Analysis of the high-frequency volume-price relationship of Bitcoin

Fang Wan[a], Chun-Xiao Nie[a,b,*]

[a] School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China
[b] Collaborative Innovation Center of Statistical Data Engineering, Technology & Application, Zhejiang Gongshang University, Hangzhou, 310018, China

Abstract

Keywords: Information flow, Transfer entropy, Cryptocurrency market, Volume-price relationship

1. Introduction

Research into the volume-price relationship provides insights into the underlying mechanisms of market efficiency and liquidity formation. By quantifying the extent to which volume influences or predicts price, analysts can develop more robust models for risk management, trading strategies, and regulatory assessments.

2. Data and Software

The primary data source used in this study is the Gemini Exchange dataset available from `https://www.cryptodatadownload.com/data/gemini/`, which provides minute-level historical data from 2017 through the most recently available period for Bitcoin trading. This dataset includes timestamped open, high, low, close (OHLC) prices, and trading volume calculated in BTC and USD. The high-frequency nature of the data enables fine-grained temporal analysis and makes it particularly suitable for exploring information flow dynamics within short time intervals. The software implemented in our analysis is R with several well-verified packages mentioned in Section 3. The general view of our data is shown as below. We consider only days with over 90% minute-level data available for the analysis, and the valid days for each year are shown in Table 1. The summary statistics of the high-frequency close price and trading volume are shown in Table 2 and Table 3 respectively. The data is then transformed into two log returns, and are denoted as $r^{(P)}$ and $r^{(V)}$ for price and volume respectively.

Table 1: Valid Days for Each Year in the Dataset

Year	Valid Days
2017	111
2018	94
2019	49
2020	195
2021	359
2022	314
2023	30
2024	140
2025	65

*Corresponding author
Email address: `niechunxiao2009@163.com; niechunxiao@zjgsu.edu.cn` (Chun-Xiao Nie)

Table 2: Summary statistics of the high-frequency Close Price(calculated in USD)

Year	Mean	Std	Skewness	Kurtosis
2017	7323.06	4905.62	0.9181	-0.4115
2018	9681.76	2826.69	0.1143	0.0823
2019	9205.53	1968.58	-0.2179	-0.4777
2020	12762.69	5245.04	0.9788	0.4220
2021	47424.68	9802.02	-0.0576	-1.1365
2022	29678.15	10009.00	0.2697	-1.5421
2023	29687.64	7180.20	0.7205	-0.5403
2024	75431.92	17850.48	0.0627	-1.1986
2025	94123.43	7461.64	-0.2667	-1.0735

Table 3: Summary statistics of the high-frequency Trading Volume(calculated in BTC)

Year	Mean	Std	Skewness	Kurtosis
2017	8.96	28.97	34.32	2278.52
2018	6.10	13.52	10.60	259.67
2019	3.18	12.65	13.87	340.38
2020	2.04	7.71	16.00	463.17
2021	1.52	4.95	26.14	1770.81
2022	1.06	4.48	25.13	1253.53
2023	0.58	2.17	25.10	1381.38
2024	0.84	1.85	8.88	194.07
2025	0.78	1.64	7.94	143.46

It is obivious that the price grows rapidly from 2017 to 2025, with the its distribution much level than normal, and slightly positive skewness. From the perspective of volume, the distribution is hight skewed and heavy-tailed, with the mean volume decreasing from 2017 to 2025. This indicates that in this fast, extensive market, the growth of trading scale of Bitcoin is imbalance with the increase of its price, which implies a non-linear relationship can be captured.

3. Method

3.1. Transfer entropy(TE) with Markov Block Bootstrap

Transfer entropy is used to capture the non-linear and directional information flow between time series, with that from $X = \{x_t\}$ to $Y = \{y_t\}$ defined as Eq. 1 Schreiber (2000).

$$TE_{y\to x} = \sum p(x_{t+1}, \mathbf{x}_t^{(k)}, \mathbf{y}_t^{(l)}) \log \frac{p(x_{t+1} \mid \mathbf{x}_t^{(k)}, \mathbf{y}_t^{(l)})}{p(x_{t+1} \mid \mathbf{x}_t^{(k)})} \tag{1}$$

where $\mathbf{x}_t^{(k)} = (x_{t-k+1}, \ldots, x_t)$ and $\mathbf{y}_t^{(l)} = (y_{t-l+1}, \ldots, y_t)$ denote the k-lag and l-lag historical state vectors of x and y respectively, with the number of states is determined as 4 by quantiles as Eq. 2. Besides, a 600-time Markov Block Bootstrap is adopted to preserve the local dynamics between the series while eliminating the global dependency, the significance of the true TE is obtained from the surrogate distribution of the generated TEs, where the smaller the p-value, the more significant the information flow.

$$x_t^{'} = \begin{cases} s_1 & x_t \le q_{0.1} \\ s_2 & x_t \in (q_{0.1}, q_{0.5}] \\ s_3 & x_t \in (q_{0.5}, q_{0.9}] \\ s_4 & x_t > q_{0.9} \end{cases} \tag{2}$$

3.2. VAR-based TE analysis

Given that VAR models are widely used to capture linear interdependencies, which can be expressed as Eq. 3:

$$\begin{bmatrix} r_t^{(P)} \\ r_t^{(V)} \end{bmatrix} = c + \sum_{i=1}^{3} A_i \begin{bmatrix} r_{t-i}^{(P)} \\ r_{t-i}^{(V)} \end{bmatrix} + \epsilon_t \tag{3}$$

Here, A_i represents the coefficient matrices capturing linear dependencies at lag i , and ϵ_t is a 2-dimensional residual vector, isolating the nonlinear parts of the $r^{(P)}$ and $r^{(V)}$ that cannot be explained by VAR. With the significance evaluated by the combination of JB test, Ljung-Box test and ARCH LM test, where smaller p-values for the JB and LM tests indicate better results, and larger p-values for the LB test indicate better results.

To analyze nonlinear components individually in the volume-price relationship, we design a implementation based on transfer entropy (TE) and vector autoregression (VAR):

1. Firstly, we compute the transfer entropy (TE) bidirectionally between $r^{(P)}$ and $r^{(V)}$, with the result denoted as $TE_{p \to v}$ and $TE_{v \to p}$. Based on these TE results, we put analysis on the directional dominance and higher moments of TE.

2. Secondly, we fit a VAR with order up to 3 between $r^{(P)}$ and $r^{(V)}$, and evaluating the significance of VAR with a combination of JB test(Eq. 4), Ljung-Box test(Eq. 5) and ARCH LM test(Eq. 6) on each residual respectively, and a primal idea towards the different components in the price-volume relationship will be obtained from the result.

$$JB = \frac{n}{6} \left(S^2 + \frac{(K-3)^2}{4} \right) \tag{4}$$

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k} \tag{5}$$

$$LM = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k} \tag{6}$$

   In these equations, S, K, n, h, and $\hat{\rho}_k$ denote skewness, kurtosis, length of the residual observed, lag order, and autocorrelation at lag k, respectively. The significance of the residuals is assessed via the Jarque-Bera (JB) test for normality and the Ljung-Box (Q) test for autocorrelation. The corresponding p-values are computed under the null hypotheses of normality (χ_2^2 distribution) and white noise (χ_h^2 distribution), respectively.

3. We then recompute bidirectional TE on the residuals to measure the information flow of the nonlinear components.By comparing TE before and after filtering, we can model the relative strength of linear and nonlinear causality and distinguish their respective roles in market dynamics.

3.3. Mann-Kendall Trend

To further evaluate trend inside the TE calculated, we applied the Mann-Kendall (MK) test to the daily TE series, The definition of the MK test contains two parts as Eq. 6 and Eq. 7, where the satistics Z itself represent both the direction and the significance of the trend, for Z rising to 1 represent a significant upward trend, and falling to -1 represent a significant downward trend.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sign}(x_j - x_i) \tag{7}$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{\mathrm{Var}(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{\mathrm{Var}(S)}} & \text{if } S < 0 \end{cases} \tag{8}$$

4. Results

4.1. Volume-price relationship of high-frequency data

Based on the transfer entropy (TE) calculated for each day, we compute Annual Summaries of TE for both direction, which are offered in Tab. 4 and Tab. 5. hype skewness and kurtosis are observed in 2017, 2021, 2022 and 2025, showing that the price-volume relation tend to get more significant when the market is in drastic volatility. By comparing the $\bar{p}$ of TE betwen the two direction, the one from price to volume is much lower, which means the information flow from price to volume is more significant, indicating the market is generally driven more by price than volume, however, On the TE value it self, We found the average of $TE_{v\to p}$ is slightly higher than $TE_{p\to v}$, which indicates that though the information flow from volume to price is less significant, a stable part of non-linear information flow from volume to price is coverd by the the dominant bidirectional linear one.(Details in Sec. 4.2).

Table 4: Moments & Significance of TE from Price to Volume ($TE_{P\to V}$)

Year	**Mean**	**Std**	**Skewness**	**Kurtosis**	$\bar{p}$
2017	0.0226	0.0052	0.2346	0.0073	0.1234
2018	0.0238	0.0111	**5.0787**	**32.0705**	0.1284
2019	0.0229	0.0056	0.9319	0.7630	0.2129
2020	0.0220	0.0054	0.5658	0.1045	0.2020
2021	0.0212	0.0251	**17.8038**	**330.1464**	0.2575
2022	0.0238	0.0267	**11.8465**	**145.9301**	0.2354
2023	0.1545	0.2303	1.2866	-0.1367	0.2873
2024	0.0920	0.1670	2.1809	3.2453	0.2624
2025	0.0373	0.0785	**5.7141**	**32.5258**	0.2614

Table 5: Moments & Significance of TE from Volume to Price ($TE_{V\to P}$)

Year	**Mean**	**Std**	**Skewness**	**Kurtosis**	$\bar{p}$
2017	0.0284	0.0066	0.6172	-0.0201	0.3550
2018	0.0299	0.0154	**6.8460**	**57.5926**	0.3710
2019	0.0257	0.0061	0.0381	-0.2896	0.3893
2020	0.0257	0.0066	0.4806	-0.0247	0.4033
2021	0.0243	0.0188	**16.3237**	**293.5880**	0.4673
2022	0.0259	0.0174	**10.6488**	**134.4553**	0.3987
2023	0.1433	0.2221	1.7369	2.1892	0.3688
2024	0.0911	0.1634	2.3158	4.2528	0.3461
2025	0.0397	0.0888	**5.5911**	**30.6203**	0.3084

Consequently, we did another comparison over the TE value pairly in a day, with the results shown in fig 1. We found that from a synchronized perspective, the ininformation flow from price to volume is still dominant the price-volume relationship. But the dominance is showing a noticable decreasing trend in recent years, showing a paradiam shift we'd describe in Sec. 4.2.

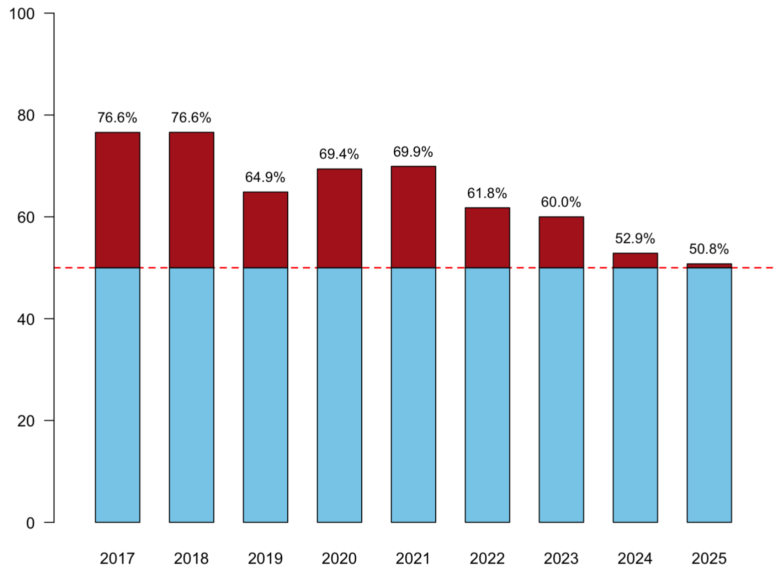

Figure 1: Ratio of Days that $TE_{p \to v} > TE_{v \to p}$ across years.

Besides, we observed specially higher upper quantiles of TE in 2023 and 2024(as shown in Tab. 6), which might be caused by certain market maniputation like The Merge of ETH, and cause the market to form a hyper connection between price and volume.

Table 6: 95% and 99% Quantiles of TE from Both Directions

Year	$Q_{0.95}^{P \to V}$	$Q_{0.99}^{P \to V}$	$Q_{0.95}^{V \to P}$	$Q_{0.99}^{V \to P}$
2017	0.0317	0.0339	0.0404	0.0457
2018	0.0315	0.0798	0.0435	0.0632
2019	0.0339	0.0369	0.0348	0.0388
2020	0.0318	0.0358	0.0374	0.0419
2021	0.0286	0.0345	0.0340	0.0411
2022	0.0314	0.0379	0.0358	0.0465
2023	**0.5931**	**0.6572**	**0.5638**	**0.7595**
2024	**0.5195**	**0.5941**	**0.4698**	**0.6575**
2025	0.0343	0.4505	0.0374	0.5187

4.2. Analysis of VAR filtered volume-price relationships

I this process, we firstly check the results of theVAR itself, with the average lag selected with minium AIC for each day, and p-values of JB test, LB test and ARCH LM test shown in Table 7. The results indicate:

1. From the period from 2017 to 2020 exhibits strong autocorrelation and significant non-normality in the VAR residuals, reflecting a volatile and structurally complex market.

2. In contrast, from 2021 onward, both the trading volume (Vr) and price (Pr) residuals display diminishing Ljung-Box and ARCH effects, with Jarque-Bera tests suggesting convergence toward normality.

3. This reflects a structural simplification of the market dynamics, with the VAR model capturing most of the linear dependencies — a potential signal of increased market efficiency or post-regime stabilization.

Moreover, we observe a consistent asymmetry between the Ljung-Box p-values of volume and price residuals, where the residual autocorrelation in trading volume is significantly stronger than that in price across most years, indicating volume series contains more linear conponents that may not be clearly explained by only volume-price relationship. But generally, the results indecating our VAR models are relatively well fitted on these data and non-linear conponents are well extracted, with residuals shown weak non-normality, ARCH effects on both series and autocorrelation at least on $r^{(P)}$.

Table 7: Annual Significance of multi-test in VAR residuals

Year	k_{AIC}	$p_{\mathrm{JB}}^{(\mathrm{Vr})}$	$p_{\mathrm{JB}}^{(\mathrm{Pr})}$	$p_{\mathrm{LB}}^{(\mathrm{Vr})}$	$p_{\mathrm{LB}}^{(\mathrm{Pr})}$	$p_{\mathrm{ARCH}}^{(\mathrm{Vr})}$	$p_{\mathrm{ARCH}}^{(\mathrm{Pr})}$
2017	3.000	**4.21e-13**	**0.0000**	**2.87e-11**	0.2444	**0.0093**	**0.0070**
2018	3.000	**1.09e-03**	**1.35e-04**	**5.44e-05**	0.3390	**0.0165**	**0.0220**
2019	3.000	**6.20e-07**	**0.0000**	**5.48e-12**	0.1455	**0.0331**	**0.0285**
2020	3.000	**5.08e-04**	**5.46e-18**	**2.48e-14**	0.2462	**0.0208**	**0.0202**
2021	2.994	**1.99e-03**	**1.06e-03**	**1.49e-04**	0.2556	**0.0101**	**0.0146**
2022	2.997	**2.20e-03**	**2.67e-03**	**1.70e-03**	0.2427	**0.0141**	**0.0121**
2023	2.633	1.49e-01	1.82e-01	1.55e-01	0.3418	0.1666	0.1917
2024	2.771	8.15e-02	8.91e-02	7.30e-02	0.2776	0.1008	0.0895
2025	2.938	**1.16e-02**	**1.05e-02**	**3.74e-03**	0.1781	**0.0284**	**0.0297**

After obtaining the residuals, we recompute the TE on the residuals, It is notable that the $\bar{p}$ of both directions degenerate to the same level, with $\bar{p}$ for $TE_{Pr\to Vr}$ higher than before(Tab. 4) and $\bar{p}$ for $TE_{Vr\to Pr}$ lower than before(Tab. 5), indecating that linear components dominates the information flow from price to volume, while non-linear components dominates the information flow from volume to price. The dynamic remains almost the same on moments of the bidirectional TE, with few

Table 8: Average p-values of TE for Residuals

Year	$\bar{p}_{Pr\to Vr}$	$\bar{p}_{Vr\to Pr}$
2017	0.2522	0.1496
2018	0.2552	0.3227
2019	0.2436	0.2734
2020	0.3914	0.3671
2021	0.3795	0.4743
2022	0.3619	0.4330
2023	0.4007	0.4486
2024	0.3118	0.3743
2025	0.2276	0.3205

Table 9: 95% and 99% Quantiles of Residual TE from VAR

Year	$Q_{0.95}^{Pr\to Vr}$	$Q_{0.99}^{Pr\to Vr}$	$Q_{0.95}^{Vr\to Pr}$	$Q_{0.99}^{Vr\to Pr}$
2017	0.0335	0.0392	0.0869	0.1455
2018	0.0373	0.0650	0.0394	0.0585
2019	0.0339	0.0372	0.0397	0.0517
2020	0.0335	0.0378	0.0325	0.0375
2021	0.0310	0.0369	0.0286	0.0331
2022	0.0309	0.0428	0.0315	0.0365
2023	0.5399	0.6538	0.5670	0.6928
2024	0.5471	0.6291	0.5268	0.6531
2025	0.0388	0.5075	0.0350	0.4129

4.3. Trend analysis

We apply the Mann-Kendall test to the yearly transfer entropy (TE) series to examine directional trends in information flow. The results reveal that TE from price to volume is more likely to show significant changes over time, with a tendency to weaken in recent years, which might be caused by the high value of Bitcoin. In contrast, TE from volume to price, when it shows a trend, is nearly always increasing. This suggests that volume is becoming a stronger leading indicator for price in the cryptocurrency market. Additionally, residual-based TE series uncover asynchronous shifts in information structure, particularly in 2021 and 2022, which may reflect behavioral changes associated with the market's transition from bull to bear conditions.

5. Conclusions

CRediT authorship contribution statement

Funding Sources

The work is supported by the characteristic & preponderant discipline of key construction universities in Zhejiang province (Zhejiang Gongshang University − Statistics). Fund Number : 1020JYN6523001G-026; 1020JYN6524001G-020

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. No conflict of interest have been declared by the authors.

References

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, *85*, 461.

Appendix

Table 10: Moments of $TE_{Pr \to Vr}$ (Residuals)

Year	Mean	Std	Skewness	Kurtosis
2017	0.0248	0.0054	0.4983	0.6252
2018	0.0273	0.0200	8.2555	75.2433
2019	0.0258	0.0055	0.1479	-0.6447
2020	0.0224	0.0060	0.8790	1.8067
2021	0.0222	0.0132	13.7429	229.5683
2022	0.0243	0.0255	11.6713	141.2074
2023	0.1271	0.1917	1.7327	2.0264
2024	0.0935	0.1726	2.3455	4.1877
2025	0.0407	0.0845	5.5176	29.5499

Table 11: Moments of $TE_{Vr \to Pr}$ (Residuals)

Year	Mean	Std	Skewness	Kurtosis
2017	0.0412	0.0292	2.8593	11.8731
2018	0.0256	0.0100	2.7837	11.6623
2019	0.0270	0.0090	1.0267	1.8689
2020	0.0226	0.0057	0.3192	0.1847
2021	0.0205	0.0149	16.2695	292.4390
2022	0.0230	0.0238	11.6990	143.4347
2023	0.1401	0.2128	1.5154	1.0093
2024	0.0933	0.1708	2.2114	3.4608
2025	0.0364	0.0755	6.0254	37.3487