MSDM5003 Final Project

Stock price return distribution of a stock market (Crypto Spot Market)

Chun Yin Wong

1 Introduction

Cryptocurrencies have gained a lot of attention in recent years due to their potential as a new form of digital money. As a result, the cryptocurrency market has grown rapidly. One interesting aspect of the cryptocurrency market is its high degree of price fluctuation, which can be quite dramatic at times. This makes it an interesting area of study for power law. In this report, we will investigate the price return distribution in the cryptocurrency market and compare the result with traditional developed stock market.

2 Data Source and preprocessing

The historical 1 minute and 5 minutes data are downloaded via Binance Restful API from 1-1-2021 to 10-12-2022 (708 trading days). Since not all cryptocurrencies are actively trading, only top 100 cryptocurrencies with the highest trading volume as of 10-12-2022 are included. Log return (Eq. 1) is then applied to the close price and then the values are normalized by eq. 2 to compare different assets in the same scale [1].

3 Choosing Δt

In order to choose an appropriate Δt in Eq2, different Δt are shortlisted for comparison. From Fig 1, we can see that both 1 minute ($\Delta t = 1$) and 5 minutes ($\Delta t = 5$) are having a similar long tail on the CDF of log returns. Hence, $\Delta t = 5$ is chosen to save the computational power.

4 The exponent of the aggregated return

4.1 Linear regression

The aggregated return with $\Delta t = 5$ contains around 17 million data points. Since from Fig 1, we can see that the assets follow very similar CDF on log return, the data points can be aggregated together for further analysis. The corresponding CDF with the aggregated data is shown in Fig2. To calculate the exponents of the power law for the tail distribution, a linear regression is applied on the CDF under log-log scale. As a result, for the positive returns, the alpha is 2.5814 and equal to 2.7660 for the negative returns.

4.2 Hill estimator

An alternative method to estimate the exponent (α) is Hill Estimator. The equation is explained in Eq 3. If k is large enough, the estimated alpha is consistent. From the Eq 3, we can observe that by taking k = 1000 the values become stable. And the estimated α from positive return is 2.714 and from the negative return is 2.620. These exponents values are very close to the exponents we observed from the linear regression above.

5 Distribution of alpha

5.1 Alpha distribution

Besides aggregating the log returns together, we also studied the distribution of alpha based on individual's cryptocurrency log returns. The power-law exponents are obtained by linear regression. The results are show in Fig.3.

5.2 The outlier of alpha

Yet, from Fig3, we can observe that the histogram has some abnormally large alpha. After investigating to those cryptocurrencies, these assets prices have an extraordinarily jump in 5 minutes. For example, DREP-USDT went up from 0.026 to 4.817 in 5 minutes, which equals to a +18400% change and 40.49 standard deviation away. These outlier movement impacted the regression results and leading to a wrong exponent. To measure the fitness of linear regression, R^2 score is introduced [2], and the equation can be found in Eq4. The R^2 of the linear regression is 0.75 for DREP-USDT. But, from Fig4, the majority of R^2 of the linear regression should be larger than 0.8.

5.3 Alpha distribution after removing the outlier

To reduce the impact of extreme volatile price movement and have a better linear regression fit, the same process is repeated except the top 10 extreme log returns are removed from the regression. The result can be found in Fig5. From it we can observe that the exponents are mainly centralized around 2.5 to 3 for both positive and negative return. And a QQ-plot [3] is done to compare the distribution of alpha and normal distribution, we can see that in Fig6 the distribution of alpha is following normal distribution. Besides, for the distribution of R^2 in Fig7 shows that the majority is larger than 0.85, which indicates the regression fitted quite well on the tail log returns.

5.4 Exponents distribution with different Δt

The same approach in 5.1 is repeated for different Δt. The exponents can be found in Table1, and we can observe that the exponent is slightly less than 3 no matter the Δt is changed. It implies that the log return price follows the inverse cubic law in different time scale, but the smaller exponent indicates a flatter distribution, with more observations at the extreme values. Besides, the histograms of fitted alpha for individual cryptocurrency are shown in Fig8. As seen in Fig8, the distributions are alike for different Δt.

6 Compare with different periods

The data is then break down into 4 different period **1)** 1/1/2021 to 27/6/2021, **2)** 27/6/2021 to 21/12/2021, **3)** 21/12/2021 to 16/6/2022 and **4)** 16/6/2021 to 10/12/2022. The market in periods 1 & 2 are bull market, but 3 & 4 are under a bear market due to some black swan events happened like the top 10 decentralized network went down and second large exchange collapsed. The same approach in 5.1 is repeated for different period. And from the result in Fig 9, we can see that the CDF of both positive and negative log price returns for different period are very similar on the tail. It implies that the log prices follow power law in different period.

7 Compare to developed market

US stock market is a well-developed market with well-established regulatory framework and market liquidity. Yet, on the other hand, cryptocurrency market is still a developing market with volatile price. We want to explore the difference of market maturity on log price return distribution. To compare it, we have downloaded the S&P 500 daily return from Yahoo Finance from 1-1-2022 to 10-12-2022. Then, the same approach in 5.1 is repeated on daily basis. As seen in the Fig 10, the distributions for both markets are similar. Yet, the crypto market has a longer tail than the S&P 500 for both positive and negative return. This indicates that the prices in crypto market contain a high degree of price fluctuation. Power law distributions are often used to describe the distribution of rare events, such as market crashes or other extreme events. Hence, the longer and fatter tail in distribution implies that the extreme events are more often than traditional market.

8 Conclusion

Our analysis has shown that both the stock and cryptocurrency markets follow an inverse cubic law, indicating the presence of underlying patterns in both markets. Additionally, our analysis indicates that the cryptocurrency market has a fatter tail in the cumulative distribution function [3](CDF) and the alpha is slightly smaller than the stock market, suggesting a higher likelihood of extreme events in crypto market. Further research can be done in the future while the cryptocurrency market is developing.

Appendix

$$R_i(t, \Delta t) = \ln\left(\frac{P_i(t+\Delta t)}{P_i(t)}\right) \qquad (1)$$

R_i is the log return, $P_i(t)$ is the close price of the instrument *i* at time *t*.

$$\mathrm{r_i(t, \Delta t)} = \frac{\mathrm{R_i(t,\Delta t)} - \mu}{\sigma} \qquad (2)$$

r_i is the normalized log return, μ is the mean of the log return data, and σ is the standard deviation of the log return data.

$$\widehat{\alpha} = \frac{k}{\sum_{i=1}^{k}\left(\frac{r_i}{r_{k+1}}\right)^{-1}} \qquad (3)$$

$\widehat{\alpha}$ is the estimated value of the power law exponent, k is the number of top k observations in the dataset, r_i is the ith largest return. This equation estimates the value of the power law exponent by taking the ratio of the number of observations to the sum of the inverse powers of the observations. The Hill estimator is often used in the analysis of financial data because it is relatively simple to implement and can provide a good approximation of the power law exponent in many cases.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (4)$$

R^2 is the R squared value, y_i is the observed value of the dependent variable for the ith observation, $\widehat{y_i}$ is the predicted value of the dependent variable for the ith observation, $\bar{y}$ is the mean of the observed values of the dependent variable, and n is the number of observations in the dataset. This equation calculates the R squared value as the ratio of the sum of the squared differences between the observed and predicted values to the sum of the squared differences between the observed values and the mean of the observed values. A higher R squared value indicates a better fit of the model to the data.

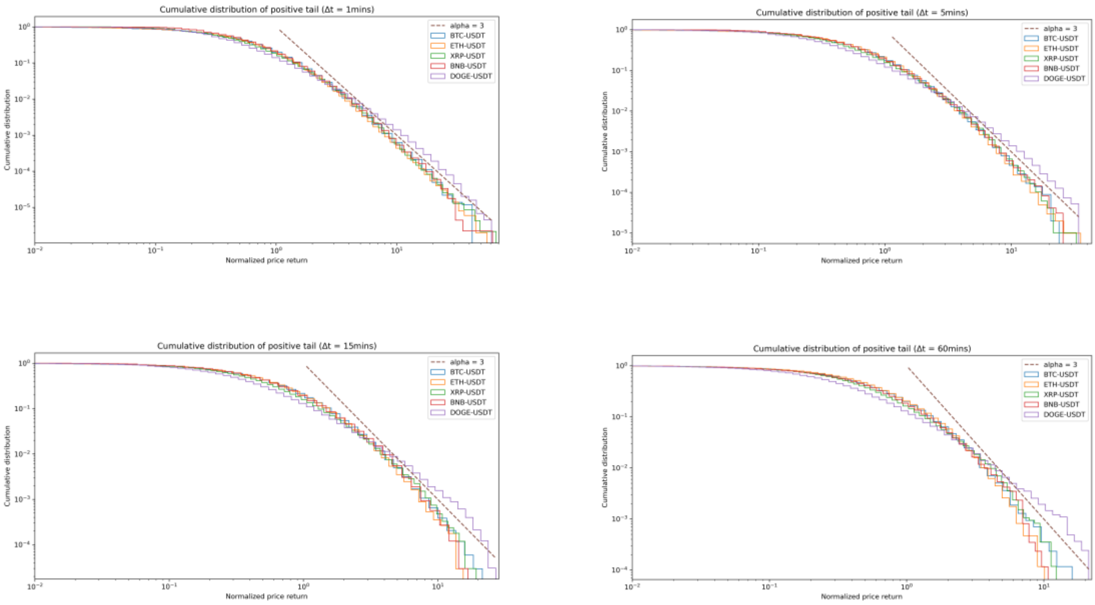
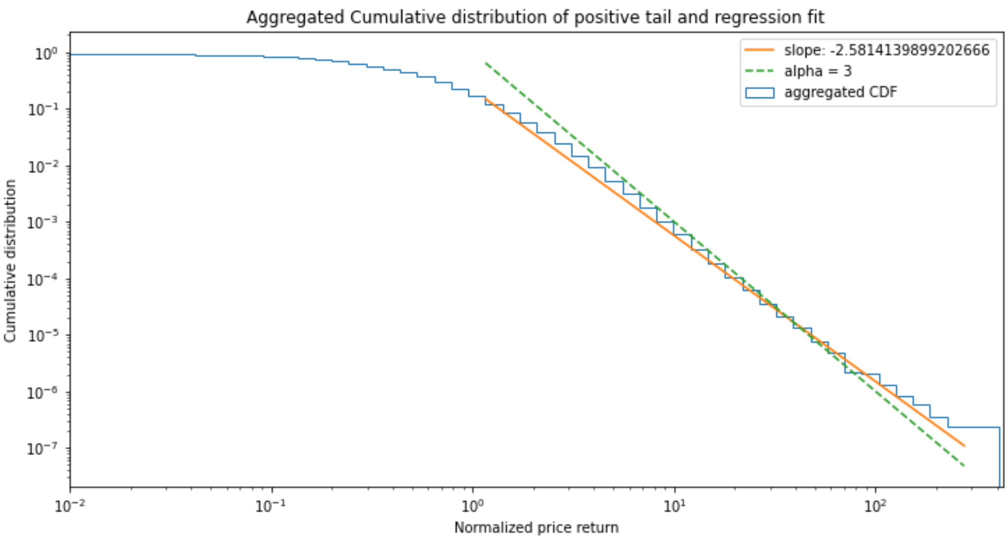


Fig. 1. CDF with different Δt. Top left: 1 min; Top right: 5 min; Bottom left: 15 min; Bottom right 60 min



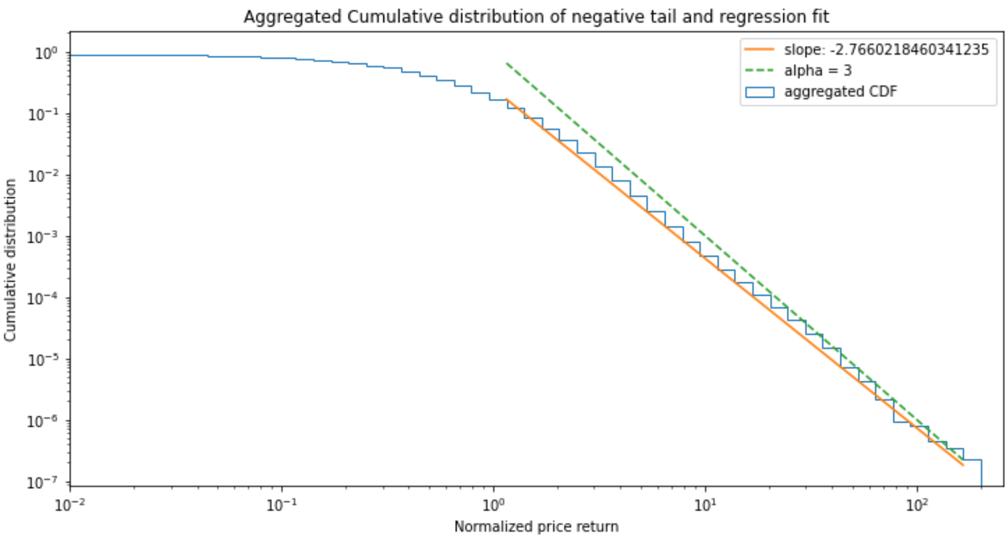


Fig. 2. Aggregated Cumulative distribution of the regression fit. Top: positive tail; Bottom negative tail

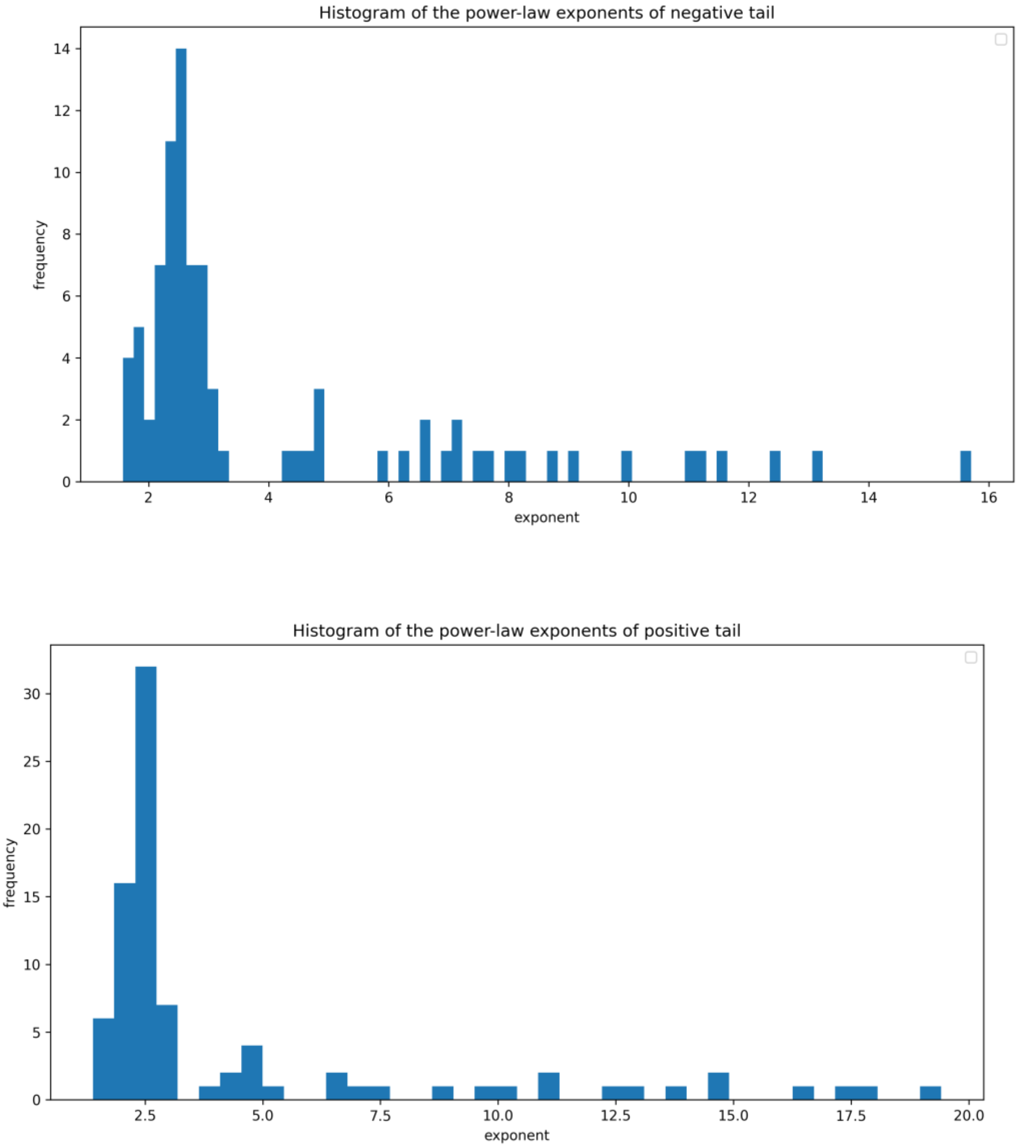


Fig. 3. Histogram of the power-law exponents before remove outlier. Top: positive tail; Bottom negative tail

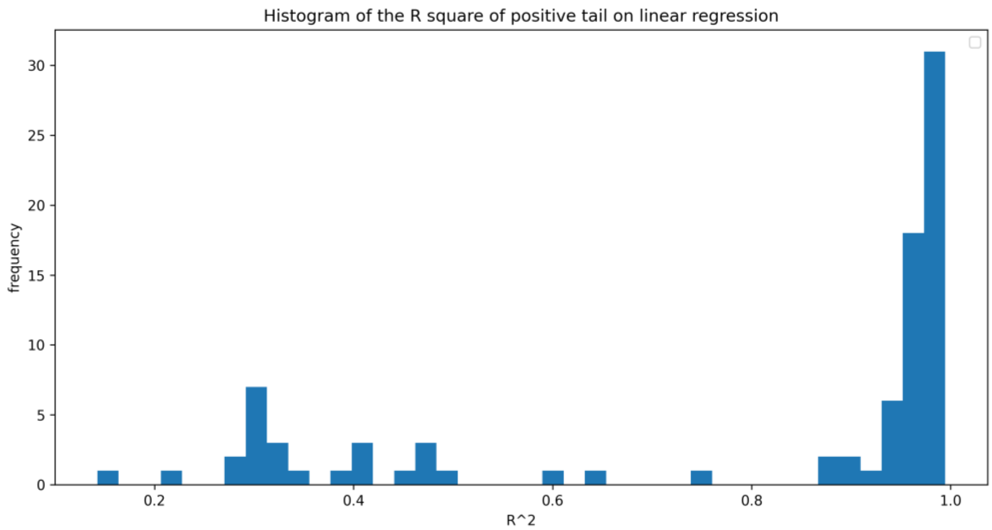
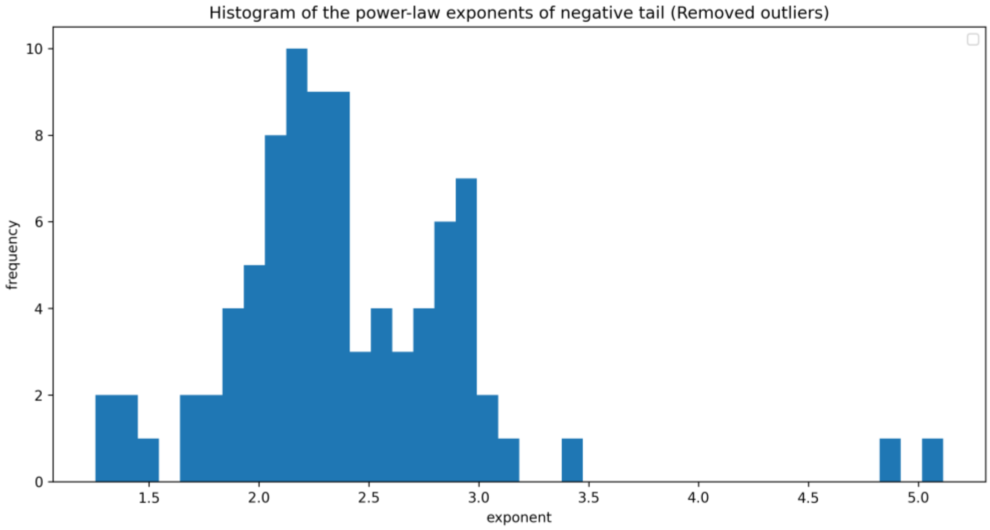


Fig. 4. Histogram of the R square of positive tail on linear regression



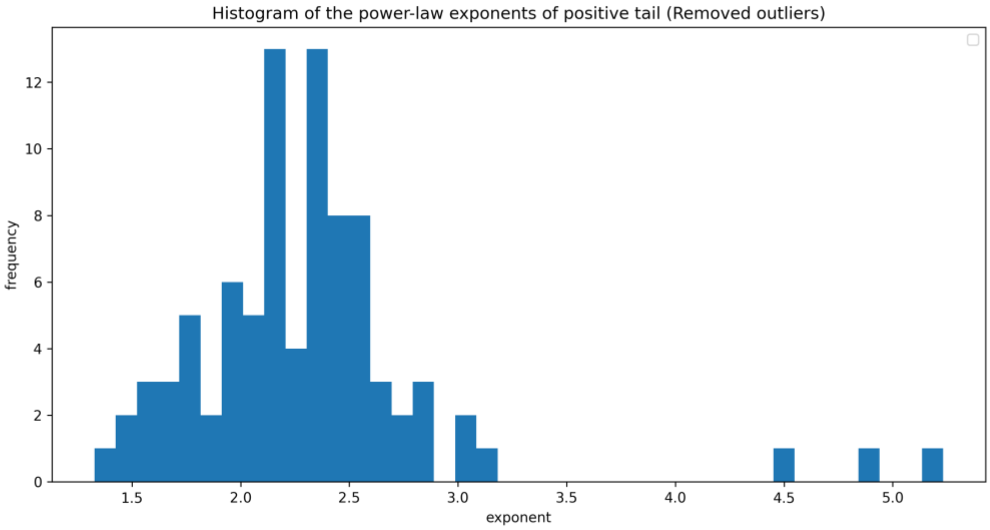
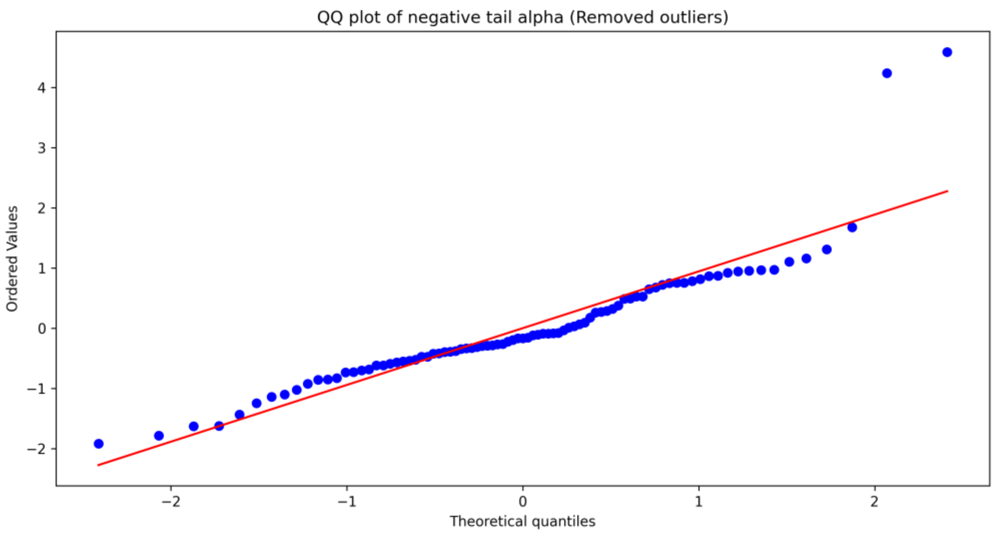


Fig. 5. Histogram of the power-law exponents after remove outlier. Top: positive tail; Bottom negative tail



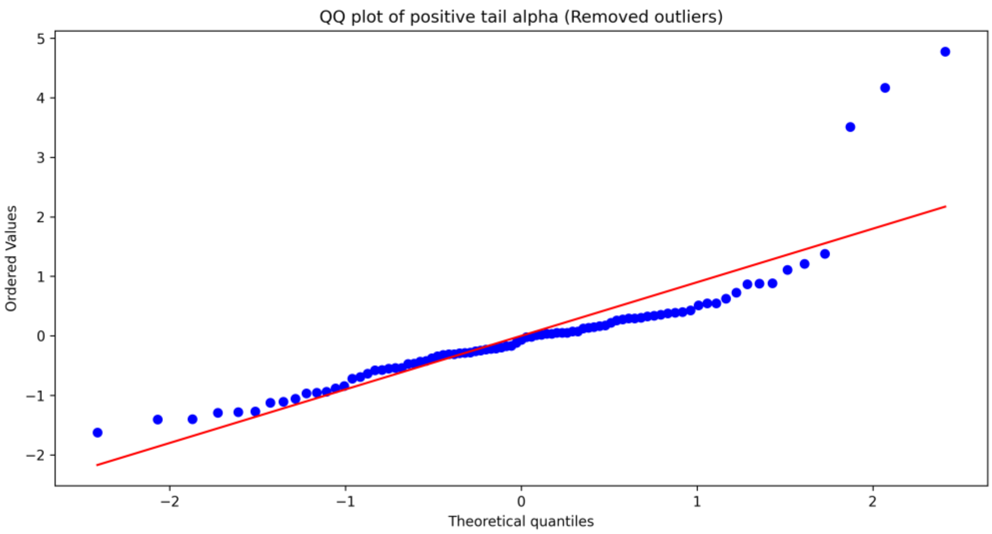


Fig. 6. QQ plot of alpha (Removed outliers) Top: positive tail; Bottom negative tail

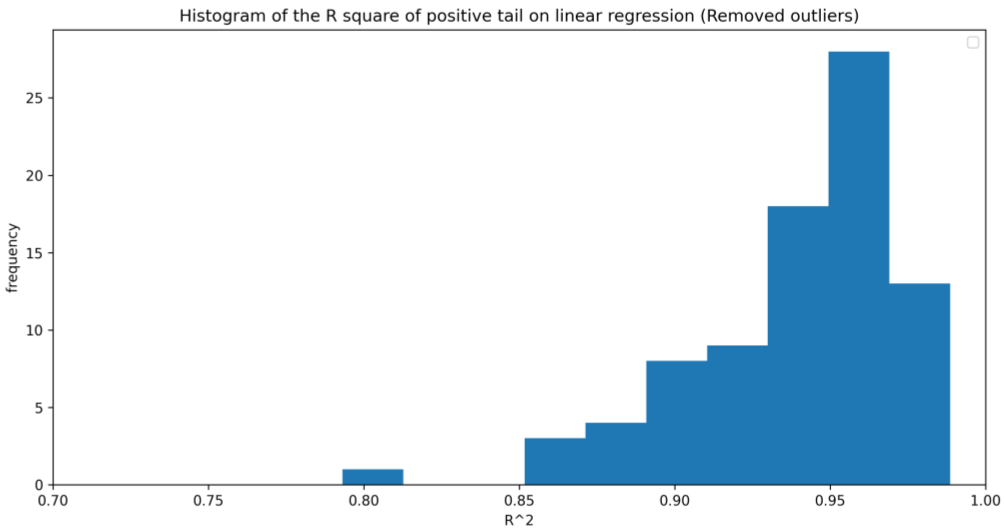


Fig. 7. Histogram of the R square of positive tail on linear regression (Removed outliers)

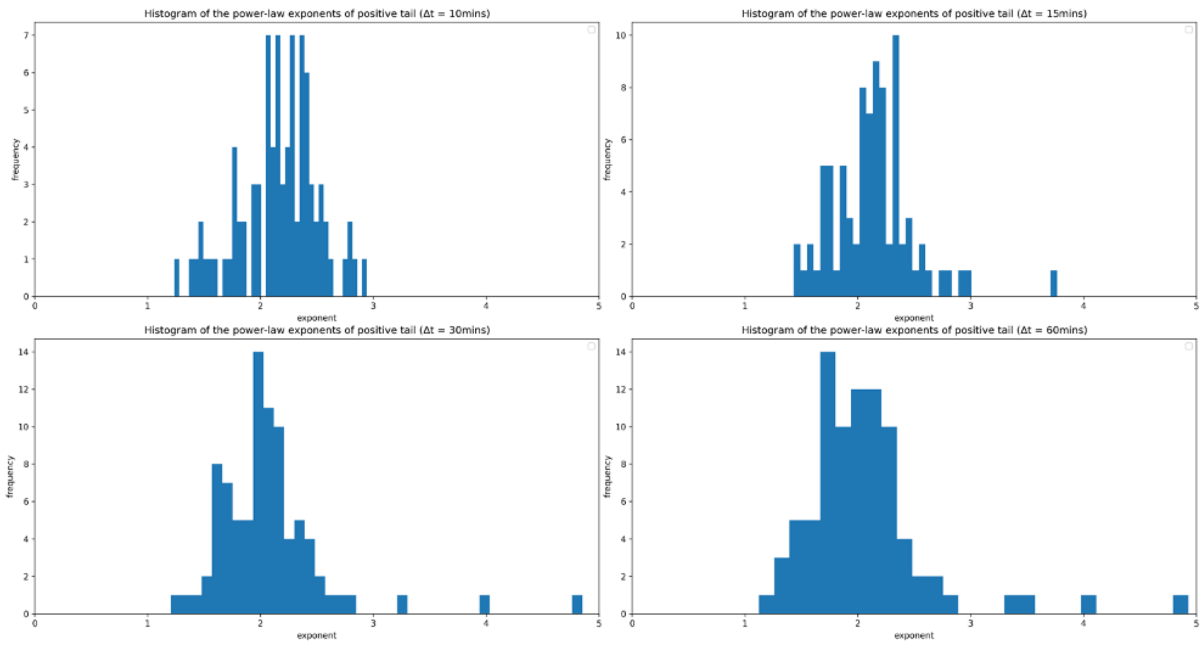


Fig. 8. Fitted alpha distribution with different Δt. Top left: 10min; Top right: 15 min; Bottom left: 30min; Bottom right 60min

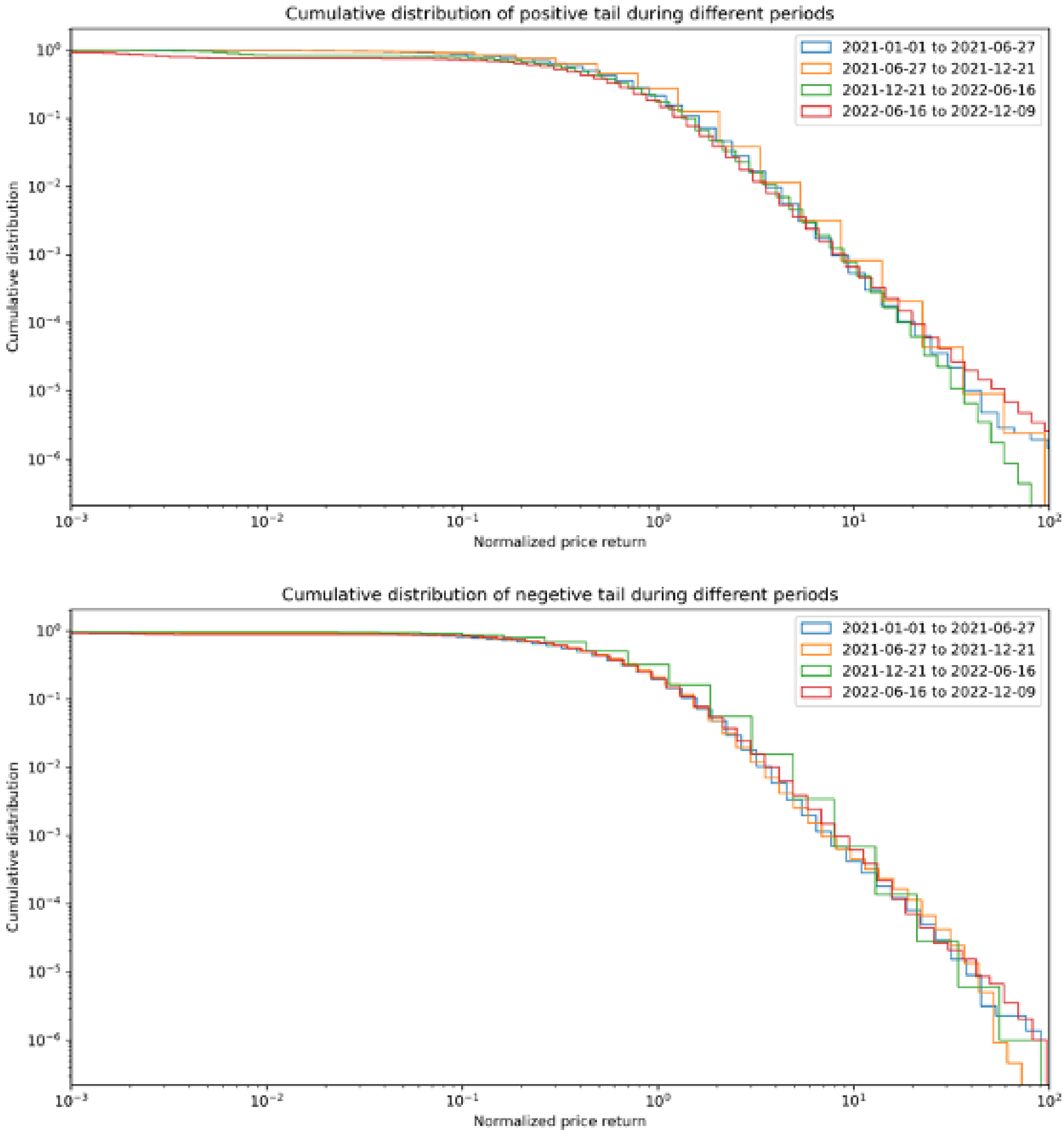


Fig. 9. Cumulative distribution of during different periods Top: positive tail; Bottom negative tail

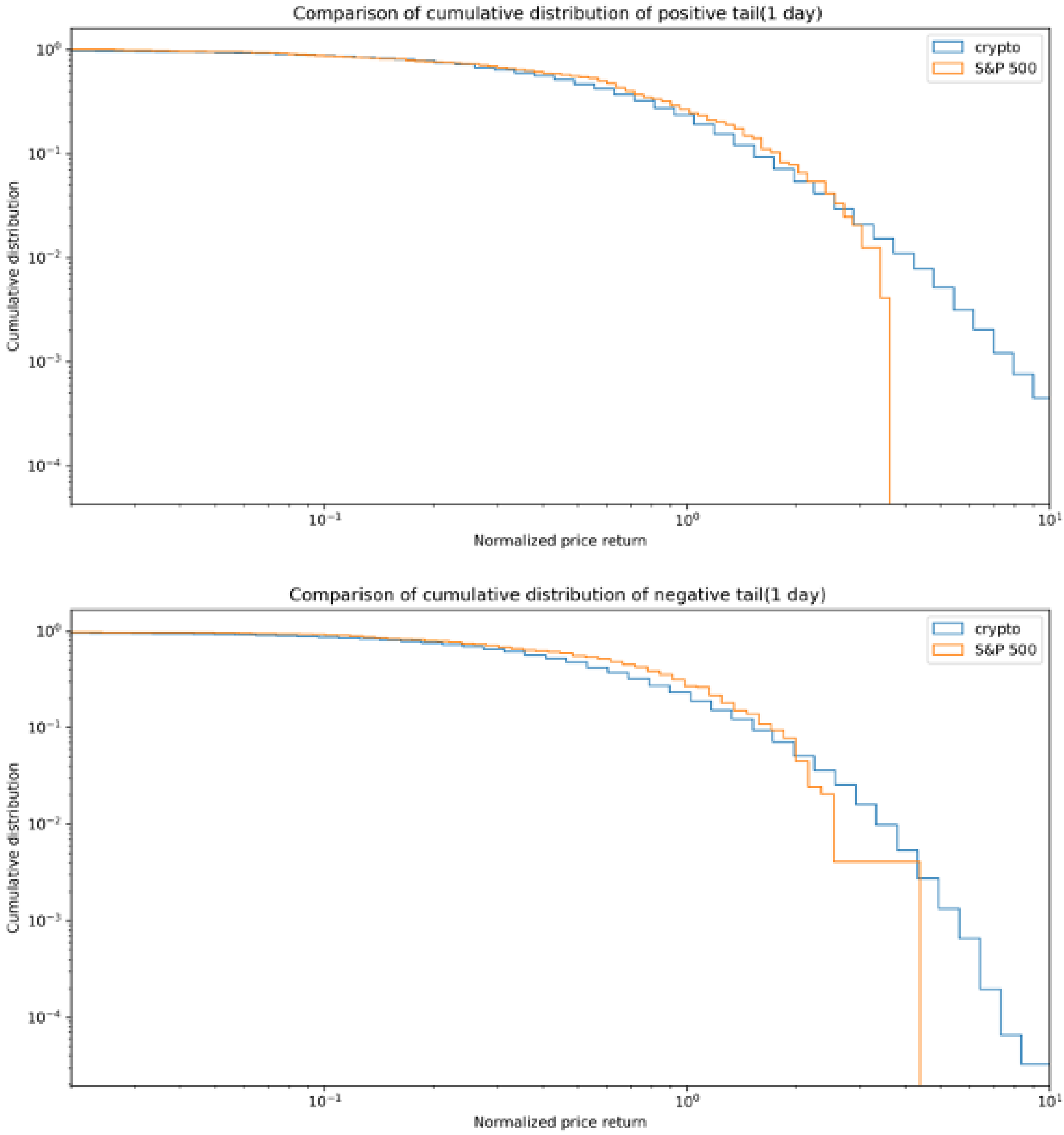


Fig. 10. Comparison of cumulative distribution with S&P 500 Top: positive tail; Bottom negative tail

	Δt = 5	**Δt = 10**	**Δt = 15**	**Δt = 30**	**Δt = 60**
Positive	2.5814	2.5824	2.6460	2.7045	2.8606
Negative	2.7660	2.7244	2.9116	2.8216	2.8413

Table 1. Fitted α in tail distribution

References

[1] J. L. Devore, Probability and Statistics for Engineering and the Sciences, 2011.
[2] A. A. S.Sinha, Econophysics : An Introduction, 2010.
[3] M. B. W. a. R. Gnanadesikan, Probability Plotting Methods for the Analysis of Data, 1968.