ORIE 5640 Statistics for Financial Engineering

Project 1

Cryptocurrency Return Analysis and Distribution Modeling

Group members:
Rui Dai (rd576)
Shaohan Wang (sw2265)
Xuanyu Ding (xd82)

March 21, 2021

1 Introduction

In recent years, a new type of currency, cryptocurrency, is attracting a significant amount of attention around the world, including the traditional financial industry. Trading volumes of cryptocurrency grows at a fast pace, and therefore, understanding the properties and distribution of cryptocurrency returns becomes increasingly important. Inspired by literature on this topic, this project conducts exploratory data analysis on the first and most popular cryptocurrency, Bitcoin (BTC), and tries to model the distribution of its log returns. In addition, we also did supplementary analysis on Ethereum (ETH) and Ripple (XRP) to check if our observations are generalizable.

2 Exploratory Data Analysis

The cryptocurrency prices in USD are acquired from the quantmod package in R with Tiingo source. The time frame is from Jan 1, 2017 to Feb 11, 2021.
The log returns are calculated from the closing prices of each trading date. In this section, our team tried to obtain summary statistics and diagnostic plots to help us better evaluate and test the fitted distributions later.

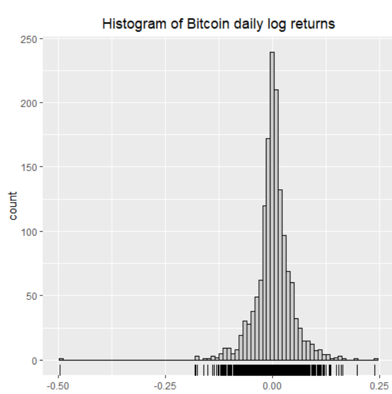
2.1 Summary Statistics

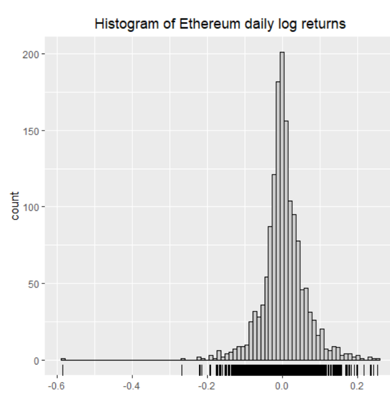
The quantiles and standard deviation of the log returns over the whole period are listed below. Since all of the average log returns are above 0, we could conclude that the prices of these three cryptocurrencies have been increasing over recent years. However, even with the overall growing pattern, we cannot ignore the high volatility and the large amount of loss during the “bad time”.

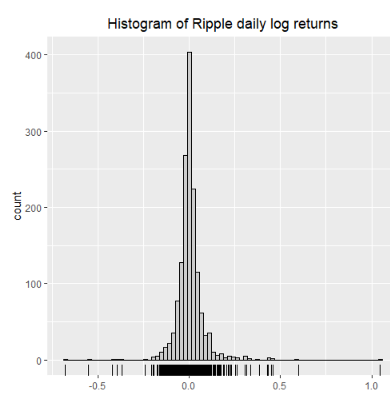
Cryptocurrency	Min	1st Quantile	Mean	3rd Quantile	Max	Std. Dev.
BTC	-49.68%	-1.50%	0.26%	2.20%	23.84%	4.35%
ETH	-58.52%	-2.15%	0.36%	2.92%	25.27%	5.79%
XRP	-68.04%	-2.45%	0.30%	2.41%	104.61%	7.83%

2.2 Histogram of Log Returns

Histogram is one of the most common methods in simulating density plots, and we could at least have some opinions about the shape of the density plots by observing the histograms of the cryptocurrencies.
At first glimpse, the plots are fairly similar to normal distributions. However, there are many extreme outliers at the left hand side, which indicates the possibility of left skewness.

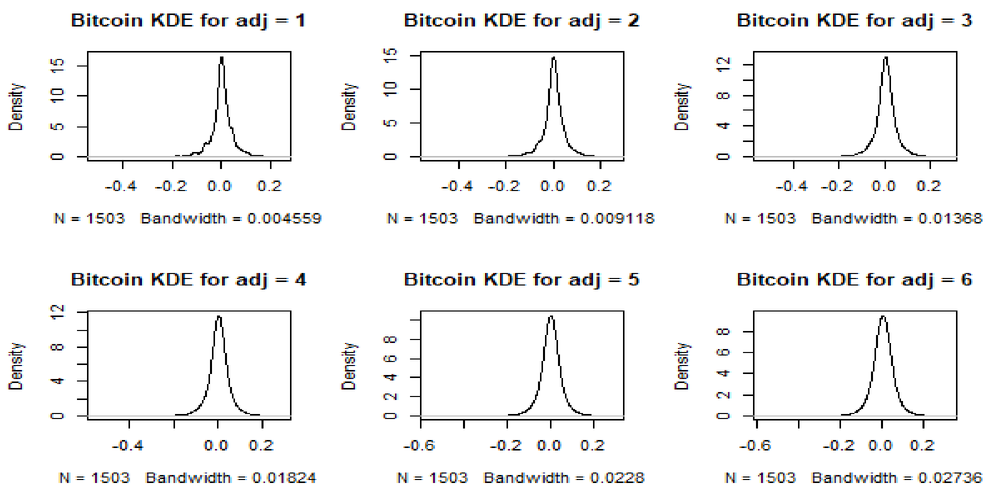






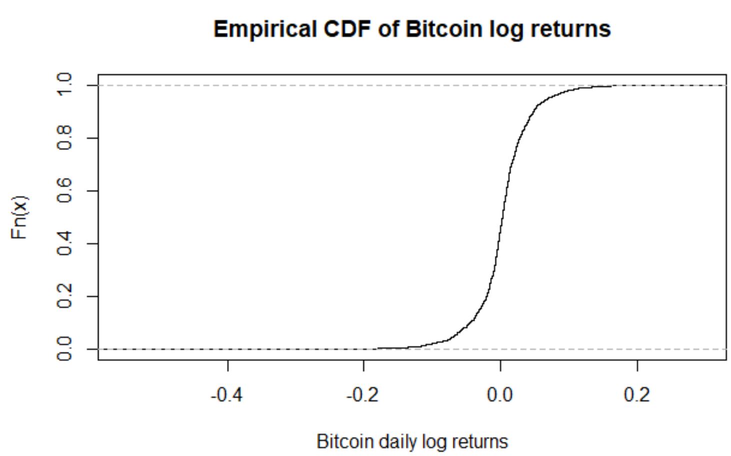
2.3 Kernel Density Estimation

Another way to estimate the probability density function is to use kernel density estimation. Here, we use the "Sheather-Jones" method to automatically select the bandwidth. The choice of the bandwidth might imply that the size of the confidence band is larger than the optimal size because we are having large variance in order to eliminate bias [1].
According to the Bitcoin log returns Kernel density plots, clearly the best choices are probably when adjusted value equal to 4 or 5.



2.4 Empirical CDF

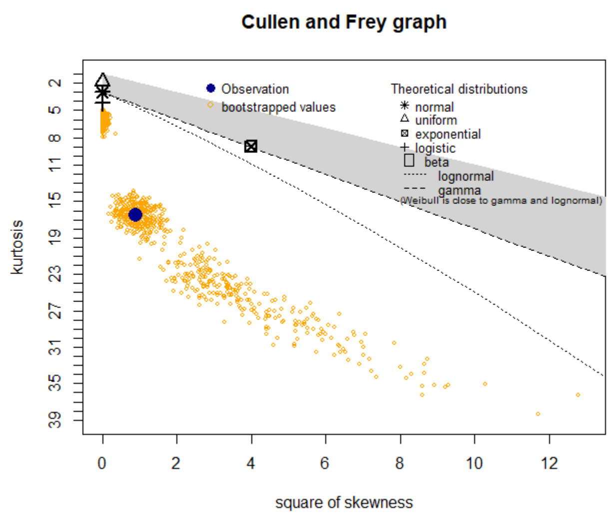
We could also estimate the actual cumulative density functions by applying the empirical CDF method. Empirical CDF will give us the cumulative probability with respect to each order statistics or the respective p-th sample quantile. Empirical CDF will help to better understand the distribution of our sample log returns and might be utilized to select the corresponding distribution models.
Here, according to Bitcoin log returns empirical CDF, the log returns might indicate the left skewness because of the negative extreme values, which is the evidence of showing normal distribution might not be a good fit to the cryptocurrency's log returns.



3 Modeling Return Distribution

3.1 Candidate Distribution

Although there is limited literature on modeling the return distribution of cryptocurrency, it is commonly agreed that the returns have heavy tails, as common financial data. From the histograms, we found that the log returns in the cases of all three cryptocurrencies show significant deviation from the normal distribution. In addition to EDA from previous section, values of skewness and kurtosis were computed on bootstrap samples, and we visualized their similarities and differences with several classical distributions.
Take Bitcoin as an example:



According to the kurtosis and skewness of Bitcoin data, five most popular parametric distributions are selected for further analysis: **normal distribution, Laplace distribution, Cauchy distribution, Student's t-distribution and the generalized hyperbolic distribution**.
The maximum likelihood method was used to fit each distribution, and this fitting method is defaulted in fitdist() in R. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to discriminate between and determine the best fitting distribution.
In addition, various other measures could be used to discriminate between non-nested models. These could include:

- Kolmogorov-Smirnov statistic [2] defined by

$$\mathrm{KS} = \sup_x \left| \frac{1}{n} \sum_{i=1}^{n} I\{x_i \leq x\} - \hat{F}(x) \right|$$

  where $I\{\}$ denotes the indicator function and $\hat{F}(\cdot)$ is the maximum likelihood estimate of $F(x)$.

- Anderson-Darling statistic [3] defined by

$$\mathrm{AD} = -n - \sum_{i=1}^{n} \left\{ \ln \hat{F}\left(x_{(i)}\right) + \ln \left[1 - \hat{F}\left(x_{(n+1-i)}\right)\right] \right\}$$

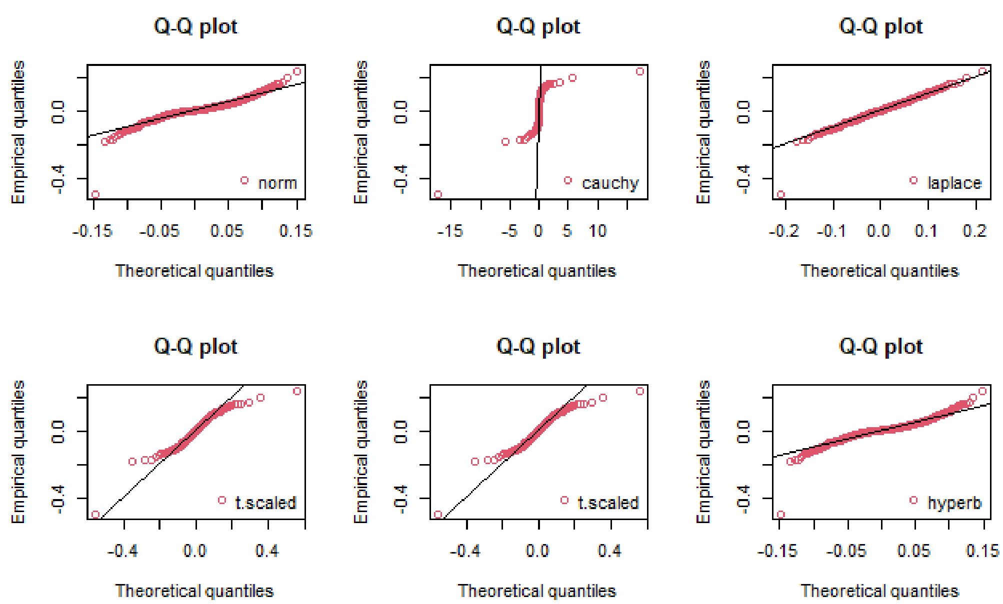
where $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ are the observed data arranged in increasing order.
Once again, the smaller the values of these statistics, the better the fit.

3.2 Application on Cryptocurrencies

Fitted distributions and results for daily log returns of Bitcoin from 1 Jan. 2017 until 12 Feb. 2021:

	Dist.	Log.L	AIC	BIC	A.D.stat.	K.S.stat.
1	Normal	2628.664	-5253.328	-5242.660	28.548632	0.10004232
2	Cauchy	2743.046	-5482.092	-5471.425	6.006283	0.03846342
3	Laplace	2834.689	-5665.378	-5654.710	1.282470	0.03259307
4	Std-t(2)	2823.498	-5640.997	-5624.996	1.627756	0.02799477
5	Std-t(7)	2823.498	-5640.997	-5624.996	1.627110	0.02794918
6	GH	2638.083	-5270.166	-5254.165	27.677831	0.09744663

See appendix for results for Ethereum and Ripple.
We found that Laplace distribution gives the best fit for Bitcoin and Ethereum, and the t distribution gives the best fit for Ripple. Q-Q plot for daily log returns of Bitcoin from 1 Jan 2017 until 12 Feb 2021 on six testing distributions:



See appendix for Q-Q Plots for Ethereum and Ripple.

4 Conclusion

Traditional financial instruments usually exhibit heavy tails [4], but we found a slightly suprising results from both EDA and distribution modeling. We found that Bitcoin whereas Ethereum have semi heavy tails and Ripple have heavy tails, thus a light tailed distribution (Laplace distribution) gives the best fit for the former two and a heavy tailed distribution (t distribution) for the latter. From the Q-Q plots for Bitcoin fitting, the normal distribution did not capture the heavy tail on both ends, the Cauchy, t, and generalized hyperbolic distribution have tails that are exceedingly heavy. The Laplace distributions captured the data well but failed to fit the left end which is likely to be an outlier. In summary, the tail behaviour of cryptocurrency returns does not follow that of traditional financial instruments and the returns generally have lighter tails.

References

[1] Y. Chen, "A tutorial on kernel density estimation and recent advances," *Biostatistics and Epidemiology*, 2017.

[2] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Giornale dell'Istituto Italiano*, 1933.

[3] T. Anderson and T. Darling, "A test of goodness of fit," *Journal of the American Statistical Association*, 1954.

[4] S. Chan, J. Chu, S. Nadarajah, and J. Osterrieder, "A statistical analysis of cryptocurrencies," *Journal of Risk and Financial Management*, 2017.

5 Appendix

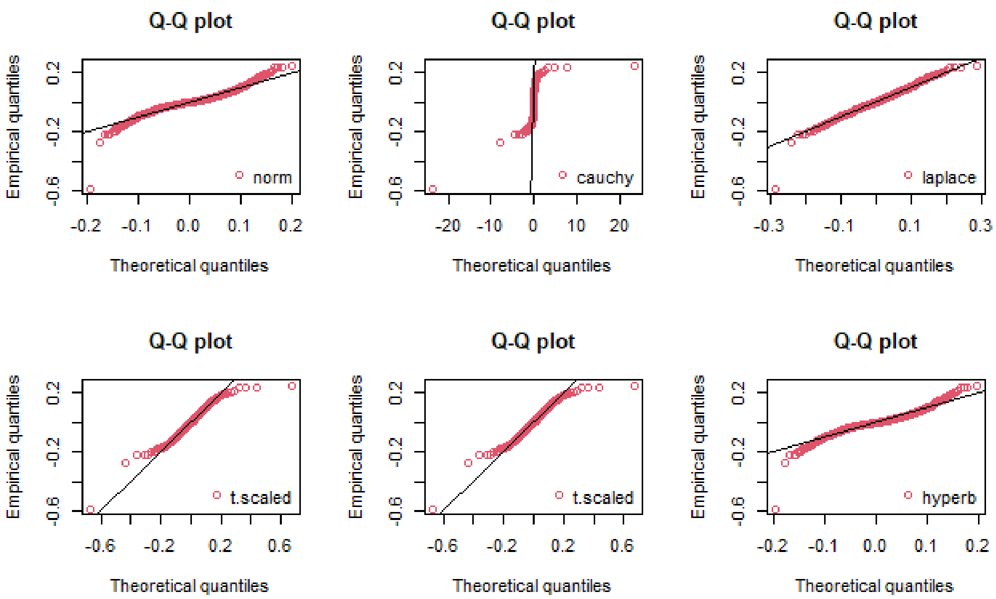
5.1 Distribution Fitting Results for Ethereum

	Dist.	Log.L	AIC	BIC	A.D.stat.	K.S.stat.
1	Normal	2184.344	-4364.688	-4354.029	25.145798	0.08768005
2	Cauchy	2263.139	-4522.278	-4511.618	7.933596	0.04264559
3	Laplace	2364.334	-4724.667	-4714.008	1.627210	0.02934265
4	Std-t(2)	2353.229	-4700.458	-4684.469	1.872362	0.02478061
5	Std-t(7)	2353.229	-4700.458	-4684.469	1.870781	0.02477683
6	GH	2192.484	-4378.968	-4362.979	24.615600	0.08482858

5.2 Distribution Fitting Results for Ripple

	Dist.	Log.L	AIC	BIC	A.D.stat.	K.S.stat.
1	Normal	1727.313	-3450.627	-3439.967	25.145798	0.08768005
2	Cauchy	2209.495	-4414.989	-4404.330	7.933596	0.04264559
3	Laplace	2195.228	-4386.455	-4375.796	1.627210	0.02934265
4	Std-t(2)	2261.221	-4516.442	-4500.453	1.872362	0.02478061
5	Std-t(7)	2261.221	-4516.442	-4500.453	1.870781	0.02477683
6	GH	2192.484	-4378.968	-4362.979	24.615600	0.08482858

5.3 Q-Q plots for Ethereum



5.4 Q-Q plots for Ripple