Information disclosure on crypto assets by Chainsight

Tatsuyoshi Okimoto[*], Shumpei Koike[†], and Toyotaka Sakai[‡]

July 14, 2023

Abstract

Chainsight is a novel blockchain project that records on-chain data of crypto assets as easy-to-use time series data. We present a series of economic, political, and social indicators of crypto assets that can be created using those data. Each of these indicators represents key information about crypto assets. Even if data about a crypto asset is publicly available, it does not necessarily mean that important information about that crypto asset is publicly available. This is because data has no value unless it is transformed into information. Chainsight promotes the democratic ideal of web3 by facilitating this transformation.

1 Introduction

Since Satoshi Nakamoto created Bitcoin with blockchain technology, thousands of crypto assets based on this technology have been released (Nakamoto 2008). As long as the transactions of a crypto asset are recorded on a public blockchain, anyone can monitor them. Because of this property, blockchain technology has been evaluated in terms of transparent information sharing, which is essential to democratic governance. However, things are not that simple.

A blockchain is a database of countless numbers and letters from which most people cannot directly extract useful information. For example, although the database enables us to obtain the distribution of a crypto asset over addresses, obtaining it requires time and effort, and extracting useful information from the obtained distribution necessitates knowledge of statistics and econometrics. In order for transparent information sharing to be properly effective, useful information must be extracted from the data on the blockchain, and such information must be made available to the public. Given that comparative analysis is a fundamental methodology of sciences, the information should be provided for multiple crypto assets. This is what Chainsight hopes to achieve.

More specifically, information on crypto assets very often takes the form of economic, political, and social indicators about the assets. During the development process of Chainsight, it is expected that various such indicators will be provided. This paper describes some of them.

We explain non-financial, and financial indicators in Sections 2 and 3, respectively. Finally, we conclude our discussion in Section 4.

[*]Professor of Economics, Keio University and Economic Advisor, HorizonX
[†]CTO, HorizonX
[‡]Professor of Economics, Keio University and Economics Advisor, HorizonX

2 Non-financial indicators

On-chain data show transactions of a crypto asset between addresses and reveal its quantity in each address. This information is useful for understanding various relational properties of addresses and network structure. To accomplish this, we first introduce basic definitions of a crypto asset.

The finite set of *addresses* is $A = \{a_1, \ldots, a_N\}$, and the *quantity* of a crypto asset in address $a \in A$ is $q(a) \geq 0$. The *ratio* of the quantity of the crypto asset in address $a \in A$ is

$$\text{ratio}(a) \equiv \frac{q(a)}{\sum_{a' \in A} q(a')}.$$

Similarly, the ratio for a set of addresses $B \subset A$ is

$$\text{ratio}(B) \equiv \frac{\sum_{a \in B} q(a)}{\sum_{a \in A} q(a)}.$$

A *binary relation* $\rightarrow$ on A is a subset of $A \times A$. A binary relation $\rightarrow$ on A is *reflexive* if $(a, a) \in \rightarrow$ for all $a \in A$. When $(a, b) \in \rightarrow$, we interpret that a has ever sent money to b and write $a \rightarrow b$.

2.1 Centrality

The Herfindahl-Hirschman index (Herfindahl 1950; Hirschman 1954) is

$$\text{HHI}(A) \equiv \sum_{a \in A} \text{ratio}(a)^2.$$

When calculating all squares in A is computationally intensive, we may choose to restrict our attention to a subset of addresses $B \subset A$ and calculate

$$\text{HHI}(B) \equiv \sum_{a \in B} \text{ratio}(a)^2.$$

Typical candidates of B might be

$$B = \{a \in A : q(a) \geq r\} \text{ where } r > 0 \text{ is a minimum quantity for calculation,}$$
$$B = \{a \in A : \text{ratio}(a) \geq s\} \text{ where } s \in (0, 1] \text{ is a minimum share for calculation.}$$

2.2 Criminal holdings

Consider any reflexive binary relation $\rightarrow$ on A. For any $a, b \in A$, *a is connected to b with respect to $\rightarrow$* if there exists a sequence of addresses $a_1, a_2, ..., a_K \in A$ such that $a_1 = a$, $a_K = b$, and $a_k \rightarrow a_{k+1}$ for all $k = 1, ..., K - 1$. For any subset $B \subset A$, let

$$\bar{B} \equiv \{a \in A : \exists b \in B,\ a \text{ is connected to } b \text{ with respect to } \rightarrow\}$$

be the set of addresses connected to at least one address in B with respect to $\rightarrow$.[1] Note that $B \subset \bar{B}$ because of the reflexivity of $\rightarrow$. Given any $a \in A$, we can determine if $a \in \bar{B}$ or not using the Breadth-First Search algorithm. When B is the set of addresses that have ever been used in crimes, ratio($\bar{B}$) can be seen as the ratio of asset quantities in the set of "contaminated" addresses with respect to B. Then

$$\mathrm{ratio}(\bar{B}) \equiv \frac{\sum_{b \in \bar{B}} q(b)}{\sum_{a \in A} q(a)}$$

is the ratio of asset quantities in such contaminated addresses.

This way of distinguishing contaminated and non-contaminated addresses is called the union-find algorithm. This algorithm is employed by Foley et al. (2019) to analyze the ratio of contaminated addresses of Bitcoin. In traditional finance, where bank account information is not publicly available, this type of ratio can hardly be calculated.

2.3 Decentralization of transaction network

Network structure is another very interesting piece of information that can be gleaned from on-chain data. For example, one network might resemble a star network centered around a hub node, while another might have many P2P connections.

Consider any reflexive binary relation $\rightarrow$ on A. For any $a \in A$, let

$$\mathrm{adj}(a) \equiv \{b \in A : a \rightarrow b \text{ or } b \rightarrow a\}$$

be the set of "adjacent" addresses of a with respect to $\rightarrow$. Then $|\mathrm{adj}(a)|$ is the number of adjacent addresses of a. In a very centralized network, there exists a "center" $a^* \in A$ for which $\mathrm{adj}(a^*) = A$ and $\mathrm{adj}(a) = \{a, a^*\}$ for most of $a \in A$. Meanwhile, in a much decentralized network, the numbers $|\mathrm{adj}(a)|$ are very close to each other.

The *Gini coefficient*, which is usually used to measure income differences, can be used to evaluate the degree of network decentralization:

$$\frac{\sum_{a \in A} \sum_{b \in A} |\mathrm{adj}(a) - \mathrm{adj}(b)|}{2 \cdot |A| \cdot \sum_{a \in A} |\mathrm{adj}(a)|}.$$

Let

$$P(k) \equiv \frac{|\{a \in A : |\mathrm{adj}(a)| = k\}|}{|A|}$$

be the proportion of addresses whose number of adjacents is k. Baraba and Albert (1999) observe that the distribution of $P(k)$ often follows a power-law distribution, say, $P(k) \sim k^{-\gamma}$ for some exponent parameter $\gamma > 0$. As γ becomes higher, this distribution converges to zero more quickly as k increases. Conversely, as γ becomes smaller, this distribution becomes more long-tailed. Thus γ can be seen as a measure to evaluate the centrality of a given network. Exponent γ can be estimated by regressing

$$\ln P(k) = \alpha + (-\gamma) \ln k + \varepsilon,$$

[1] Because crypto assets can be usually sent from anyone to anyone, one may want to define $a \rightarrow b$ only when a large enough quantity has been sent from a to b.

where the error term ε is assumed to follow a normal distribution. For other interesting indicators of network decentralization, we refer to the survey by Bonchev and Buck (2005).

3 Financial indicators

Using on-chain data provided by Chainsight, we can also compute indicators of liquidity and volume volatility, as in traditional finance. For instance, the prices of crypt assets are available from DEX. Therefore, price volatility and volume volatility can be comuted to show the crypto asset's market value. In this section, we introduce major risk and liquidity measures that can help assess the stability of a crypto asset.

3.1 Realized volatility

Realized volatility (RV) is one of the most widely adopted measures in finance to assess the volatility and risk of an asset. Historically, Andersen and Bollerslev (1998) propose RV as a model-free estimator of daily return volatility. Subsequently, Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) demonstrate more formally that RV is an unbiased and highly efficient estimator of daily return volatility.

To define RV, consider intraday asset return data denoted by r_{ti}, where t represents the day and i indicates the period within a day. RV is defined using intraday asset returns as

$$RV_t \equiv \sqrt{\sum_{i=1}^{N} r_{ti}^2}, \tag{1}$$

where N is the number of observations per day.

Barndorff-Nielsen et al. (2008) further introduce a realized semivolatility, which focuses on measuring volatility based only on downwards (upwards) moves using intraday data.[2] Specifically, the downside realized semivolatility (DRSV) and upside realized semivolatility (URSV) are defined respectively as follows:

$$DRSV_t \equiv \sqrt{\sum_{i=1}^{N} \min(0, r_{ti})^2}, \tag{2}$$

$$URSV_t \equiv \sqrt{\sum_{i=1}^{N} \max(0, r_{ti})^2}. \tag{3}$$

By computing the RV as well as DRSV and URSV for asset prices and volumes, we can compare the general risk as well as the downside and upside risk of each asset in terms of price and liquidity.

As an empirical example, RV (1), SRSV (2), and URSV (3) are calculated using 5-minute intraday data for Eitherum (ETH), Bitcoin (BTC), Binance Coin (BNB), and Uniswap (UNI), for the year 2022. Table 1 summarizes the average values of the RV, SRSV, and URSV for each asset, as well as the results of hypothesis tests for no difference between SRSV and URSV. In addition, the top-left panel of Figure 1 illustrates the RV of each crypto asset. As can

[2] Precisely speaking, Barndorff-Nielsen et al. (2008) propose a semivariance, which is a square of semivolatility in this paper.

be seen, RVs of all assets are highly correlated, showing strong co-movements. UNI tends to have the highest RV, followed by ETH, while BTC and BNB appear to have relatively smaller RV, compared to UNI and ETH. The bottom two panels of Figure 1 display the SRSV (left) and URSV (right). Their general patterns are similar to the pattern of RV, with UNI having the highest semivolatility followed by ETH, while BTC and BNB exhibit relatively lower semivolatility. The results of hypothesis tests support these observations, indicating no significant difference between the means of SRSV and URSV.

3.2 Value at Risk (VaR)

Value at Risk (VaR) is another well-known risk measure in finance. It has gained widespread acceptance and been adopted by regulators, such as the US Securities and Exchange Commission and the UK Securities and Futures Authority, as a key indicator of financial risk. Additionally, the Basel Committee employs the VaR to evaluate the market risk component of bank capital requirements.

In essence, $\alpha\%$ VaR quantifies the maximum potential loss on an asset or portfolio over a specified time horizon with a probability $\alpha\%$. Similarly, we could compute $(100-\alpha)\%$ VaR potential maximum gain with probability $\alpha\%$ to evaluate the upside risk for short-sellers. Using the intraday data, we can compute the realized (ex-post) VaR for each day, enabling a comparison of the potential risk and stability of every crypto asset.

Using 5-minute intraday data, we calculate the realized 1% VaR and 99% VaR of each crypto asset on a daily basis. To facilitate comparison, we multiply the 1% VaR by -1. Table 2 provides the averages of these VaRs along with the results of hypothesis tests for no difference between the two VaRs. With 5-minute intraday data, these values can be interpreted as the average realized maximum losses and gains with a probability 1% each day within a 5-minute period throughout 2022. The results indicate that there is no significant asymmetry between the losses and gains. In addition, the findings suggest that UNI has the highest VaR, followed by ETH, BNB, and BTC, which is consistent with the results of the RV, but the differences between UNI and the other crypto assets are more pronounced. Figure 2 depicts the dynamics of 1% VaR (left) and 99% VaR (right). As can be seen, UNI consistently exhibits the highest VaR for the majority of the days, while BTC tends to have the lowest VaR most of the days.

3.3 Turnover ratio

One of the commonly used measures of liquidity is trading volume. The trading volume can be measured as the quantity of assets traded or the value of assets traded. Trading volume is considered a proxy for liquidity, as it reflects the trading activity in the market. However, it is important to note that assets with a larger supply tend to have higher trading volumes, which may not necessarily mean higher liquidity.

To address this concern, we calculate the turnover ratio (TR) as a ratio between trading volume and circulating supply. If intraday data are available, the turnover ratio can be computed as

$$TR_t \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{TV_{ti}}{CS_t},$$

where TV_{ti} is a trading volume in period i at time t and CS_t is a circulating supply at time

t. TR can also be calculated using the ratio between the traded market value and the market capitalization, which is called the price turnover ratio (PTR). If the same price is used to compute the traded market value and market capitalization, TR and PTR will yield the same value. TR and PTR are regarded as more appropriate liquidity measures than trading volume as they account for the market size of each asset, as discussed by Gabrielsen et al. (2011) and Le and Gregoriou (2020).

We calculate the TR based on 5-minute intraday data for each crypto asset, which will be equivalent to the PTR if the same price is used to compute the trading volume value and market capitalization. Table 3 reports the average values and standard errors of the TR for each asset, as well as the results of hypothesis tests testing no difference in TR between ETH and other assets. The results indicate that ETH has the highest TR. While BTC has a comparable TR, the hypothesis test suggests that the TR of ETH is statistically significantly higher than that of BTC. The TRs of BNB and UNI are significantly smaller than those of the other two assets, with UNI having a much smaller TR, meaning that BNB and UNI have lower liquidity compared to ETH and BTC. Figure 3 displays the dynamics of TR for each asset in separate panels, highlighting that the varying magnitudes of the TR are quite different across crypto assets. Generally, the TR of each asset shares a similar pattern with a high correlation and a decreasing trend throughout 2022.

3.4 Return to volume ratio

Amihud (2002) proposes a return to volume ratio (RVR) as a measure of illiquidity, based on the notion that a liquid market allows traders to complete trading with minimal impact on prices. Specifically, the RVR is defined as the ratio between the returns and the market value of the traded assets. If intraday data are available, the RVR can be calculated as

$$RVR_t \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{|r_{ti}|}{TMV_{ti}},$$

where TMV_{ti} is traded market value of assets in period i at time t.

Amihud (2002) demonstrates a strong relationship between the RVR and price impact estimated by microstructure data. Goyenko et al. (2009) compare various illiquidity measures and confirm that the RVR outperforms others in capturing cross-sectional or time-series variations of static price impact.

Using 5-minute intraday data, the RVR is computed for each asset throughout 2022. Table 4 presents the average and its standard error along with the hypothesis test results examining no difference in RVR between ETH and other assets. The results indicate that BTC has the lowest RVR followed by ETH. Although the difference is small, the null hypothesis of no difference in RVR between ETH and BTC is rejected even at the 1% significance level, meaning that BTC has lower illiquidity than ETH. In contrast, BNB and UNI exhibit significantly higher RVR compared to ETH, particularly with UNI having the far highest RVR. This implies that illiquidity could be a serious concern for UNI. More specifically, it indicates that the price of UNI is much more sensitive compared to other assets and the same amount of trade could affect the Uniswap price significantly more than the others. Figure 4 illustrates the dynamics of RVR for each asset in separate panels, considering the substantial difference in magnitudes across the four assets. Although each asset displays varying levels of illiquidity, they still share similar dynamics with a high correlation. More notably, RVR for all assets exhibits a similar increasing trend with a more significant rise towards the end of 2022.

4 Conclusion

We have presented several indicators, depicting various key aspects of crypto assets. These indicators can be calculated using on-chain data along with off-chain data, depending on what they are. Compared to traditional assets and traditional finance, crypto assets and decentralized finance are often praised for their democratic governance. However, democratic governance only makes sense when critical information is publicly available. Even if transaction data are on public blockchains, such information disclosure is not essentially achieved unless the data are processed into information that ordinary people can understand. This process will be implemented by Chainsight, bringing true democratization to the world of crypto assets and decentralized finance.

References

Amihud, Y. (2002) "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets*, Vol. 5, pp. 31–56.

Andersen, T. G., and Bollerslev, T. (1998) "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecast," *International Economic Review*, Vol. 39(4), pp. 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003) "Modeling and Forecasting Realized Volatility," *Econometrica*, Vol. 71(2), pp. 529–626.

Barndorff-Nielsen, O. E., and Shephard, N. (2002) "Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Series B*, Vol. 64, pp. 253–280.

Barndorff-Nielsen, O. E., Kinnebrock, S., and Shephard, N. (2008) "Measuring Downside Risk-Realized Semivariance," The Oxford-Man Institute, University of Oxford, Working paper.

Baraba, A.-L. and Albert, R. (1999) "Emergence of Scaling in Random Networks," *Science*, pp. 509–512.

Bonchev, D. and Buck, G. A. (2005) "Quantitative Measures of Network Complexity," in *Complexity in Chemistry, Biology, and Ecology*, eds by Bonchev, D. and Rouvray, D. H., Springer.

Foley, S., Karlsen, J. R., and Putniņš, T. J. (2019) "Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies?" *Review of Financial Studies*, Vol. 32(5), pp. 1798–1853.

Gabrielsen, A., Marzo, M. and Zagaglia, P. (2011) "Measuring Market Liquidity: An Introductory Survey," *SSRN Electronic Journal*.

Goyenko, R. Y., Holden, C. W. and Trzcinka, C. A. (2009) “Do Liquidity Measures Measure Liquidity?” *Journal of Financial Economics*, Vol. 92, pp. 153–181.

Herfindahl, O. C. (1950) *Concentration in the US Steel Industry.* Ph.D. dissertation, Columbia University.

Hirschman, A. O. (1945) *National Power and the Structure of Foreign Trade.* University of California Press, Berkeley.

Le, H. and Gregoriou, A. (2020) “How Do You Capture Liquidity? A Review of the Literature on Low-Frequency Stock Liquidity,” *Journal of Economic Surveys*, Vol. 34(5), pp. 1170–1186.

Nakamoto, S. (2008) “Bitcoin: A Peer-to-Peer Electronic Cash System,” https://assets.pubpub.org/d8wct41f/31611263538139.pdf

Appendix: Tables

Table 1: Summary of RV, DRSV, and URSV

		ETH	BTC	BNB	UNI
RV	Mean	1.2768	0.7182	0.9527	2.1320
	Std. Dev.	1.6833	0.9515	1.5726	2.3027
DRSV	Mean	0.6285	0.3572	0.4802	1.0816
	Std. Dev.	0.8129	0.4704	0.7521	1.1874
URSV	Mean	0.6483	0.3610	0.4725	1.0504
	Std. Dev.	0.9191	0.5106	0.8452	1.1786
Test results of	t-stat	−0.8987	−0.2981	0.4991	1.0979
DRSV=URSV	P-value	0.3694	0.7658	0.6180	0.2730

Table 2: Summary of 1%VaR and 99%VaR

		ETH	BTC	BNB	UNI
1% VaR	Mean	0.6668	0.5040	0.5773	0.8816
	Std. Dev.	0.3641	0.2748	0.3185	0.3876
99% VaR	Mean	0.6840	0.5083	0.5645	0.8738
	Std. Dev.	0.3965	0.3024	0.3818	0.4114
Test results of	t-stat	−1.4756	−0.4943	1.3931	0.6203
−1% VaR= 99% VaR	P-value	0.1409	0.6214	0.1645	0.5355

Table 3: Summary of TR

TR	Mean	0.0067	0.0062	0.0004	1.15E-05
	Std. Dev.	0.0044	0.0043	0.0003	1.06E-05
Test results of no	t-stat		5.991	29.04	29.24
difference from ETH	P-value		5.02E-09	8.34E-97	1.48E-97

Table 4: Summary of RVR

		ETH	BTC	BNB	UNI
RVR	Mean	2.18E-10	7.78E-11	1.43E-08	4.09E-05
	Std. Dev.	1.48E-10	5.34E-11	8.14E-09	8.70E-05
Test results of no	t-stat		25.90	−33.47	−8.982
difference from ETH	P-value		1.31E-84	3.94E-113	1.46E-17

Appendix: Figures

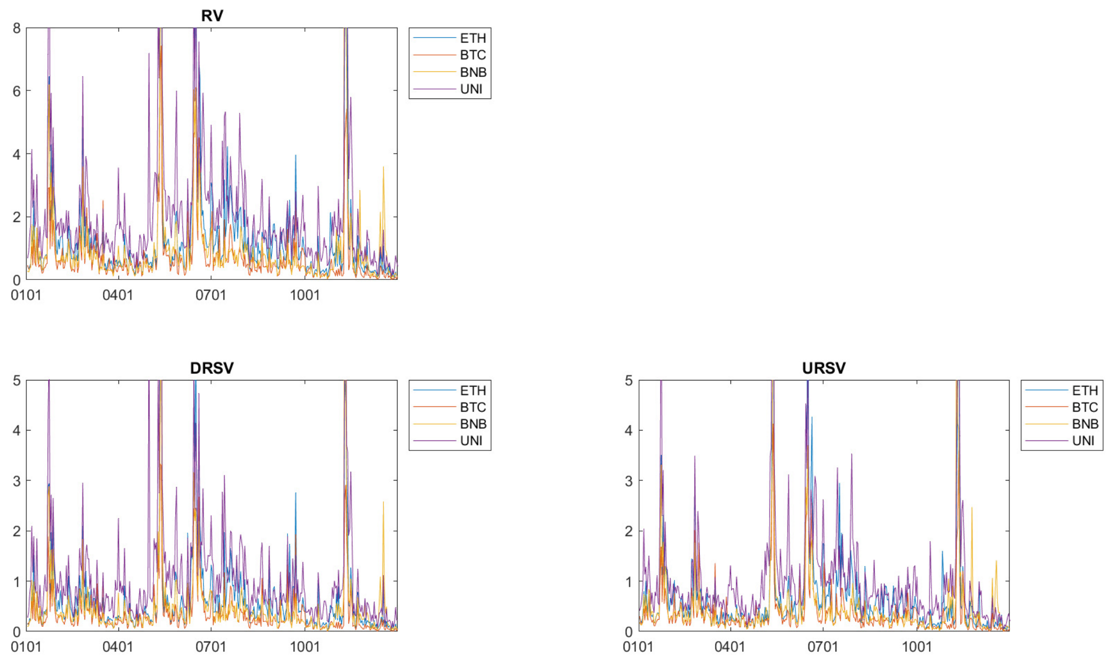

Figure 1: Dynamics of RV, DRSV, and URSV

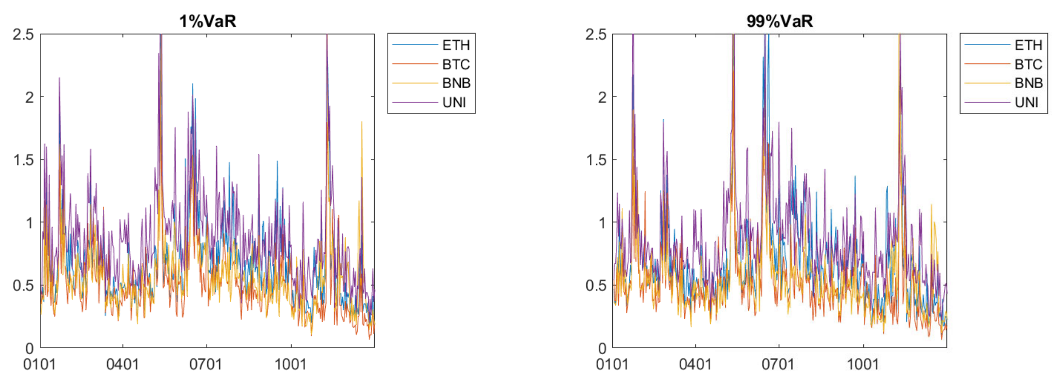

Figure 2: Dynamics of 1% VaR and 99% VaR

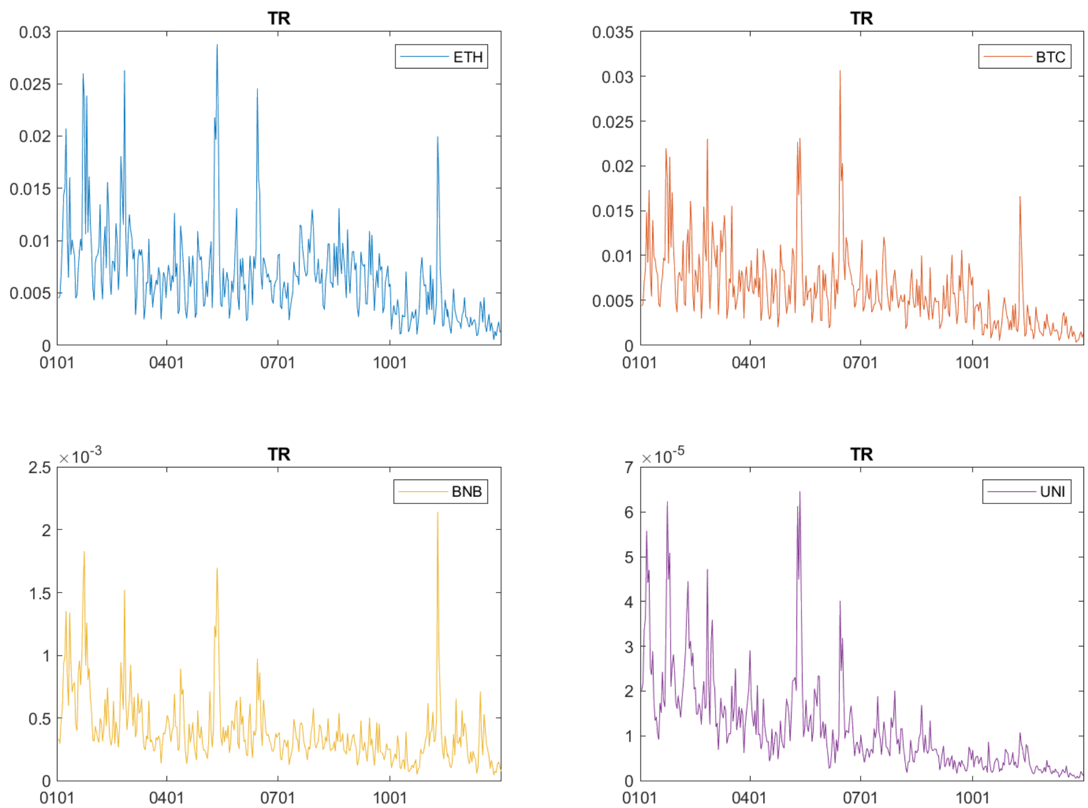

Figure 3: Dynamics of TR

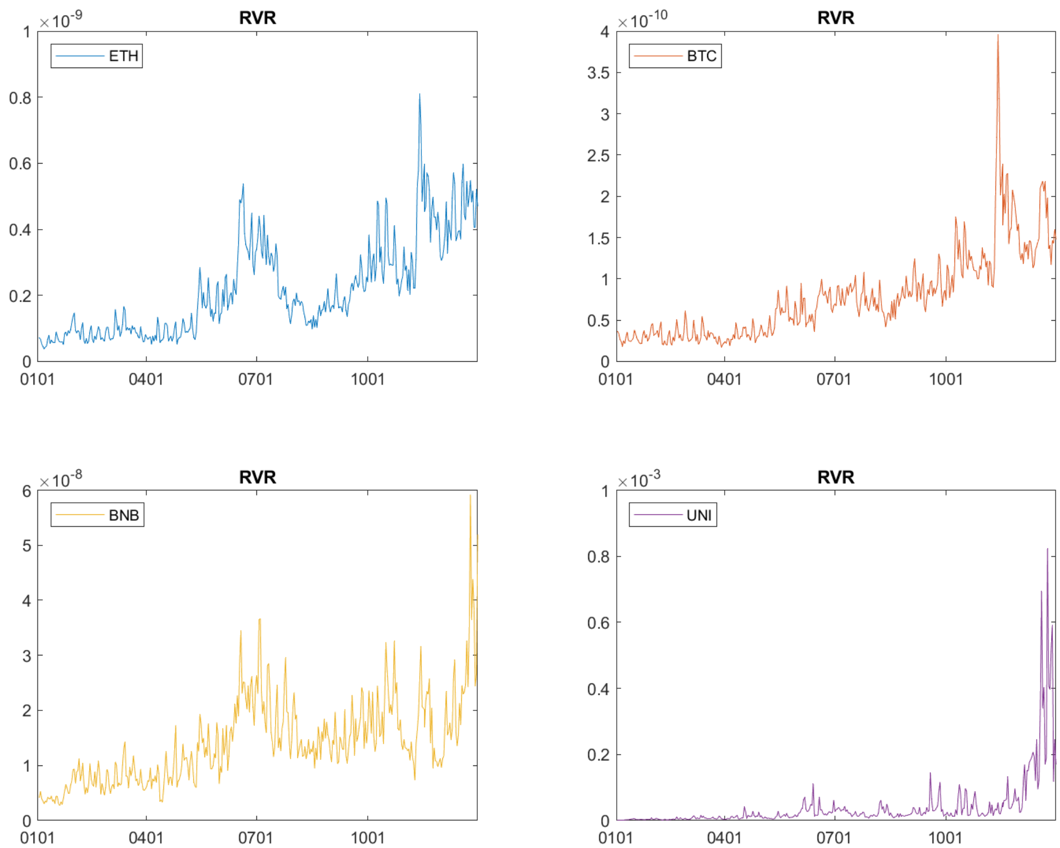

Figure 4: Dynamics of RVR