Final report

Team 2

December 6, 2025

Crypto 시장에서 고빈도 데이터 기반 거래소간 정보선도력 측정이 가능하다. Cointegration 기반 정보선도력 추정시 noise에 의한 bias가 발생할 수 있다. 따라서 High-frequency data anaysis 방법을 적용하여 noise를 제거 후, BTC-USDT 가격에 대해 거래소간 정보선도력을 추정하고 결과를 검증한다.

Contents

1 Introduction

1.1 Literature Research

Cointegration based methods can be applied to high-frequency crypto price data, To observe information leadership across exchange. High-frequency, intraday event pattern[3][4] can be observed in crypto market, which is highly fragmented.

We focus on daily information lead-lag effect among BTC-USDT exchanges, although it known that there are various time scales of cross-exchange price discovery process. [1][5][2].

Jump and noise hadling can be an important[7] for ILS(Information Leadership Share) estimation. Truncation and pre-averaging was selected for log-price time series of BTC-USDT from 2020-08-19 to 2020-09-03. Relatively simple methods and easily available data, volume report, fee data was selected to minimize our implementation mistake.

Our result shows that the truncation and the pre-averaging have effect on noise profile and the cointegration analysis result. We verify that some HFT research methods such as truncation and pre-averaging can be applied to the ILS measure.

1.2 Data

10-level quote data of the folowing BTC-USDT exchange was collected : Bitfinex, FTX, Huobi, OKEX, Probit, Upbit. Data sampling frequency is about 1 second. Date period is from 2020-08-19 to 2020-09-03. Mid price was extracted frmo the quote data for simplicity, while use of the micro-price[8] can be an improvement.

Exchange Name	Coverage of the collected data	Fake-adjusted daily volume [6] as of 2019-05-28	Standard transaction fee	Test Pass Score[6]
Bitfinex	0.923985	189054133	0.002	5
FTX	0.930627	50444971	unknown	4
Huobi	0.943173	3006542737	0.001	4.5
OKEX	0.941697	4728575030	0.002	4
Probit	0.931365	unknown	unknown	unknown
Upbit	0.539852	963941563	0.0005	4.5

2 Denoising

2.1 Jump adjust

Jump-adjusted price was reconstructed by cumulatively summing the truncated log-return.

2.1.1 Truncation with static threshold

Truncation with threshold parameter $\tau = \sigma$ was applied.

$$\tilde{r}_t = \Psi_\tau(r_t) = \mathrm{sgn}(r_t)\min\left(|r_t|, \tau = \sigma_t\right)$$

where $\sigma = \sqrt{\frac{1}{N}\sum_{t=0}^{N} r_t^2}$

Truncated series satisfies the sub-Gaussian condition, by the Hoeffding's lemma.

$$E[\exp(\lambda r_t)] \leq \exp\left(\frac{\lambda^2\tau^2}{2}\right) = \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

2.1.2 Truncation with MLE

Truncation treshold τ was dynamically optimzed based on the daily volatility. It was assumed that the microstructure noise small enough compared to the jump.

$$\frac{1}{N}\sum_{t=0}^{N} r_t^2 = \sigma^2_{\text{true}} + \sigma^2_{\text{noise}} + \sigma^2_{\text{jump}} \sim \sigma^2_{\text{daily}} + 0 + \sigma^2_{\text{jump}}$$

To Minimize jump effect σ^2_{jump}, loss function was set to the following :

$$L(\tau) = \left(\sum_{t=0}^{N} \text{Huber}((\tilde{r}_t^\tau)^2 - \sigma^2_{\text{daily}})\right)$$

For each day, τ was optimized to minimize the loss. The τ was used for the truncation.

2.2 Microstructure noise adjust

2.2.1 Pre-Averaging

$N = 60 * 60 * 24 = 86400, K = \sqrt{N} = 293$ and Window function $g(x) = \min(x, 1-x)$ was chosen. From the pre-averaged return equation

$$\bar{Y}(t_r) = \sum_{l=1}^{K-1} g(\frac{l}{k})[Y_{t_{r+l}} - Y_{t_{r+l-1}}] = \sum_{l=1}^{K-1} g(\frac{l}{k})Y_{t_{r+l}} - \sum_{l=1}^{K-1} g(\frac{l}{k})Y_{t_{r+l-1}} = Y^{\text{PR}}_{t_{r+l}} - Y^{\text{PR}}_{t_{r+l-1}}$$

Pre-averaged price $Y_t^{\text{PR}} = \sum_{l=1}^{K-1} g(\frac{l}{k})Y_t$ is defined.

Pre-averaging was applied to the jump-adjusted price to obtain noise-adjusted price.

2.2.2 Fractional differentiation

To check the effect of discretization noise as a source of microstructure noise, Fractional differenentiation or order 0.2 was appplied to minimize the discretization effect.

Fractional differentiation is defined as

$$\frac{\partial^n x_t}{\partial t^n} = x(t) * w(s) \quad \text{where} \quad w(s) = \left(\frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)}(-1)^s\right)$$

Order $n = 0.2$ was chosen to ensure stationary with minimum differenentiation.

Resulting time series was cumulatively summed and used as noise-adjusted price.

3 노이즈 크기 측정

3.1 Spectral decomposition

3.2 AC(1)

4 Indicator

5 Conclusion

References

[1] Yukihiro Aiba, Naomichi Hatano, Hideki Takayasu, Kouhei Marumo, and Tokiko Shimizu. Triangular arbitrage as an interaction among foreign exchange rates. *Physica A: Statistical Mechanics and its Applications*, 310(3-4):467–479, 2002.

[2] Jakob Albers, Mihai Cucuringu, Sam Howison, and Alexander Y Shestopaloff. Fragmentation, price formation and cross-impact in bitcoin markets. *Applied Mathematical Finance*, 28(5):395–448, 2021.

[3] Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998, 2017.

[4] Albert J Menkveld and Bart Zhou Yueshen. The flash crash: A cautionary tale about highly fragmented markets. *Management Science*, 65(10):4470–4488, 2019.

[5] Robbie Morrison, Natasha CHL Mazey, and Stephen C Wingreen. The dao controversy: the case for a new species of corporate governance? *Frontiers in Blockchain*, 3:25, 2020.

[6] Alameda Research. Investigation into the legitimacy of reported cryptocurrency exchange volume, 2019.

[7] Shulin Shen and Eric Zivot. In defense of information leadership share: A response to shrestha and lee (2023). *Available at SSRN 4752194*, 2024.

[8] Sasha Stoikov. The micro-price: a high-frequency estimator of future prices. *Quantitative Finance*, 18(12):1959–1966, 2018.