The Statistics of Bitcoin and Cryptocurrencies

Jörg Osterrieder[1]

[a]Zurich University of Applied Sciences, School of Engineering, Switzerland

This manuscript was compiled on November 18, 2016

We show the statistical properties of the most important cryptocurrencies, of which Bitcoin is the most prominent example. We characterize their exchange rates versus the US Dollar by fitting parametric distributions to them. It is shown that returns are clearly non-normal, with standard heavy-tailed distributions giving good descriptions of the data. The results are important for investment and risk management purposes.

Bitcoin | cryptocurrency | statistics | parametric distributions

1. Introduction and motiviation

Bitcoin, the first decentralized cryptocurrency, has gained 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 (2). 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.

(3) 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, (2) 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).

Normal distribution and qq-plots. In figure 1, we are showing the histogram of daily returns of the Bitcoin / USD exchange rate. The red line which is overlayed, 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.

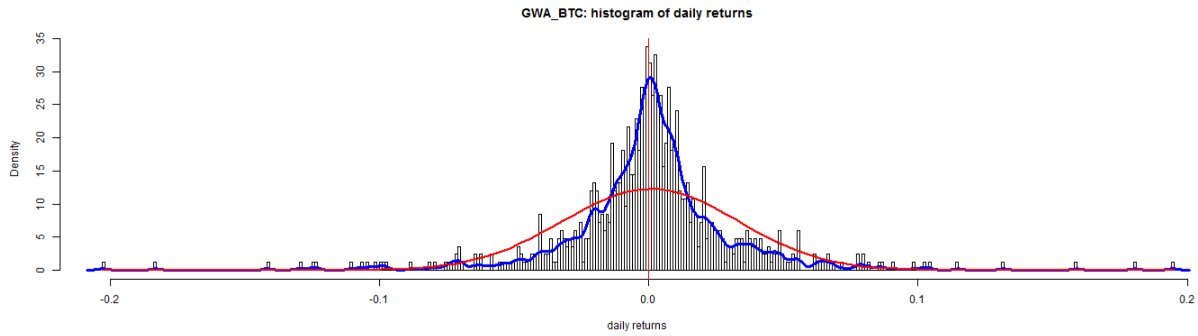

Fig. 1. Histogram of Bitcoin/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

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

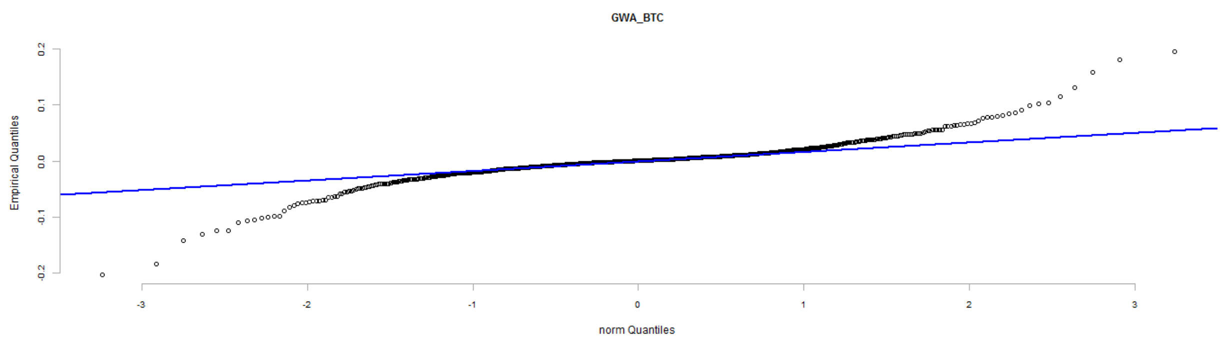

Fig. 2. QQ-Plot of the Bitcoin/USD exchange rate versus the normal distribution

The following six figures 3, 4, 5, 6, 7, 8 show the empirical histogram of our six other cryptocurrencies, Dash, Litecoin, MaidSafeCoin, Ripple, Doge and Monero. We have chosen to restrict the x-axis to returns between -0.2 and 0.2, thus omitting occasional large outliers.

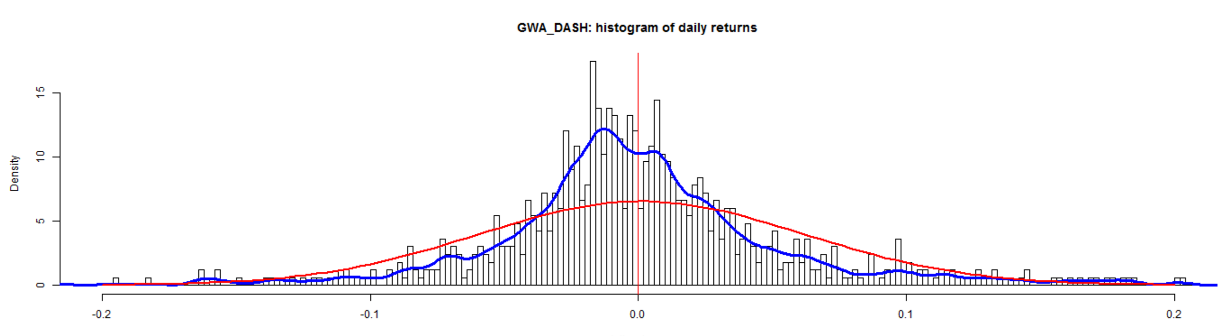

Fig. 3. Histogram of Dash/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

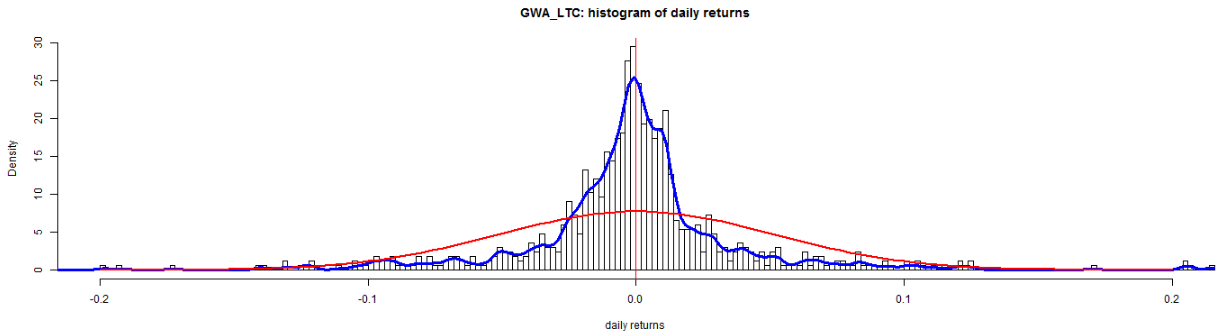

Fig. 4. Histogram of LTC/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

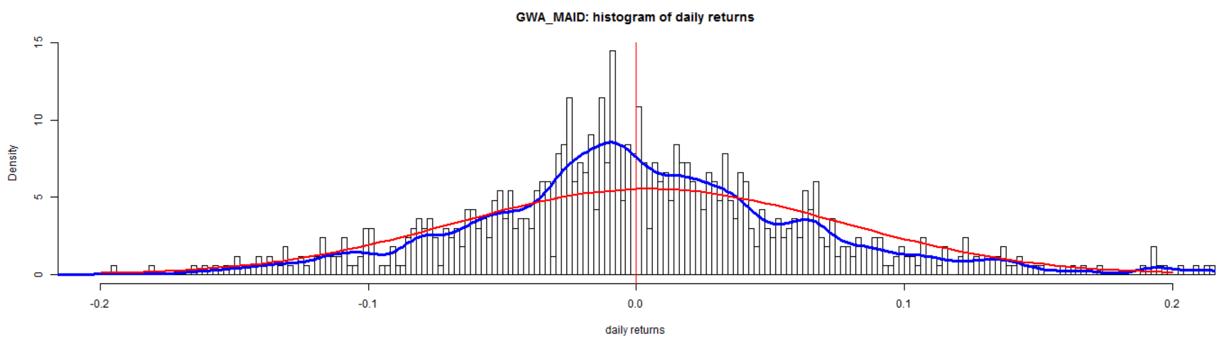

Fig. 5. Histogram of MAID/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

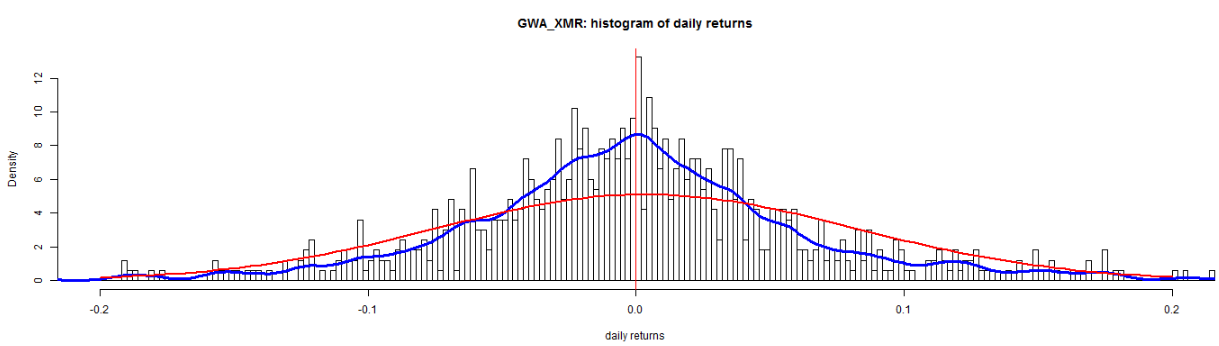

Fig. 6. Histogram of XMR/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

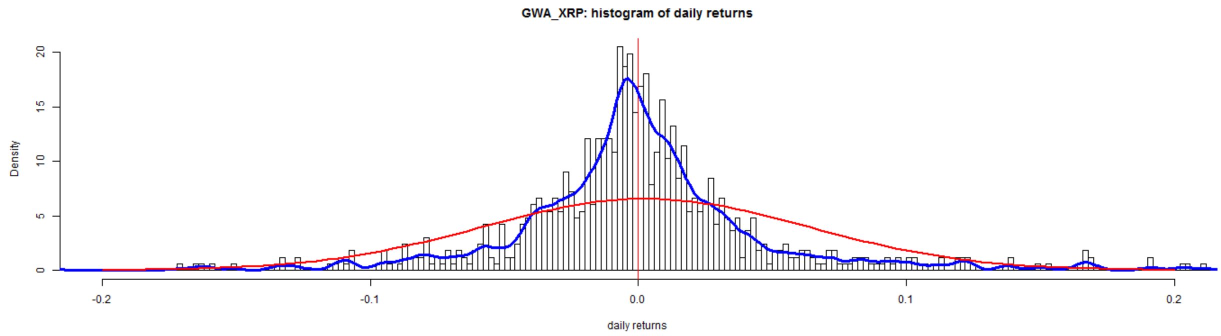

Fig. 7. Histogram of XRP/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

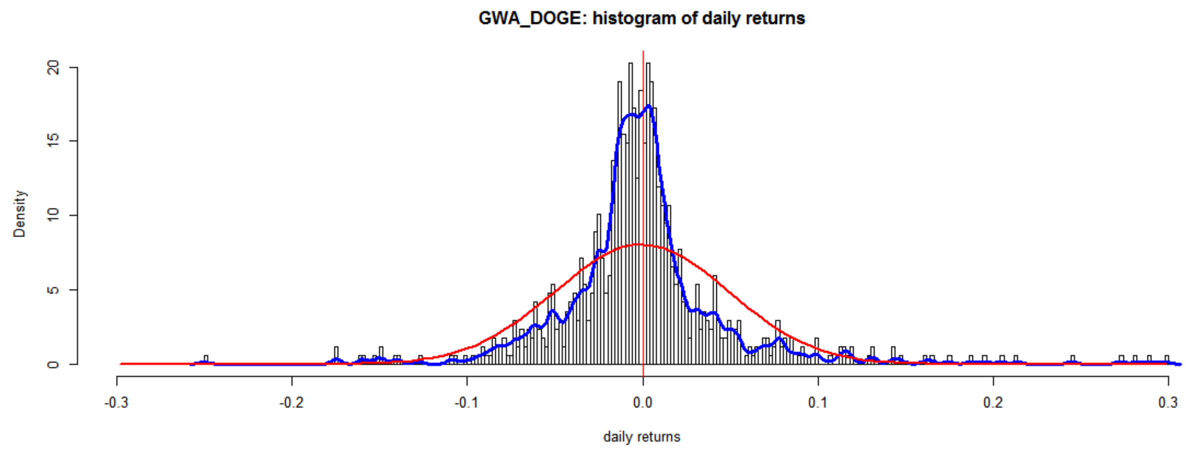

Fig. 8. Histogram of Doge/ USD exchange rate with fitted normal distribution and Gaussian KDE overlayed

We see a substantial deviation from the normal distribution in all six plots.

In the following six charts, 9, 10, 11, 12, 13, 14 we show the qq-plot of the empirical returns versus the quantiles of the standard normal distribution. Again, we observe large deviations from the normal distribution both on the left and the right tail for all cryptocurrencies.

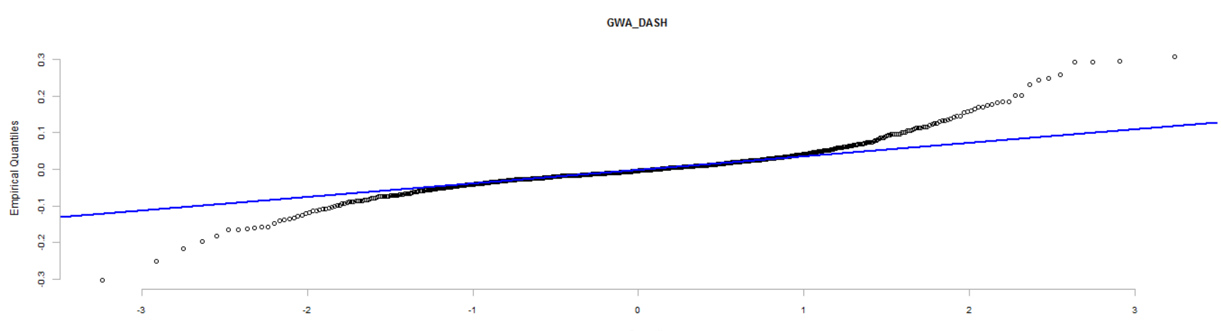

Fig. 9. QQ-Plot of the DASH/USD exchange rate versus the normal distribution

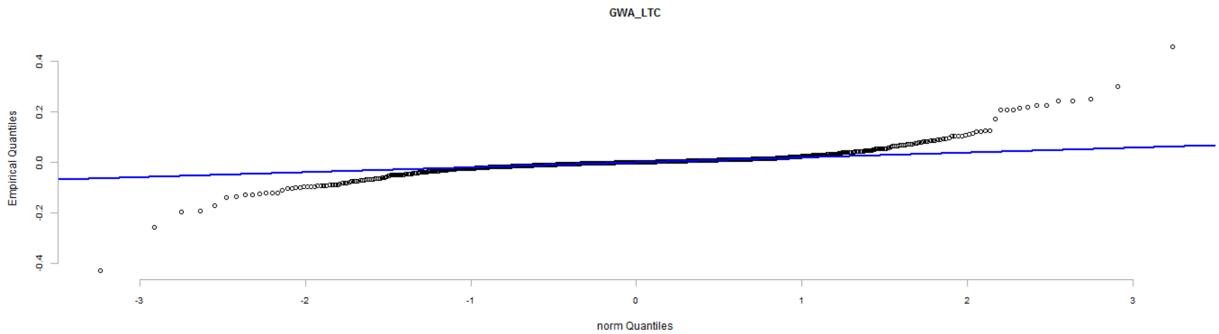

Fig. 10. QQ-Plot of the LTC/USD exchange rate versus the normal distribution

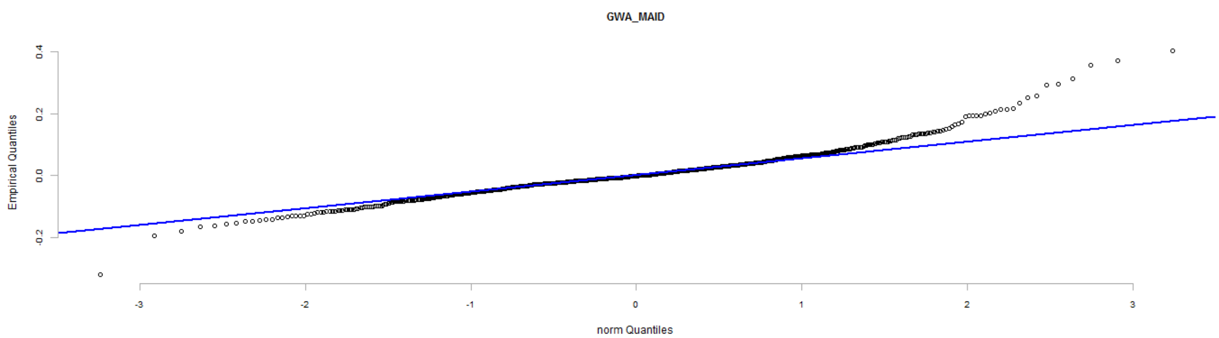

Fig. 11. QQ-Plot of the MAID/USD exchange rate versus the normal distribution

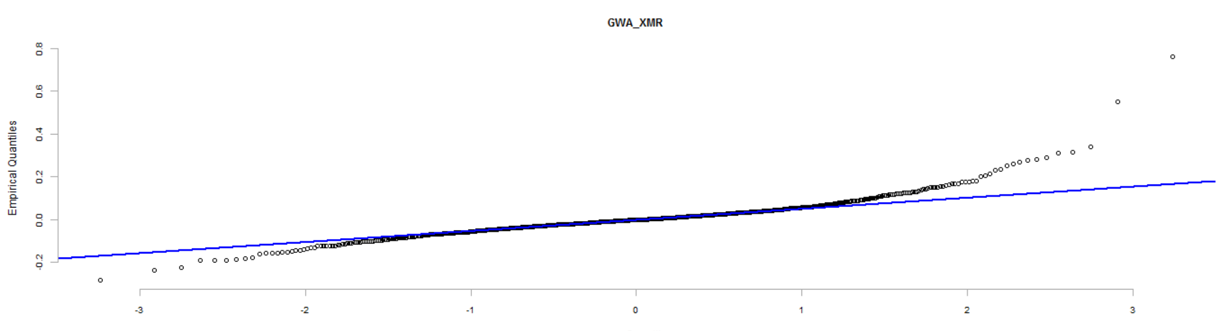

Fig. 12. QQ-Plot of the XMR/USD exchange rate versus the normal distribution

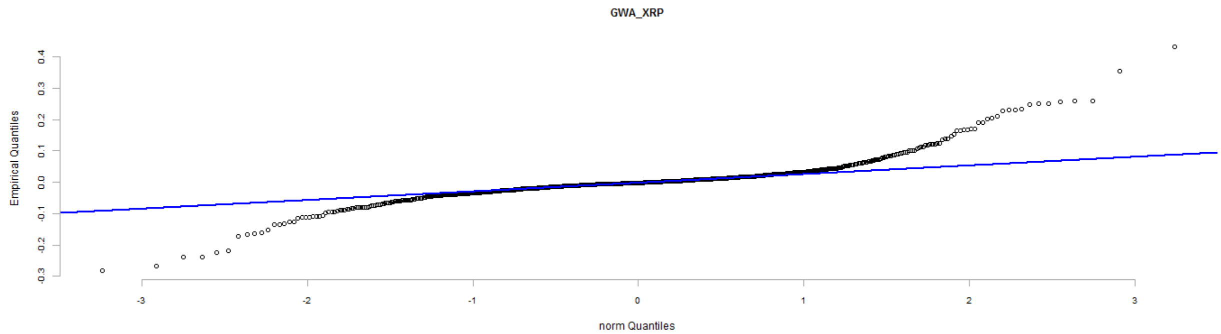

Fig. 13. QQ-Plot of the XRP/USD exchange rate versus the normal distribution

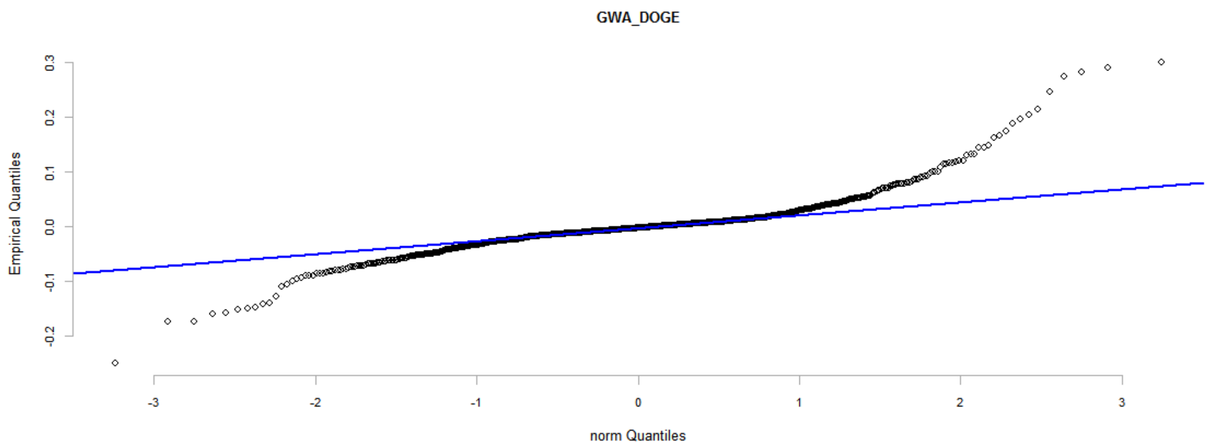

Fig. 14. QQ-Plot of the Doge/USD exchange rate versus the normal distribution

Those charts are mostly taken from (4). For more a detailed analysis of those returns, please see (5) and (4).

This short look at the normal distribution and the seemingly substantial devation 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.

Heavy-tailed distributions. Two recent papers on fitting of distributions to exchange rate data (traditional fiat currencies) are (6) and (7). (6) fitted the generalized Lambda, skew t, normal inverse Gaussian and normal distributions as well as the Johnson's family of distributions to the data. (7) 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. (3) 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 the 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.

We start by giving a short theoretical introduction and show the density function in all cases. Next, we use maximum-likelihood to fit those distributions to the data.

Normal distribution

The normal distribution is probably the most important continuous distribution. The density is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

with $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

Student's t distribution

In statistics, the t-distribution was first derived as a posterior distribution in 1876 by Helmert and Lüroth. The t-distribution also appeared in a more general form as Pearson Type IV distribution in Karl Pearson's 1895 paper. In the English-language literature the distribution takes its name from William Sealy Gosset's 1908 paper in Biometrika under the pseudonym "Student".

Student's t-distribution has the probability density function given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where ν is the number of degrees of freedom and Γ is the gamma function.

Skewed generalized t distribution

In probability and statistics, the skewed generalized t distribution is a family of continuous probability distributions. The distribution was first introduced by Panayiotis Theodossiou in 1998.

The density is given by

$$f(x) = \frac{p}{2\nu\sigma q^{\frac{1}{p}} B\left(\frac{1}{p}, q\right) \left(\frac{|x-\mu+m|^p}{q(\nu\sigma)^p(\lambda sign(x-\mu+m)+1)^p} + 1\right)^{\frac{1}{p}+1}}$$

where B is the beta function, μ is the location parameter, $\sigma > 0$ the scale parameter, $-1 < \lambda < 1$ is the skewness parameter and $p > 0$, $q > 0$ are the parameters that control the kurtosis. m and ν are given by

$$m(\nu, \sigma, \lambda, p, q) = \frac{2\nu\sigma\lambda q^{\frac{1}{p}} B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)}$$

and

$$\nu(\lambda, q, p) = \frac{q^{-\frac{1}{p}}}{\sqrt{(3\lambda^2+1)\frac{B\left(\frac{3}{p}, q-\frac{2}{p}\right)}{B\left(\frac{1}{q}, p\right)} - 4\lambda^2 \frac{B\left(\frac{2}{p}, q-\frac{1}{p}\right)^2}{B\left(\frac{1}{p}, q\right)^2}}}$$

Special and limiting cases of the skewed generalized t distribution include the skewed generalized error distribution, the generalized t distribution, the skewed t, the skewed Laplace distribution, the generalized error distribution, a skewed normal distribution, the student t distribution, the skewed Cauchy distribution, the Laplace distribution, the uniform distribution, the normal distribution and the Cauchy distribution.

Generalized hyperbolic distribution

The generalized hyperbolic distribution (GH), introduced by Ole Barndorff-Nielsen, is a continuous probability distribution defined as the normal variance-mean mixture where the mixing distribution is the generalized inverse Gaussian distribution. Its probability density function is given in terms of modified Bessel function of the second kind, denoted by K_λ.

Its density is given by:

$$f(x) = \frac{\left(\frac{\gamma}{\delta}\right)^\nu}{\sqrt{2\pi}K_\lambda(\delta\gamma)} e^{\beta(x-\mu)} \frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\frac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{\frac{1}{2}-\lambda}}$$

for $x \in \mathbb{R},\ \alpha \in \mathbb{R},\ \lambda \in \mathbb{R},\ \beta \in \mathbb{R},\ \delta \in \mathbb{R},\ \mu \in \mathbb{R},\ \gamma = \sqrt{\alpha^2 - \beta^2}$.

Hyperbolic distribution

The hyperbolic distribution is a continuous probability distribution characterized by the logarithm of the probability density function being a hyperbola. The hyperbolic distributions form a subclass of the generalised hyperbolic distributions.

Its density is given by:

$$f(x) = \frac{\gamma}{2\alpha\delta K_1(\delta\gamma)} e^{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)}$$

for $x \in \mathbb{R},\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R},\ \delta \in \mathbb{R},\ \mu \in \mathbb{R},\ \gamma = \sqrt{\alpha^2 - \beta^2}$. K_1 denotes a modified Bessel function of the second kind.

Asymmetric Variance Gamma

The variance-gamma distribution, generalized Laplace distribution or Bessel function distribution is a continuous probability distribution that is defined as the normal variance-mean mixture where the mixing density is the gamma distribution. The tails of the distribution decrease more slowly than the normal distribution. The variance-gamma distributions form a sub-class of the generalised hyperbolic distributions.

The density is given as

$$f(x) = \frac{\gamma^{2\lambda} |x-\mu|^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}(\alpha |x-\mu|)}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-\frac{1}{2}}} e^{\beta(x-\mu)}$$

for $x \in \mathbb{R},\ \alpha \in \mathbb{R},\ \lambda \in \mathbb{R},\ \beta \in \mathbb{R},\ \mu \in \mathbb{R},\ \gamma = \sqrt{\alpha^2-\beta^2}$. K_1 denotes a modified Bessel function of the second kind and Γ denotes the gamma function.

Asymmetric Normal Inverse Gaussian Distribution

The normal-inverse Gaussian distribution (NIG) is a continuous probability distribution that is defined as the normal variance-mean mixture where the mixing density is the inverse Gaussian distribution. The NIG distribution was noted by Blaesild in 1977 as a subclass of the generalised hyperbolic distribution discovered by Ole Barndorff-Nielsen.

$$f(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}} e^{\delta\gamma+\beta(x-\mu)}$$

for $x \in \mathbb{R},\ \alpha \in \mathbb{R},\ \delta \in \mathbb{R},\ \beta \in \mathbb{R},\ \mu \in \mathbb{R},\ \gamma = \sqrt{\alpha^2-\beta^2}$. K_1 denotes a modified Bessel function of the second kind and Γ denotes the gamma function.

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: (8), (9), (10), (11), (12), (13), (14), (15), (16), (6), (7) Probably the most comprehensive comprehensive collection of distributions used to analyze any exchange rate data set is given by (3).Each distribution was fitted by the method of maximum likelihood. The log-likelihood of the fits are given in table 1.

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

Table 1. Log-likelihood

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 criterium to differentiate between the log-likelihood of the fit. Looking at both the table 1 above and figure 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:

Cryptocurrency	μ	σ	μ s.e.	σ s.e.
Bitcoin	0.0005888929	0.0321963547	0.0011089014	0.0007841117
Ripple	0.001618149	0.057547360	0.001984392	0.001403177
Litecoin	0.0002167247	0.0451107296	0.0015555424	0.0010999346
Monero	0.0007732781	0.0688290549	0.0023776602	0.0016812597
Doge	0.0005758394	0.0494780495	0.0017071549	0.0012071408
Dash	0.001142796	0.058707804	0.002024407	0.001431472
MaidSafeCoin	0.003258315	0.066295580	0.002288778	0.001618410

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

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

Cryptocurrency	μ	σ	df	μ s.e.	σ s.e.	df s.e.
Bitcoin	0.0007832907	0.0142276950	1.8167153214	0.0006278324	0.0007873350	0.1683602724
Ripple	−0.001746360	0.023136868	1.655060658	0.001039305	0.001258018	0.138970085
Litecoin	−0.0008376863	0.0144748002	1.3760266099	0.0006633390	0.0008281613	0.1048934398
Monero	−0.002154319	0.047822804	3.411781959	0.001986291	0.002282549	0.473698699
Doge	−0.003587645	0.021107224	1.751907628	0.000936940	0.001177574	0.157810490
Dash	−0.003717149	0.031316854	2.214337293	0.001370649	0.001592906	0.222622694
MaidSafeCoin	−0.0001129375	4.5699208998	0.0510499146	0.0020651955	0.0023668137	0.8109761548

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

In figures 15, 16, 17, 18, 19, 20, 21 we are showing the empirical histogram of the seven cryptocurrencies together with the fitted density of the seven distributions that we have fitted to the data. All charts shows 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.

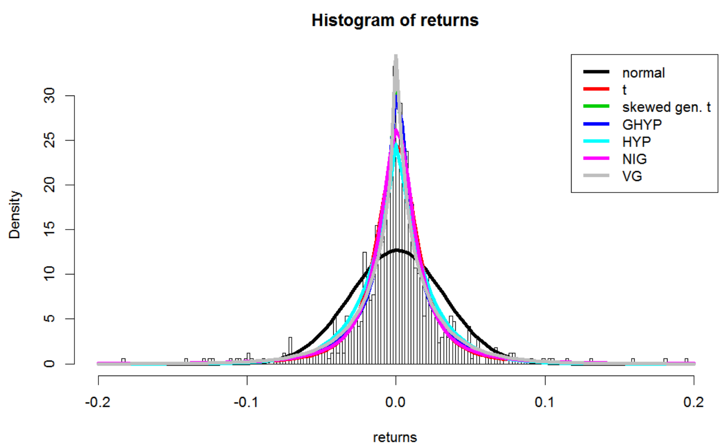

Fig. 15. Histogram of Bitcoin/ USD exchange rate with fitted densities overlaid

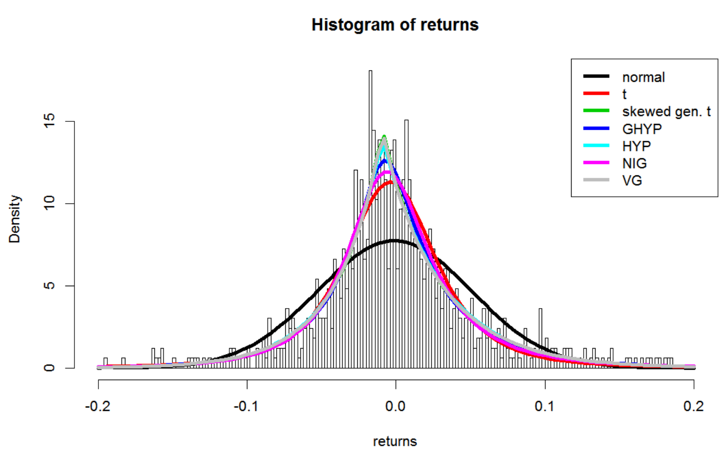

Fig. 16. Histogram of Dash/ USD exchange rate with fitted densities overlaid

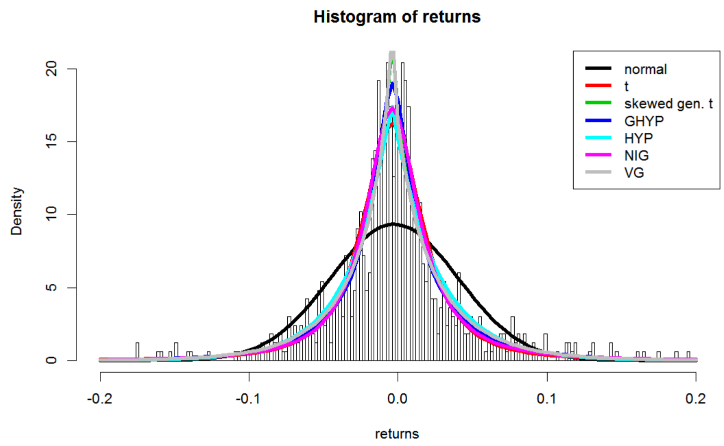

Fig. 17. Histogram of Doge/ USD exchange rate with fitted densities overlaid

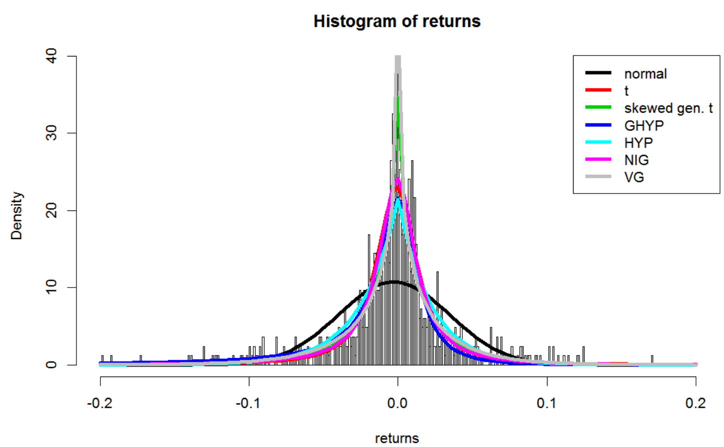

Fig. 18. Histogram of LTC/ USD exchange rate with fitted densities overlaid

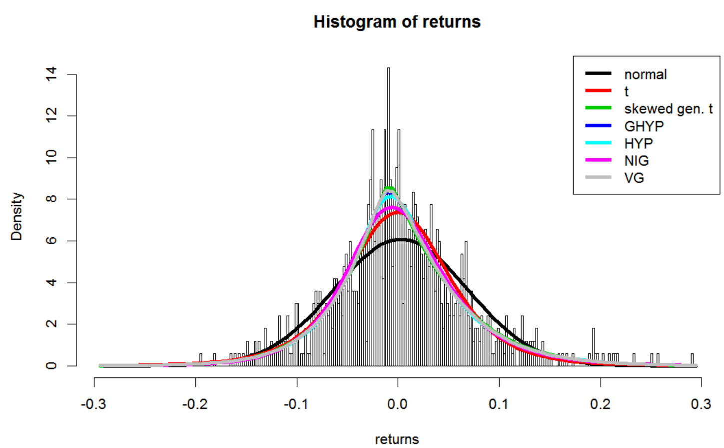

Fig. 19. Histogram of Maid/ USD exchange rate with fitted densities overlaid

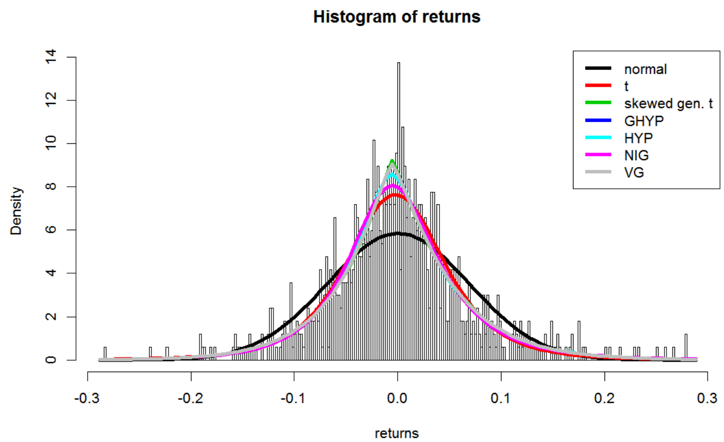

Fig. 20. Histogram of XMR/ USD exchange rate with fitted densities overlaid

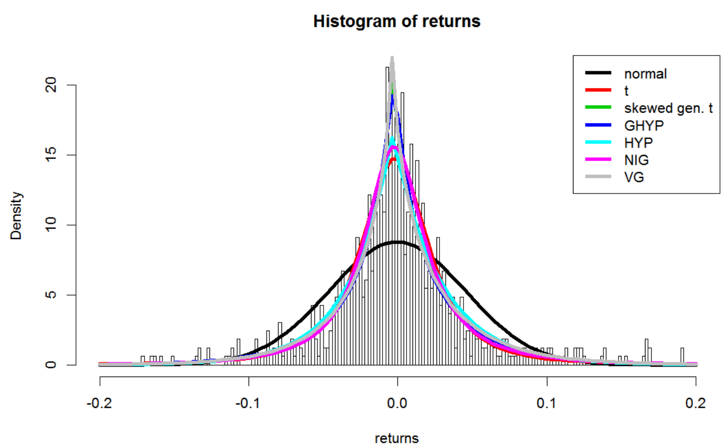

Fig. 21. Histogram of XRP/ USD exchange rate with fitted densities overlaid

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

CoinMarketCap (2016) Crypto-currency market capitalizations.

Chu J, Nadarajah S, Chan S (2015) Statistical analysis of the exchange rate of bitcoin. *PLOS ONE* 10(7):e0133678.

Osterrieder J, Lorenz J, Strika M (2016) Bitcoin and cryptocurrencies — not for the faint-hearted.

Osterrieder J, Lorenz J (2016) A statistical risk assessment of bitcoin and its extreme tail behaviour.

Corlu CG, Corlu A (2015) Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance* 15(11):1851–1864.

Nadarajah S, Afuecheta E, Chan S (2015) A note on "modelling exchange rate returns: Which flexible distribution to use?". *Quantitative Finance* 15(11):1777–1785.

Coppes RC (1995) Are exchange rate changes normally distributed? *Economics Letters* 47(2):117–121.

Bauer C (2000) Value at risk using hyperbolic distributions. *Journal of Economics and Business* 52(5):455–467.

Pipień M (2004) Garch processes with skewed-t and stable conditional distributions. bayesian analysis for pln/usd exchange rate. *Folia Oeconomica Cracoviensia* 45:45–62.

Linden M (2005) Estimating the distribution of volatility of realized stock returns and exchange rate changes. *Physica A: Statistical Mechanics and its Applications* 352(2):573–583.

Alexander C, Lazar E (2006) Normal mixture garch (1, 1): Applications to exchange rate modelling. *Journal of Applied Econometrics* 21(3):307–336.

Gurrola P (2007) 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.

Kiliç R (2007) Conditional volatility and distribution of exchange rates: Garch and figarch models with nig distribution. *Studies in Nonlinear Dynamics & Econometrics* 11(3).

Lopez Herrera F, Rodriguez Benavides D, Ortiz Arango F (2011) The stochastic volatility of the peso-dollar exchange rate: The floating regime in mexico. *Investigacion Economica* 70(276):19–+.

Nakajima J (2013) 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.