Report: Statistical Analysis of cryptocurrencies

Mouktar ABDILLAHI

1 Introduction

The purpose of this paper is to analyze statistical properties of some well-known cryptocurrencies such as Bitcoin Cash and Ethereum Classic. First, we do a Time Series Analysis on these 9 cryptocurrencies then we characterize their exchange rates versus the U.S. Dollar by fitting parametric distributions to them. We will see that returns are mostly non-normal, however, no single distribution fits well jointly to all the cryptocurrencies analysed. Finally, we will analyze these 9 cryptocurrencies as a portfolio using the Modern Portfolio Theory to try and find the best portfolio in terms of return and risk. The results are important for investment and risk management purposes.

2 Data

In this paper, the datasets used are the historical global price indices of nine different cryptocurrencies. For our analysis, we had to use hourly data from 2nd February 2018, 11 AM until 14th June 2018, 7 PM. However, we noticed that we had missing data for two hours in the different datasets and that for the Ontology cryptocurrency (ONT) data was not available for the first two and a half months. This was not a problem initially as we were analyzing each cryptocurrency separately in the first parts of this study, however, action had to be done when treating these cryptocurrencies as a portfolio in the last part. The nine cryptocurrencies that are part of our analysis are: Bitcoin Cash(BCH), Ethereum Classic (ETC), Dash, Ontology(ONT), Icon (ICX), Stellar (XLM), Tether (USDT), Cardano(ADA) and QTUM. It should be noted that we have here cryptocurrencies of different sizes and backgrounds. For example, we have Bitcoin Cash (BCH) that "is a cryptocurrency created in August 2017, arising from a fork of Bitcoin Classic. Bitcoin Cash increases the size of blocks, allowing more transactions", as stated in its Investopedia definition [1]. So BCH is a cryptocurrency with a high price today because of its particular relation with the well-known Bictoin. On the other side, we have more recent cryptocurrencies such as ADA who have been launched in September of 2017. We can also give the example of USDT whose "cryptocoins in circulation are backed by an equivalent amount of traditional fiat currencies, like the dollar, the euro or the Japanese yen, which are held in a designated bank account" and hence we noticed that the price slowly fluctuates around 1.

We have plotted the prices of the different cryptocurrencies in the same plot using a log scale on the y-axis.

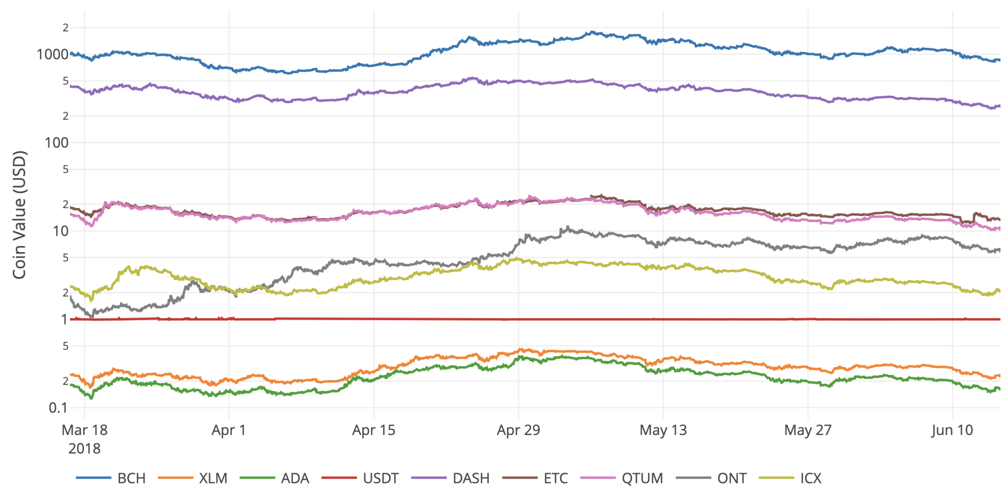


Figure 1: Cryptocurrency Prices

3 Time Series Analysis

Starting from here, we will use BCH to explain our approach and the different steps we went through. However, the results and the findings for the other cryptocurrencies will also be presented quickly.

For the first part of the analysis, we have done a time series analysis of our dataset. One of the important characteristics of time series is stationarity. For a stationary process, parameters such as mean and variance, if they are present, also do not change over time. Since stationarity is an assumption underlying many statistical procedures used in time series analysis, it is very important to spot if it is the case here. We proceeded to a decomposition of our time series in trend, seasonality and residual, in order to determine visually if the time series was stationary and we also did a Dickey-Fuller, which is one of the statistical tests for checking stationarity. Here the null hypothesis is that the Time Series is non-stationary. Here, the results of this test show that we can assume that the time series is not stationary as the 'Test Statistic' is higher than the critical values up to a certain confidence interval.

```
Results of Dickey-Fuller Test:
Test Statistic                  -2.722119
p-value                          0.070288
#Lags Used                      28.000000
Number of Observations Used   3891.000000
Critical Value (1%)             -3.432032
Critical Value (5%)             -2.862283
Critical Value (10%)            -2.567166
dtype: float64
```

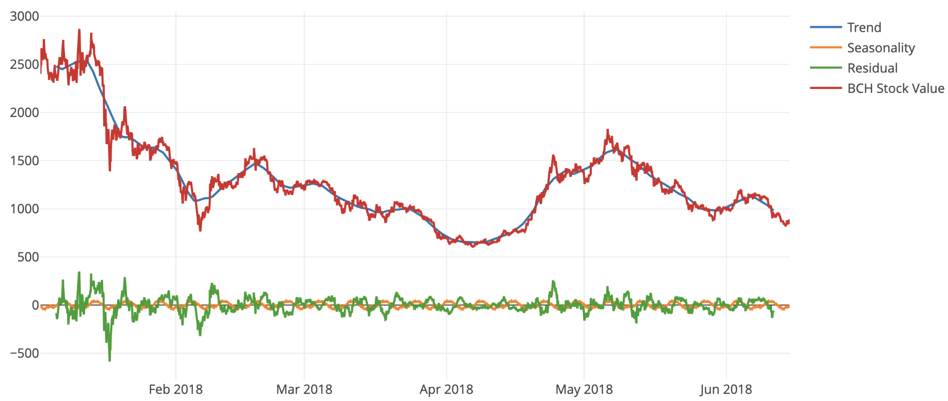


Figure 2: Trend-Seasonality-Residual Decomposition (BCH) and Dickey Fuller Test

In a time series forecast problem, we would think of techniques that can be used to take this Time Series towards stationarity and compare different models for forecasting the time series as that could be really useful for cryptocurrency trading. However, in the context of this assignement, we will not insist on this part and focus on estimating as good as possible the kind of probability distribution and the associated parameters each time series is derived from.

4 Distributions fitted

Before fitting any distributions, summary statistics of the exchange rates of the nine cryptocurrencies are given in the following table.

	BCH	**XLM**	**ADA**	**USDT**	**DASH**	**ETC**	**QTUM**	**ONT**	**ICX**
Mean	1289	0.365	0.344	1.004	534.88	22.728	24.1	5.477	4.314
Standard Deviation	472.9	0.13	0.214	0.018	226.2	7.72	11.93	2.561	2.283
Skew	1.231	1.082	1.819	4.617	1.176	0.68	1.295	-0.202	1.42
Kurtosis	1.369	1.048	3.104	24.14	0.735	-0.75	0.93	-1.18	1.151
Min	605.7	0.165	0.125	0.958	243.8	12.21	10.15	1.05	1.630
Quantile 25%	986.1	0.278	0.203	0.999	364.7	16.08	15.11	3.55	2.68
Median	1209	0.345	0.283	1.0	472.4	20.44	19.875	6.11	3.725
Quantile 75%	1486	0.432	0.371	1.0	627.6	29.473	28.813	7.58	4.59
Max	2857	0.918	1.34	1.13	1239	46.73	73.78	11.52	12.48

In our initial dataset, we only had 4000 rows and most of the time less than that. Hence, we have decided to work with the hourly log-returns instead of the daily-log returns or any other timescale and for the rest of this paper, you can assume that the dataset, we are working on is composed of the hourly log-returns for each cryptocurrency.

4.1 Methodology

[2]We have decided to try to fit six different distributions to each time series. They are specified as follows: the Normal Distribution, the Student-t Distribution, the Skew-t Distribution, the Laplace distribution, the Lognormal Distribution and the Normal Inverse Gaussian.

All but one of the distributions (the Laplace distribution and the Normal Distribution) are heavy tailed and as heavy tails are common in financial data, it was obvious for us to fit these particular distributions to our time series.

In order to do so, we have decided to do a Time Series cross-validation on our datasets in order to find the optimal parameters for each distribution fitting. A Time Series cross-validation is organized as explained in Figure 3.

To decide between two sets of parameters for a given distribution, we have decided to use the Kolmogorov-Smirnov's test Statistic. Indeed, given that the test doesn't reject the hypothesis that the distributions are the same, we can compare the Statistic to decide of the best parameters for each train and test set as defined above.

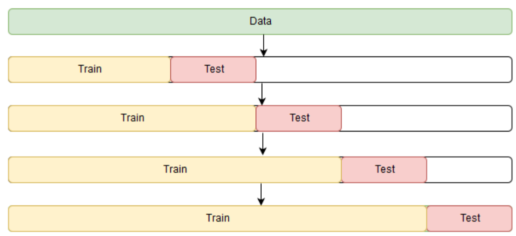


Figure 3: Time Series Nested Cross Validation

The next step is to decide of the best fitting distributions for each cryptocurrency. Discrimination among them was performed using various criteria: the Akaike Information Criteria (AIC), the Bayesian information criterion (BIC), the Constant Akaike Information Criteria (CAIC), the corrected Akaike information criterion (AICc) and the the Hannan-Quinn criterion (HQC). The five discrimination criteria use the Maximum Likelihood Estimate (MLE). In all cases, the smaller the values of the criteria the better the fit. Apart from the five criteria, various other measures could be used to discriminate between distributions.

4.2 Results

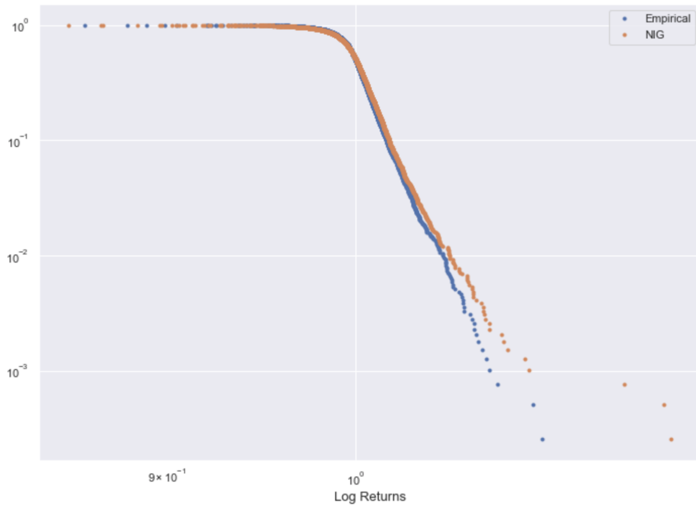
We have produced the following table (BCH) for every cryptocurrency in order to compare these information criteria and decide of the best fitting distribution.

Criteria	**Normal**	**Student-t**	**Skew-t**	**Laplace**	**LogNormal**	**NIG**
AIC	-20905	-21810	-19640	-21950	-20911	-21990
BIC	-20893	-21791	-19615	-21938	-20892	-21965
CAIC	-20891	-21788	-19611	-21936	-20889	-21961
AICc	-20905	-21810	-19640	-21950	-20911	-21990
HQC	-20901	-21803	-19631	-21946	-20904	-21981

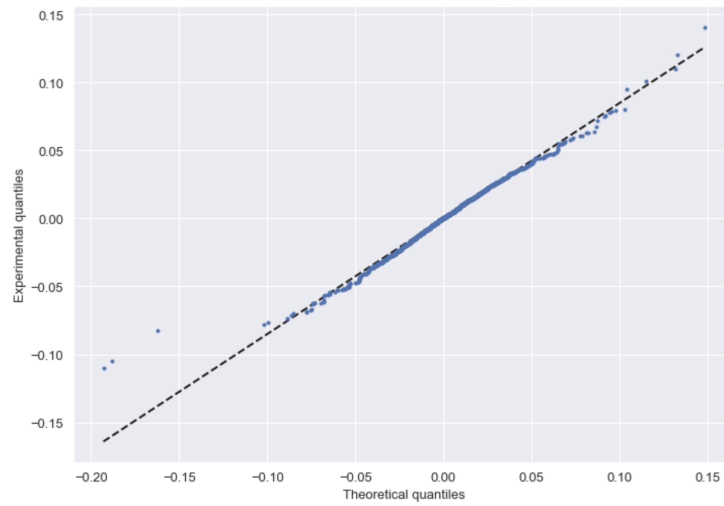
Cryptocurrency	**Best Fitting Distribution**	**Parameters Estimates**
BCH	Normal Inverse Gaussian	$\mu = -9.45333e^{-5}, \alpha = 0.39155, \beta = -0.0118, \delta = 0.01192$
XLM	Normal Inverse Gaussian	$\mu = -0.00055, \alpha = 0.35518, \beta = 0.00958, \delta = 0.01419$
ADA	Normal Inverse Gaussian	$\mu = -0.00047, \alpha = 0.51840, \beta = 0.0035, \delta = 0.01653$
USDT	Skew-t	$\mu = 7.4342e^{-19}, \sigma = 7.4342e^{-19}, \lambda = 0.80567, q = 2.30191$
DASH	Normal Inverse Gaussian	$\mu = 0.000198, \alpha = 0.41438, \beta = -0.02154, \delta = 0.00998$
ETC	Laplace	$\mu = -2.98968e^{-8}, b = 0.0155$
QTUM	Normal Inverse Gaussian	$\mu = -0.00048, \alpha = 0.40014, \beta = 0.00305, \delta = 0.01373$
ONT	Student-t	$\nu = 3.15624, \mu = -4.20713, \sigma = 0.01931$
ICX	Normal Inverse Gaussian	$\mu = -0.000297, \alpha = 0.40025, \beta = 0.00288, \delta = 0.01709$

For most of these cryptocurrencies, the best fitting distribution is actually fitting quite well the log-returns. However, in the case of USDT, the best fitting distribution is Skew-t and is not good at all, I believe because of the specificities of this cryptocurrency.

The adequacy of the best fitting distributions is assessed visually in terms of Q-Q plots, CCDF log-scale plots. In order to illustrate this, here are the plots used to do so for Bitcoin Cash.



(a) CCDF (BCH - Normal Inverse Gaussian)



(b) Q-Q Plots (BCH - Normal Inverse Gaussian)

Goodness of fit: We tested it by using the one-sample Kolmogorov-Smirnov test and the one-sample Anderson-Darling test too. The p-values of the one-sample Kolmogorov-Smirnov test for the best fitting distributions concluded that all of the best fitting distributions are adequate while the critical values of the one-sample Anderson-Darling test led us to decline the hypothesis that BCH and ADA derived from their best fitting distribution. However, both tests agreed on USDT as the 'bizarre' structure of the cryptocurrency led to no distribution being able to fit its log-returns good enough.

In the following table, you will find the results of these tests for BCH for its best fitting distribution - NIG and the interpretation that followed.

	KS Statistic	**p-value**	**AD Statistic**	**Critical value**
BCH	0.07501	$4.669e^{-10}$	34.635	3.752

For the Kolmogorov-Smirnov test, the statistic statistic is the absolute max distance (supremum) between the CDFs of the two samples. The closer this number is to 0 the more likely it is that the two samples were drawn from the same distribution and as the p-value is lower than 0.05, we can validate that. For the Anderson-Darling test, the critical region where we reject the hyptothesis that the the two samples were drawn from the same distribution is where the statistic is higher than the critical value and that is the case here.

4.3 Tail Distribution

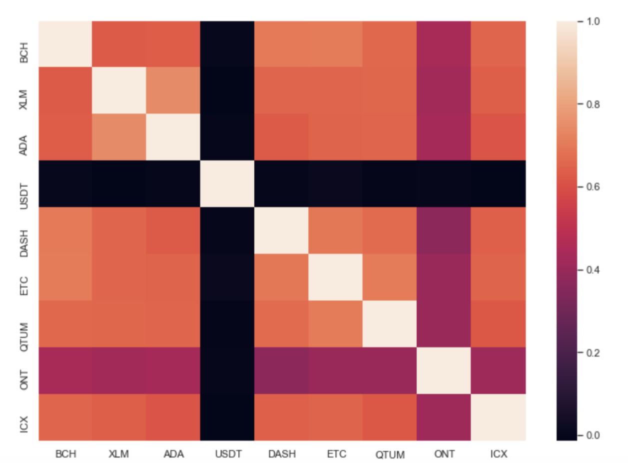
For this part of our study, we had to determine two values to characterize the best possible the tails of each cryptocurrency distribution. In order to determine the tail exponent and the adequate confidence interval, we have proceeded as we did in class for the case study 2 using a Bootstrap method. We also had to determine the threshold for the selection of the upper tail and the lower tail of each distribution. For this, we have followed the methodology of a research paper called 'The power-law tail exponent of income distributions' [3] that minimizes the Mean Squared Error of the Hills estimator for a series of thresholds. We summed up the results for this study only for Bitcoin Cash:

	Threshold	**Lower Bound**	**Tail Exponent**	**Upper Bound**
Right tail	0.08	2.102379	2.181561	2.274971
Left tail	0.11	-0.746635	-0.709862	-0.700828

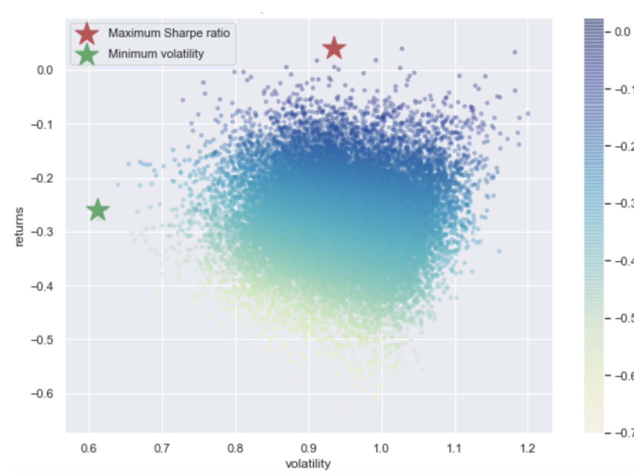
However, this method was not the best one to define the threshold in our case as for most of the cryptocurrencies, the higher the threshold and the lower the MSE, so testing something else to fit the tails of the distributions could be valuable.

5 Portfolio

For this last chapter, we have decided to study the nine cryptocurrencies not individually but as a portfolio. In order to do that, we have applied the principles of Modern Portfolio Theory to determine the weight to attribute to each cryptocurrency. We used the Maximum Sharpe Ratio to decide of the weights for the portfolio. Besides the correlation heatmap and a scatterplot of Returns and Volatility for simulated weights, in the following table, we have summarized the results of our study. The Portfolio 1 is composed of the nine datasets, in the Portfolio 2, we took out three cryptocurrencies that had very low prices (XLM, ADA and USDT) and in the Portfolio 3, we have only kept the three cryptocurrencies that had the highest prices in order to try and get the highest return possible in the timescale of 5 months that we had data for. You will see that the portfolio 3 has a negative return so, obviously, we wouldn't recommend building your portfolio this way. Besides that, we did the same study as in Chapter 4 to find the best fitting distribution for each portfolio and we have presented the results in the table too. We have also calculated the historical and the parametrical VaR and CVaR for each of these portfolio at 95% and 99% only at time horizon 1 here (values calculated with the log-returns not the returns).



(a) Correlation Heatmap



(b) Simulated Portfolio Optimization

		Portfolio 1	Portfolio 2	Portfolio 3
Weights	BCH	5.85%	0.52%	0.81%
	XLM	26.51%	XXX	XXX
	ADA	1.13%	XXX	XXX
	USDT	21.2%	XXX	XXX
	DASH	0.59%	0.46%	0.22%
	ETC	9.74%	23.27%	98.97%
	QTUM	0.09%	0.87%	XXX
	ONT	25.67%	33.27%	XXX
	ICX	9.21%	41.62%	XXX
Scores	Returns	0.04	0.05	-0.18
	Volatility	0.94	1.17	1.07
Distribution	Best Distribution	Laplace	NIG	Laplace
	Parameters	$\mu = 0.00026$, b = 0.01069	$\mu = -0.0012, \alpha = 0.8561$, $\beta = 0.0863, \delta = 0.0143$	$\mu = 0.00027$ b = 0.01796
VaR & CVaR	Historical VaR - 5%	0.018823	0.021516	0.032860
	Historical CVaR - 5%	0.001571	0.001960	0.003144
	Historical VaR - 1%	0.030828	0.034571	0.057484
	Historical CVaR - 1%	0.000496	0.000899	0.001313
	Parametrical VaR - 5%	0.019786	0.023143	0.035216
	Parametrical CVaR - 5%	0.024794	0.029032	0.044067
	Parametrical VaR - 1%	0.027954	0.032748	0.049651
	Parametrical CVaR - 1%	0.032015	0.037524	0.056828

References

[1] *Bitcoin Cash*, available at `https://www.investopedia.com/terms/b/bitcoin-cash.asp` and same for every cryptocurrency studied.

[2] Stephen Chan, Jeffrey Chu, Saralees Nadarajah and Joerg Osterrieder *A Statistical Analysis of Cryptocurrencies*, (Journal of Risk and Financial Management, Manchester, 2017).

[3] F. Clementia, T. Di Matteo and M. Gallegati, *The power-law tail exponent of income distributions*, (Science Direct, Rome, 2006).