance Research Letters*, 28, 68–73.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 48, 1–48.
- 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.
- Valdez, D., Ten Thij, M., Bathina, K., Rutter, L. A., & Bollen, J. (2020). Social media insights into US mental health during the COVID-19 pandemic: Longitudinal analysis of Twitter data. *Journal of Medical Internet Research*, 22(12), e21418.
- Vidal-Tomás, D., & Ibañez, A. (2018). Semi-strong efficiency of Bitcoin. *Finance Research Letters*, 27, 259–265.
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter “I hope it is not as bad as I fear.” *Procedia-Social and Behavioral Sciences*, 26, 55–62.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.