Saha, Kunal

Preprint

An investigation into the dependence structure of major cryptocurrencies

Suggested Citation: Saha, Kunal (2018) : An investigation into the dependence structure of major cryptocurrencies, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:
http://hdl.handle.net/10419/181878

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sollten die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
An investigation into the dependence structure of major cryptocurrencies

Kunal Saha

Institute for Financial Management and Research, Chennai 600034, INDIA

Abstract

This paper attempts to examine the dependence structure of four major cryptocurrencies chosen by current market capitalisation. It is a well known fact that there is huge volatility in the prices of these cryptocurrencies. The Vine Copula model is used to get some insights about the dependence structure in these asset prices. This is done using daily closing price from August 2015 to May 2018. This information can be used to calculate risk based metrics such as expected shortfall of a portfolio of these currencies. This analysis becomes more important as complex financial instruments (e.g. indices) based on these currencies are being introduced.

Keywords: Vine Copula, Cryptocurrencies, Expected shortfall

JEL Classification: C51, C52, C58, G11

Declarations of interest: none

1. Introduction

Cryptocurrencies have attracted a lot of attention about their projected use as digital currencies not governed by any central authority. Since 2009, when Bitcoin was invented, much debate has happened to ascertain if these are just momentary asset bubbles or actual disruptions in the way we know and use money.

*Corresponding author

Email address: kunal.saha@ifmr.ac.in (Kunal Saha)
However, the markets seem to have accepted them. As per data from website coinmarketcap.com, the total reported market cap of all cryptocurrencies is more than US$ 340 billion as of June 1, 2018. This is still lower than the all-time high market cap of US$ 835 billion.

There is a lot of financial innovation riding on this wave. There are more than 1600 cryptocurrencies available today. These can be traded via many exchanges or OTC trading desks. A few exchanges have introduced cryptocurrency based indices (e.g. CRIX). Several Index funds (e.g., Crypto20, Coinbase Index Fund) have also been introduced. As more investors and institutions get interested in adding these assets to their portfolios, the need to measure risk dynamics increases. This paper makes an attempt to study the dependence structure among major cryptocurrencies. This can be used in applications like measuring risk estimators such as Expected Shortfall and also to optimise portfolio performance.

2. Review of Literature

Despite growing interest and use, not much research has been published since most cryptocurrencies are very new and not much time series price/volume data is available. Most research on volatility dynamics focuses on the oldest and biggest cryptocurrency (by market capitalisation), Bitcoin.

Baek et al. [2015] examined the relative volatility between Bitcoin and S&P 500 Index daily return data. Fry et al. [2016] analysed the Bitcoin and Ripple cryptocurrency markets for Negative bubbles and shocks.

to June 2017.


It appears that there are no studies that do a higher dimensional dependence modelling between cryptocurrencies. Moreover, in the past few months, the price dynamics of cryptocurrencies have changed completely. Bitcoin prices, that had reached US$ 20000, have now crashed to less than US$ 8000. Very few studies have been published since the prices took a downward turn.

This paper attempts to analyse the latest price series data of four major cryptocurrencies to find new insights and add to the current academic literature on the subject.

3. Data and Methodology

3.1. Data

This paper uses returns data of the four major cryptocurrencies (as per Market cap) namely, Bitcoin(BTC), Ether(ETH), Ripple(XRP) and Litecoin(LTC) for a sample period from August 2015 to May 2018. This constitutes a total of 1028 data points. The data was sourced from the website coinmarketcap.com and is publicly available for download. The four cryptocurrencies constitute about 70% of the total value of cryptocurrencies available. Although Bitcoin Cash is among the top five cryptocurrencies, it is not used in this analysis since it was introduced recently (August 2017).
3.2. Methodology

This paper builds upon the Copula GARCH approach pointed out in Pfaff (2016). The Vine Copula theory is employed in this analysis.

1. Specify and estimate the GARCH models for each loss factor.
2. Determine the standardised residuals and calculate the pseudo-uniform variables.
3. Estimate the Vine copula model.
4. Use the dependence structure determined by the estimated copula to create N sets of random variates (e.g., 100,000) for the pseudo-uniformly distributed variables.
5. Compute the quantiles for these Monte Carlo draws.
6. Use these quantiles along with the weight vector to calculate the N portfolio return cases. The following weight vector is chosen: \[ \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.2 \end{bmatrix} \]. Any weight vector can be used as long as the weights sum up to one.
7. The simulated portfolio returns are then calculated by assigning the above weights to the assets.
8. These returns are then sorted and the expected shortfall (CVaR) at the 95% confidence level is reported as the median of the largest 5% losses.
9. This CVaR is compared with the CVaR for a t-copula model and CVaR based on a historical returns approach to comment on the suitability of this more complex approach.

Copula GARCH Model was introduced in 2002. Since then, it has been extensively used to model dependence, especially in the field of empirical finance.

The marginal distributions of jointly distributed random variables along with inherent dependence structure are contained in their joint distribution function,

\[
F(x_1, \ldots, x_d) = P[X_1 \leq x_1, \ldots, X_d \leq x_d] = C(F_1(x_1), \ldots, F_d(x_d))
\]
The above expression shows that a copula is the distribution function in \( \mathbb{R}^d \) space of a \( d \)-element random vector with standard uniform marginal distributions \( U(0,1) \).

There are several copula construction methods with each having its own strengths and weaknesses. Since we have a four dimensional dataset, the vine copula structure is chosen. Vine Copulas build upon bivariate copulas and provide more flexibility\cite{18} in high-dimensional dependence modeling. Hence, this analysis can also be used to scale for higher dimensional datasets. A cryptocurrency index may have quite a few currencies incorporated in it.

The concept of Vine Copula\cite{19} was first introduced in 1994. However, the graphical dependency models called Vines\cite{20} were introduced in 2002. They are defined\cite{18} as follows:

\( \nu \) is a regular vine (R-vine) on \( d \) elements, with \( \varepsilon(\nu) = \varepsilon_1 \cup \ldots \cup \varepsilon_{d-1} \) being the edges of \( \nu \), if

1. \( \nu = T_1, \ldots, T_{d-1} \) [consists of \( d - 1 \) trees];
2. \( T_1 \) is a connected tree with nodes \( N_1 = 1, \ldots, d \), and edges \( \varepsilon_1 \); for \( l = 2, \ldots, d - 1 \), \( T_l \) is a tree with nodes \( N_l = \varepsilon_{l-1} \) [edges in a tree become nodes in the next tree];
3. (proximity) for \( l = 2, \ldots, d - 1 \), for \( \{n_1, n_2\} \in \varepsilon_l \), \( \oplus(n_1 \Delta n_2) = 2 \), where \( \Delta \) denotes symmetric difference and \( \oplus \) denotes cardinality [nodes joined in an edge differ by two elements].

A regular vine is called a canonical (C-vine) if tree \( T_1 \) has a unique node of degree \( d - l \) (the maximum degree) for \( l = 1, \ldots, d - 2 \). A regular vine is called a drawable vine (D-vine) if all nodes in \( T_1 \) have degree not more than two.

We now move onto the concepts for capturing the dependence between risk factors.

Kendall’s rank correlation coefficient (\( \rho_\tau \)), is defined as

\[
\rho_\tau = \mathbb{E}[\text{sign}((X_1 - X_2)(Y_1 - Y_2))]
\]

The upper and lower tail dependencies for two random variables \( (X,Y) \) with
marginal distributions $F_X$ and $F_Y$ are defined as:

$$
\lambda_u = \lim_{q \to 1} P(Y > F_Y^{-1}(q)|X > F_X^{-1}(q)) \quad \text{and} \quad \lambda_l = \lim_{q \to 0} P(Y \leq F_Y^{-1}(q)|X \leq F_X^{-1}(q))
$$

The expectation and variance equations for ARMA-GARCH models are as follows:

$$
X_{k,t} = \mu_{k,t} + \epsilon_{k,t}
$$

$$
\mu_{k,t} = \mu_k + \sum_{i=1}^{p_{1,k}} \phi_i (X_{k,t-i} - \mu_k) + \sum_{j=1}^{q_{1,k}} \theta_j \epsilon_{k,t-j}
$$

$$
\epsilon_{k,t} = \sigma_{k,t} Z_{k,t}, \quad \text{where} \quad Z \sim \mathcal{D}_{k,t}(0,1)
$$

$$
\sigma_{k,t}^2 = \alpha_0 + \sum_{i=1}^{p_{2,k}} \alpha_i \epsilon_{k,t-i}^2 + \sum_{j=1}^{q_{2,k}} \beta_j \sigma_{k,t-j}^2
$$

where $X_k$ is the return/loss of the $k^{th}$ asset in the portfolio, $k = 1, \ldots, K$

4. Results and Interpretation

4.1. Descriptive Statistics

The raw price series is plotted first, which itself is an indicator of the wild fluctuations in the prices since 2015. These are plotted separately since the absolute price differences across the four assets are very high. We can also spot their tendency to move together.
The returns are plotted together and we can see the basic stylised fact of volatility clustering. This confirms the usefulness of GARCH modeling.

![Log Returns Time Series Plot of Cryptocurrencies](image)

Figure 2: Log Returns Time Series Plot of Cryptocurrencies

Finally, we also look at the descriptive statistics of the four cryptocurrencies.

<table>
<thead>
<tr>
<th>Variable</th>
<th>BTC</th>
<th>ETH</th>
<th>XRP</th>
<th>LTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of Observations</td>
<td>1028</td>
<td>1028</td>
<td>1028</td>
<td>1028</td>
</tr>
<tr>
<td>Mean</td>
<td>0.32%</td>
<td>0.52%</td>
<td>0.42%</td>
<td>0.32%</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.11%</td>
<td>8.24%</td>
<td>7.91%</td>
<td>5.94%</td>
</tr>
<tr>
<td>Minimum</td>
<td>-20.75%</td>
<td>-130.21%</td>
<td>-61.63%</td>
<td>-39.52%</td>
</tr>
<tr>
<td>Maximum</td>
<td>22.51%</td>
<td>41.23%</td>
<td>102.74%</td>
<td>51.03%</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.2526</td>
<td>-3.5223</td>
<td>3.0630</td>
<td>1.3764</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.8310</td>
<td>63.2822</td>
<td>38.6743</td>
<td>13.0949</td>
</tr>
</tbody>
</table>

Table 1: Descriptive Statistics of Daily Returns

We see that the cryptocurrencies vary greatly in their Skewness. Moreover, all are very highly leptokurtic.

4.2. GARCH Estimation

It was observed that the best fit is provided by ARFIMA(0,d,0)-GJR-GARCH(1,1) model[21]. GJR-GARCH also captures the leverage effect and hence, is a richer model. The model is fitted using Skewed Generalized Error Distribution (SGED). The final specification of the fGARCH-gjrGARCH model is as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \sigma_{t-j}^2 (|z_{t-j}| - \eta_j z_{t-j})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$
Diagnostics are run to ensure that relevant coefficients are significantly different from zero and the stability requirement for GARCH(1,1) models is not violated, i.e., $\alpha_1 + \beta_1 < 1$. Further, we can see significant deviation from normality from the estimate of the degrees of freedom parameter $\nu$.

Weighted Ljung-Box Test on the standardised residuals and the squared standardised residuals show no serial correlation. Table 2 shows the fitted GARCH model. Table 3 provides the main diagnostics results.

More diagnostics tests were run but are not included with an intention to keep the paper brief. More results can be provided on request.

An ARFIMA(0, d, 0) is used for the ETH series to ensure that the stability requirement is met. For the other series, the ARFIMA(0, d, 0) was estimated to be zero.

| Parameter | Estimate | SE  | t-stat | P(>|t|) |
|-----------|----------|-----|--------|--------|
| $\mu$     | -0.1836  | 0.0252 | -7.2958 | 0      |
| $\omega$  | 0.10391  | 0.0494 | 2.1027  | 0.0354 |
| $\alpha_1$| 0.1367   | 0.0189 | 7.2198  | 0      |
| $\beta_1$ | 0.8598   | 0.0181 | 47.3589 | 0      |
| $\eta_11$ | 0.1725   | 0.0607 | 2.6426  | 0.0044 |
| skew      | 1.3348   | 0.0265 | 59.0166 | 0      |
| shape     | 0.7556   | 0.0271 | 27.8102 | 0      |
| arfima    | 0.0000   | 0.0009 | 0.0003  | 0.9997 |

| Parameter | Estimate | SE  | t-stat | P(>|t|) |
|-----------|----------|-----|--------|--------|
| $\mu$     | -0.1857  | 0.0417 | -4.4553 | 0.0000 |
| $\omega$  | 2.8616   | 0.6508 | 4.3968  | 0.0000 |
| $\alpha_1$| 0.2922   | 0.0298 | 9.7937  | 0      |
| $\beta_1$ | 0.7055   | 0.0213 | 33.0516 | 0      |
| $\eta_11$ | -0.0382  | 0.0190 | -2.0067 | 0.0447 |
| skew      | 0.9106   | 0.0134 | 67.9057 | 0      |
| shape     | 0.9144   | 0.0447 | 20.4700 | 0      |
| arfima    | 0.0000   | 0.0000 | 23.8800 | 0      |

Table 2: Fitted GARCH estimates
Weighted Ljung-Box Test on Standardised Residuals (H₀: No serial correlation)

d.o.f=0

<table>
<thead>
<tr>
<th></th>
<th>Bitcoin statistic</th>
<th>p-value</th>
<th>Ether statistic</th>
<th>p-value</th>
<th>Ripple statistic</th>
<th>p-value</th>
<th>Litecoin statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log[1]</td>
<td>4.5420</td>
<td>0.0331</td>
<td>4.8250</td>
<td>0.0280</td>
<td>4.1890</td>
<td>0.0407</td>
<td>0.0031</td>
<td>0.9557</td>
</tr>
<tr>
<td>Log[2*(p+q)+(p+q)-1][2]</td>
<td>4.6310</td>
<td>0.0510</td>
<td>5.3430</td>
<td>0.0329</td>
<td>4.2090</td>
<td>0.0663</td>
<td>0.0297</td>
<td>0.9725</td>
</tr>
<tr>
<td>Log[4*(p+q)+(p+q)-1][5]</td>
<td>5.0930</td>
<td>0.1061</td>
<td>6.8920</td>
<td>0.0511</td>
<td>5.8560</td>
<td>0.0972</td>
<td>0.1261</td>
<td>0.9970</td>
</tr>
</tbody>
</table>

Weighted Ljung-Box Test on Squared Standardised Residuals (H₀: No serial correlation)

d.o.f=2

<table>
<thead>
<tr>
<th></th>
<th>Bitcoin statistic</th>
<th>p-value</th>
<th>Ether statistic</th>
<th>p-value</th>
<th>Ripple statistic</th>
<th>p-value</th>
<th>Litecoin statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log[1]</td>
<td>1.1630</td>
<td>0.2809</td>
<td>0.0109</td>
<td>0.9166</td>
<td>0.0011</td>
<td>0.9730</td>
<td>0.0309</td>
<td>0.8603</td>
</tr>
<tr>
<td>Log[2*(p+q)+(p+q)-1][5]</td>
<td>2.1090</td>
<td>0.5929</td>
<td>0.0369</td>
<td>0.9997</td>
<td>0.3024</td>
<td>0.9835</td>
<td>0.1142</td>
<td>0.9975</td>
</tr>
<tr>
<td>Log[4*(p+q)+(p+q)-1][9]</td>
<td>2.7260</td>
<td>0.8032</td>
<td>0.0639</td>
<td>1.0000</td>
<td>0.5072</td>
<td>0.9984</td>
<td>0.1862</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

Table 3: GARCH Diagnostics

4.3. Copula Estimation

The residuals from the GARCH model are used to complete the Copula modeling exercise. First, the estimated Kendall’s τ are estimated.

<table>
<thead>
<tr>
<th>τ</th>
<th>BTC</th>
<th>ETH</th>
<th>XRP</th>
<th>LTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTC</td>
<td>1</td>
<td>0.1884</td>
<td>0.2157</td>
<td>0.5057</td>
</tr>
<tr>
<td>ETH</td>
<td>0.1884</td>
<td>1</td>
<td>0.2181</td>
<td>0.2296</td>
</tr>
<tr>
<td>XRP</td>
<td>0.2157</td>
<td>0.2181</td>
<td>1</td>
<td>0.2860</td>
</tr>
<tr>
<td>LTC</td>
<td>0.5057</td>
<td>0.2296</td>
<td>0.2860</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: Estimated Kendall’s τ

With higher dimensional data, the number of possible R-vines increases dramatically. There are n!/2 × 2\(^{(n+1)/2}\) possible regular vines\(^{22}\) on n nodes. Among these, there are n!/2 distinct C-vine trees\(^{23}\) and n!/2 distinct D-vine trees.

It can be seen that enumeration and selection of the best possible Vine Copula construction can get computationally difficult quickly. Hence, a method based on kendall’s Tau proposed by Dissmann et al\(2013\)\(^{24}\) is used. Bayesian Information Criterion is used for choosing the individual pair copula constructions. Moreover, the model chosen is more frugal than the one indicated by Akaike information criterion.
The Vine copula relationships are estimated\cite{25} and reported as follows:

<table>
<thead>
<tr>
<th>Tree</th>
<th>Edge</th>
<th>Copula</th>
<th>Parameter 1</th>
<th>Parameter 2</th>
<th>( \tau )</th>
<th>Upper Tail Dependence</th>
<th>Lower Tail Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4,1</td>
<td>Gumbel</td>
<td>2.03</td>
<td>0</td>
<td>0.51</td>
<td>0.59</td>
<td>-</td>
</tr>
<tr>
<td>4,2</td>
<td>Gumbel</td>
<td>1.31</td>
<td>0</td>
<td>0.24</td>
<td>0.30</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4,3</td>
<td>Gumbel</td>
<td>1.41</td>
<td>0</td>
<td>0.29</td>
<td>0.37</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2,1;4</td>
<td>Frank</td>
<td>0.61</td>
<td>0</td>
<td>0.07</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3,2;4</td>
<td>Gumbel</td>
<td>1.16</td>
<td>0</td>
<td>0.14</td>
<td>0.18</td>
<td>-</td>
</tr>
<tr>
<td>3,1:2,4</td>
<td>Independence</td>
<td>-</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Type: C-vine
Log Likelihood: 627.76
AIC: -1245.51
BIC: -1220.84

1 = Bitcoin 2 = Ether 3 = Ripple 4 = Litecoin

Table 5: Estimated Vine Copula Structure

A goodness of fit test\cite{26} is used to check if the fitted t-Copula is indeed not satisfactory. The null hypothesis, \( H_0 : C \in C_0 \) where \( C_0 \) is specified as t-Copula is rejected at 1% significance level. For the chosen Vine Copula, the above null hypothesis could not be rejected even at 10% significance level.

Hence, it can be concluded that a t-Copula is indeed a mis-specification for our dataset. More importantly, our Vine Copula construction appears to be a good fit for the given dataset.

4.4. Application

Now that we have a copula structure at hand, we need to see how useful it is for a real world application.

Calculating Estimated Shortfall (also known as Conditional Value at Risk) is an important aspect of portfolio management. Estimated Shortfall (at a specified \( \alpha \)) is the mean/median of the lowest \( (1-\alpha)\% \) of returns. For instance, if ES(95) is estimated to be 8%, this would suggest that in the worst 5% of the returns, the average loss would be 8%.

The Expected Shortfall calculation is repeated for a mis-specified Copula to check the difference in Estimated Shortfall. A t-copula is fitted\cite{27} to the residuals. Further, Estimated Shortfall is also calculated using historical returns. The results are compared.
<table>
<thead>
<tr>
<th>Method</th>
<th>Aug 2015 - May 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vine Copula</td>
<td>16.12%</td>
</tr>
<tr>
<td>t-copula</td>
<td>5.06%</td>
</tr>
<tr>
<td>Historical</td>
<td>3.82%</td>
</tr>
</tbody>
</table>

Table 6: Comparison of Estimated Shortfall calculated using various methods

We can see that there is quite a bit of difference between the CVaR values for the different models chosen. Given the high kurtosis values seen earlier, we already know the return distributions will have fat tails. Now, the above calculated CVaR agrees with that observation.

Hence, we can say that a Vine Copula construction is useful in measuring the risk of our chosen portfolio.

5. Conclusion

This paper uses Vine Copula modeling on a basket of cryptocurrencies to analyse the dependence structure. We see that the Vine Copula based model provides the highest Estimated Shortfall.

Given the high volatility in all the cryptocurrencies along with the positive association between the cryptocurrencies, investors should be careful about their investment decisions in any cryptocurrency. Moreover, from the point of view of a portfolio manager or an investment analyst, this paper shows that risk estimates vary greatly depending on the model chosen.

An argument can be made that risk measures based on naive methods or misspecified models underestimate the inherent risk in a portfolio. As mentioned earlier, this analysis becomes more important as new financial instruments based on cryptocurrencies are being introduced.

Apart from the cryptocurrency assets, the use of Vine Copula modeling has attracted much attention due to the inherent flexibility which is useful to provide a better fit to real life data. Further research using other multivariate models
to compare and contrast model performance will help us draw more definitive conclusions.

6. Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References


