

The Statistics of Bitcoin and Cryptocurrencies

Joerg Osterrieder

Zurich University of Applied Sciences, School of Engineering, 8401 Winterthur, Switzerland

joerg.osterrieder@zhaw.ch

Abstract. Cryptocurrencies became popular with the emergence of Bitcoin and have shown an unprecedented growth over the last few years. As of November 2016, more than 720 cryptocurrencies exist, with Bitcoin still being the most popular one. We show the statistical properties of the most important cryptocurrencies. We characterize their exchange rates versus the US Dollar by fitting parametric distributions to them, including the Student t distribution, the generalized hyperbolic distribution as well as the asymmetric normal inverse Gaussian and the asymmetric variance gamma distribution. Our findings show that cryptocurrencies exhibit strong non-normal characteristics, with standard heavy-tailed distributions such as the Student t distribution giving good descriptions of the data. This is the first study that looks at the parametric distribution of cryptocurrency returns. The results are important for investment and risk management purposes.

Keywords: Bitcoin, Cryptocurrencies, Heavy-tailed distributions, generalized hyperbolic distribution

1 Introduction

Bitcoin, the first decentralized cryptocurrency, has gained large interest from media, academics and the finance industry. Since 2009, many cryptocurrencies have been created, with, as of November 2016, more than 720 in existence. Bitcoin represents over 81% of the total market of cryptocurrencies [1]. The combined market capitalization of all of them is about 14 bn USD (as of November, 17, 2016). The top 15 cryptocurrencies represent more than 97% of the market. In our analysis, we are focusing on those cryptocurrencies that have existed for more than two years and are within the top 15 currencies by market capitalization. Those are: Bitcoin, Ripple, Litecoin, Monero, Dash, MaidSafeCoin, Doge. Those seven represent 90% of the total market capitalization. [2] give a statistical analysis of the exchange rate of Bitcoin. They fit fifteen of the most popular parametric distributions to the log-returns of the exchange rate of Bitcoin versus the US dollar. Our goal is to extend their analysis by characterizing the statistical properties of the seven most important cryptocurrencies.

The paper is organized as follows. In section two, we give an overview of the data which we used and the sources from which it was retrieved. Section three looks at statistical properties of cryptocurrencies by fitting some of the most common parametric distributions to the data. The last section four concludes and summarizes our findings.

2 Data

We are using historical global price indices for cryptocurrencies from the database BNC2 from quandl.com, which shows aggregated cryptocurrency prices from multiple exchanges providing a weighted average cryptocurrency price. Daily data is downloaded from June 23, 2014 until the end of October 2016. On purpose, we have chosen to start only in June 2014, because that allows us to analyze seven out of the top fifteen cryptocurrencies by market capitalization, as of November 2016. For the current market capitalization of cryptocurrencies, (4) is a good reference.

The seven cryptocurrencies that we have chosen out of the top fifteen are: Bitcoin, Dash, Litecoin, MaidSafeCoin, Monero, Doge and Ripple. Most notably, we have omitted Ethereum, with an initial release in July 30, 2015, Ethereum Classic, which only started trading in 2016, Augur and NEM, which also only came into existence in 2015. In total, our choice of cryptocurrencies represents 90% of the market capitalization as of November 2016.

3 Statistics of Cryptocurrencies

Before we fit parametric distributions to the data, let us first plot the empirical histograms of the seven cryptocurrencies versus the US dollar. Next, we will fit the following seven distributions to the exchange rate returns: Normal, Student's t (t), skewed generalized Student's t (gen. t), generalized hyperbolic (GHYP) with their special cases hyperbolic (HYP), asymmetric normal inverse gaussian (asymm. NIG) and asymmetric variance gamma distribution (asymm. VG).

3.1 Normal distribution and qq-plots

In Fig. 1, we are showing the histogram of daily returns of the Bitcoin / USD exchange rate. The red line which is overlaid, shows a normal distribution with mean and standard deviation taken from the empirical Bitcoin/ USD exchange rates. The blue line is showing a Gaussian kernel-density estimator with bandwidth multiple 0.5. We see a substantial deviation from the normal distribution.

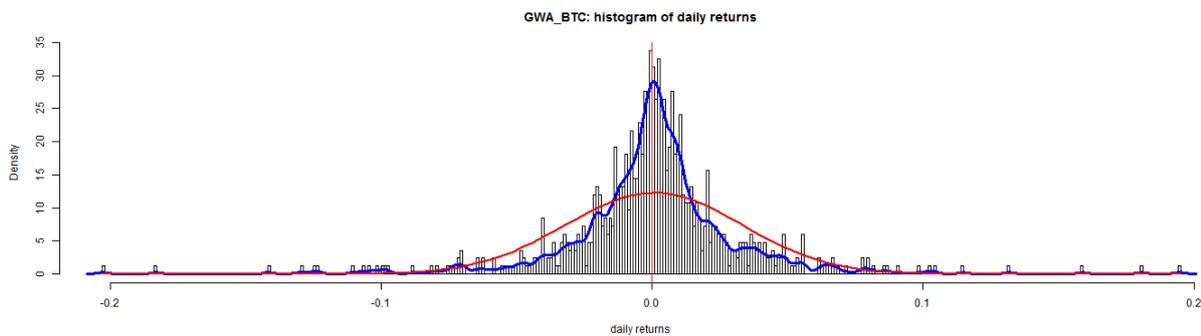


Figure 1: Histogram of Bitcoin/ USD exchange rate with fitted normal distribution and Gaussian KDE overlaid

The next Fig. 2 shows the qq-plot of the empirical Bitcoin returns versus the quantiles of the standard normal distribution.

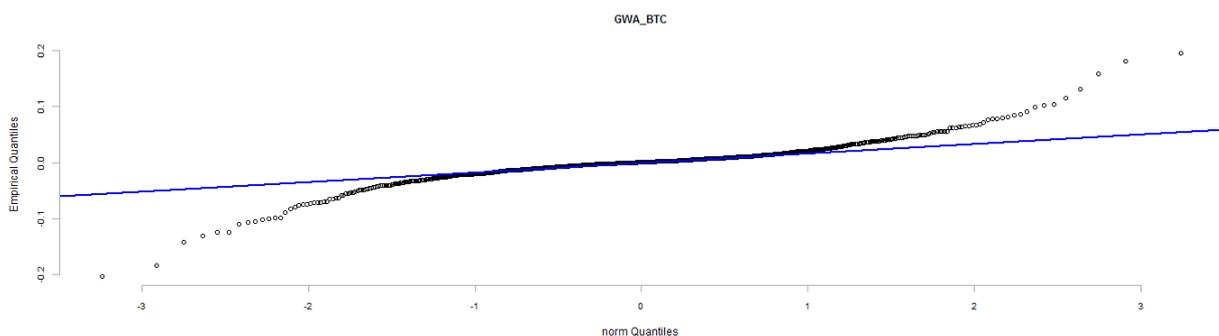


Figure 2: QQ-Plot of the Bitcoin/USD exchange rate versus the normal distribution

This short look at the normal distribution and the seemingly substantial deviation from it, justifies our desire to fit other distributions to the data, since the normal distribution does not accurately reflect the properties of the data. In particular, there is clear evidence of heavy tails, which will guide us in selecting appropriate distributions in the next section.

3.2 Heavy-tailed Distributions

Two recent papers on fitting of distributions to exchange rate data (traditional fiat currencies) are [3] and [4]. [3] fitted the generalized Lambda, skew t , normal inverse Gaussian and normal distributions as well as the Johnson's family of distributions to the data. [4] fitted the Student's t , asymmetric

Student's t, hyperbolic, generalized hyperbolic, generalized Lambda, skew t, normal inverse Gaussian and normal distributions to the data. [2] has fitted both heavy tailed and light-tailed distributions to the Bitcoin exchange rate: The normal, logistic, Laplace, exponential power, skew normal, skewed exponential power and asymmetric exponential power distributions have light tails. The Student's t, skew t, generalized t, skewed Student's t, asymmetric Student's t, normal inverse gamma, hyperbolic and generalized hyperbolic distributions have heavy tails.

In our analysis, and motivated by Section two, we restrict ourselves to heavy-tailed distributions. We do not think that light-tailed distributions will accurately represent the data, even if in one case or the other, a certain light-tailed distribution might give a better fit than one of the heavy-tailed distributions. Our approach is consistent with most of the literature on exchange rates. In addition, the results in those papers which we will cite in the next paragraph, all show that heavy-tailed distributions describe exchange rate returns more accurately.

For the purpose of our analysis, we will choose the following distributions: Normal (as default case, even though it is light-tailed), Student's t (t), skewed generalized Student's t (gen. t), hyperbolic (HYP), generalized hyperbolic (GHYP), asymmetric normal inverse Gaussian (asymm. NIG) and the asymmetric variance gamma distribution (asymm. VG). The skewed generalized Student's t distribution encompasses both light and heavy-tailed distributions.

The choice of these distributions is not arbitrary. They have been used by many authors in the past to describe exchange rate returns of traditional fiat currencies. Some of the papers are: [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Probably the most comprehensive collection of distributions used to analyze any exchange rate data set is given by [2]. Each distribution was fitted by the method of maximum likelihood. The log-likelihood of the fits are given in Table 1.

Table 1: Log-likelihood of fitted distributions

Cryptocurrency	normal	t	gen. t	HYP	GHYP	asymm. NIG	asymm. VG
Bitcoin	1700	1884	1896	1873	1896	1892	1890
Ripple	1208	1420	1429	1399	1431	1429	1421
Litecoin	1413	1696	1707	1649	1710	1706	1692
Monero	1054	1100	1106	1106	1106	1106	1106
Doge	1333	1527	1535	1513	1539	1537	1525
Dash	1191	1310	1319	1314	1322	1321	1317
MaidSafeCoin	1086	1114	1124	1122	1123	1121	1123

From this, we can see that the generalized hyperbolic always leads to the highest log-likelihood. This is not too surprising, since the GHYP distribution has the largest number of free parameters, in particular, the hyperbolic, asymmetric normal inverse Gaussian and the asymmetric variance gamma distribution are special cases of it. In addition, we see that the normal fit leads to a log-likelihood which is substantially lower than all the other fits. Lastly, once you have chosen a heavy-tailed distribution, the results are all very similar and we therefore argue to choose the simplest distribution which would be the asymmetric Student t distribution. We leave it to the reader to compute the AIC and BIC criterion to differentiate between the log-likelihood of the fit. Looking at both the Table 1 above and Fig. 1 we are confident that it is obvious that the results for any of the heavy-tailed distributions are very close to each other, also from a statistical point of view.

Therefore, we restrict ourselves to showing the fitted parameters for the normal and the t distribution. The fitted parameters for the normal distribution are in Table 2:

Table 2: Parameters and standard errors for the fit of the normal distribution

Cryptocurrency	μ	Σ	μ s.e.	σ s.e.
Bitcoin	0.00058889	0.03219635	0.00110890	0.00078411
Ripple	0.00161814	0.05754736	0.00198439	0.00140317
Litecoin	0.00021672	0.04511072	0.00155554	0.00109993
Monero	0.00077327	0.06882905	0.00237766	0.00168125
Doge	0.00057583	0.04947804	0.00170715	0.00120714
Dash	0.00114279	0.05870780	0.00202440	0.00143147
MaidSafeCoin	0.00325831	0.06629558	0.00228877	0.00161841

The fitted parameters for the t distribution are in Table 3:

Table 3: Parameters and standard errors for the fit of the t distribution

Cryptocurrency	μ	Σ	Df	μ s.e.	σ s.e.	df s.e.
Bitcoin	0.00078	0.014227	1.816715	0.000627	0.000787	0.168360
Ripple	-0.0017	0.023136	1.655060	0.001039	0.001258	0.138970
Litecoin	-0.0008	0.014474	1.376026	0.000663	0.000828	0.104893
Monero	-0.0021	0.047822	3.411781	0.001986	0.002282	0.473698
Doge	-0.0035	0.021107	1.751907	0.000936	0.001177	0.157810
Dash	-0.0037	0.031316	2.214337	0.001370	0.001592	0.222622
MaidSafeCoin	-0.0001	4.569920	0.051049	0.002065	0.002366	0.810976

In Fig. 3 we are showing the empirical histogram of Bitcoin together with the fitted density of the seven distributions that we have fitted to the data. All charts show that the exchange rate returns are non-normal and that any of the heavy-tailed distributions describes the data quite well. The spike of the variance-gamma distribution around the mean is usually too large. For simplicity reasons, we therefore opt for using a Student's t distribution to describe the data. Those charts for the other six cryptocurrencies are available on request.

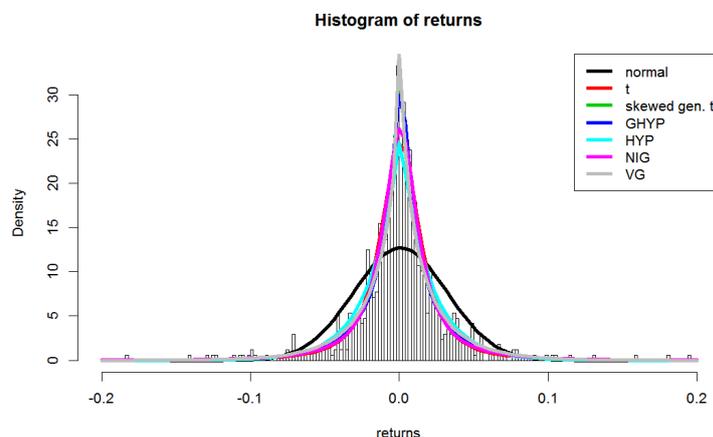


Figure 3: Histogram of Bitcoin/ USD exchange rate with fitted densities overlaid

4 Conclusion

Cryptocurrencies exhibit heavy tails. We have fitted both the normal distribution and a set of heavy-tailed distributions to exchange rates of seven of the top 15 cryptocurrencies. The choice of those seven was governed by the necessity of having data for at least two years in September 2016. Once you have chosen a heavy-tailed fit, the results are good and very similar. Therefore, we opt for choosing a Student *t* distribution when fitting exchange rate returns of cryptocurrencies. Implications for this are in the area of risk management, where you need to compute value-at-risk and expected shortfall as well as for investment purposes. To our knowledge, this is the first study looking at the statistical properties of cryptocurrencies, going beyond Bitcoin and the traditional fiat currencies.

References

- [1] CoinMarketCap: Crypto-currency market capitalizations (2016), <https://coinmarketcap.com/>
- [2] Chu, J., Nadarajah, S., Chan, S.: Statistical analysis of the exchange rate of bitcoin. *PLOS ONE* 10(7), e0133678 (07 2015)
- [3] Corlu, C.G., Corlu, A.: Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance* 15(11), 1851–1864 (2015)
- [4] Nadarajah, S., Afuecheta, E., Chan, S.: A note on modelling exchange rate returns: Which flexible distribution to use? *Quantitative Finance* 15(11), 1777–1785 (2015)
- [5] Coppes, R.C.: Are exchange rate changes normally distributed? *Economics Letters* 47(2), 117–121 (1995)
- [6] Bauer, C.: Value at risk using hyperbolic distributions. *Journal of Economics and Business* 52(5), 455–467 (2000)
- [7] Pipien, M.: Garch processes with skewed-*t* and stable conditional distributions. Bayesian analysis for pln/usd exchange rate. *Folia Oeconomica Cracoviensia* 45, 45–62 (2004)
- [8] Linden, M.: Estimating the distribution of volatility of realized stock returns and exchange rate changes. *Physica A: Statistical Mechanics and its Applications* 352(2), 573–583 (2005)
- [9] Alexander, C., Lazar, E.: Normal mixture garch (1, 1): Applications to exchange rate modelling. *Journal of Applied Econometrics* 21(3), 307–336 (2006)
- [10] Gurrola, P.: Capturing fat-tail risk in exchange rate returns using su curves: A comparison with the normal mixture and skewed student distributions. *The Journal of Risk* 10(2), 73 (2007)
- [11] Kilic, R.: Conditional volatility and distribution of exchange rates: Garch and figarch models with nig distribution. *Studies in Nonlinear Dynamics & Econometrics* 11(3) (2007)
- [12] Lopez Herrera, F., Rodriguez Benavides, D., Ortiz Arango, F.: The stochastic volatility of the peso-dollar exchange rate: The floating regime in Mexico. *Investigacion Economica* 70(276), (2011)
- [13] Nakajima, J.: Stochastic volatility model with regime-switching skewness in heavy-tailed errors for exchange rate returns. *Studies in Nonlinear Dynamics and Econometrics* 17(5), 499–520 (2013)