# QUANTITATIVE RISK MANAGEMENT

# ASSIGNMENT 2

**Pan Pan Wu**
Student Number: 2662780
pan-pan.wu@student.uva.nl

**Anna-Maria Angelova**
Student Number: 2683090
angelova8668@gmail.com

June 16, 2020

**ABSTRACT**

In this report we performed analysis by using PCA, FA, copulas and EVT on our portfolio consisting of 8 assets. Each method shows their strengths in analysing our portfolio from different perspectives.

## Contents

**1 Introduction** **1**

**2 Principal Component Analysis (PCA)** **1**

**3 Factor analysis (FA)** **4**

**4 Copulas** **6**

4.1 Bivariate Copulas - Theoretical Framework . . . 6

4.1.1 Gaussian Copula . . . 6

4.1.2 Student-t Copula . . . 6

4.1.3 Gumbel Copula . . . 6

4.1.4 Frank Copula . . . 6

4.2 Bivariate Copulas - Results . . . 7

4.3 Model Selection: Goodness of Fit . . . 7

4.4 Pseudo-samples and density plots . . . 8

4.5 Fitted Copulas . . . 10

**5 Extreme Value Theory (EVT)** **11**

**6 Bibliography** **14**

# 1 Introduction

In this assignment we create a portfolio of 8 assets - S&P 500 (GSPC), Nikkei 225 (N225), Bitcoin (BTCUSD=X), Invesco DB Commodity Index Tracking Fund (DBC), Tesla (TSLA), Lufthansa (LHA.DE), Adidas (ADS.DA), and Heineken (HEIA). We use synchronized historical data for the period from 31-03-2011 to 31-03-2020, meaning that we keep the dates, only for which all assets have available quotes simultaneously. This results in 2093 dates. While we have USD, JPY and EUR exposure in the portfolio, exchange rates are not considered.

# 2 Principal Component Analysis (PCA)

The principle of PCA can be expressed as

$$C = w1 * Y1 + w2 * Y2 + w3 * Y3 + w4 * Y4$$

where C is the component which explains the variances, w(i) is the weights of each variable (asset) and Y(i) is the value of each variable.

This can be interpreted as: each principle component is a hyperplane expanded by the linear combination of each variable weighted by their corresponding weights.

We applied principal component analysis (PCA) on our portfolio. We applied it using full components, which is equal to the number of assets, so 8 components. From figure 1 we can see that the first PC explained over 60% of the variances and the first two components together explained over 80% of the variances.

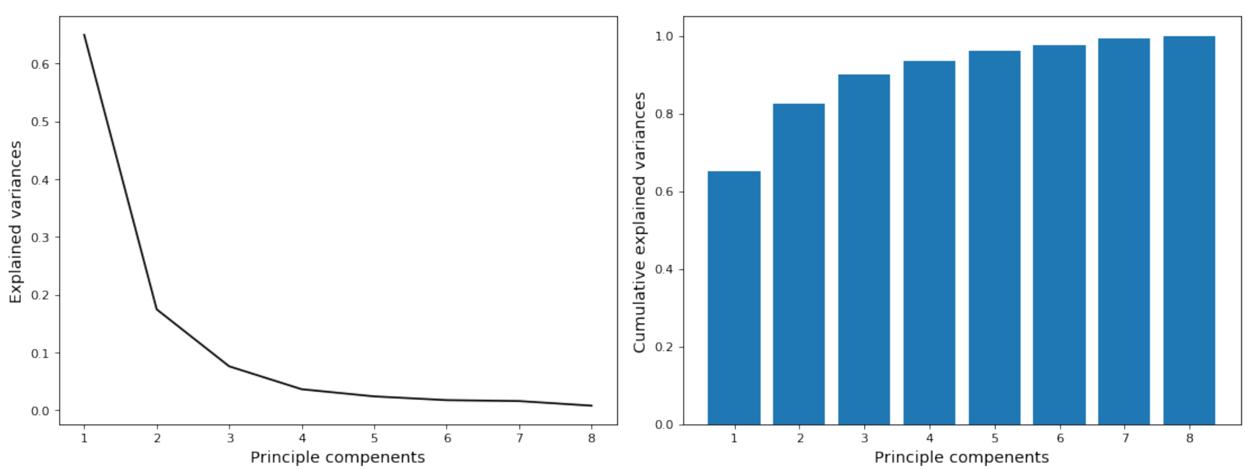



**Figure 1.** (Left) Explained variances by each principle component. (Right) The cumulative explained variances with respect to the number of principal component.

Using the scree plot as shown in figure 2 , we can determine the number of PC to retain. As the rule of thumb is to keep the PC which has eigenvalue larger than 1, therefore, 3 or 4 PCs would be optimal for our portfolios. Bitcoin (BTC) has the highest loading among all the asset and is higher than 0.9, this indicates that Bitcoin explains nearly all the variances within the first PC. Loadings of other assets are nearly 0 (figure 3 left). Tesla (TSLA) has the highest loading (over 0.8) in the second PC and other assets except Bitcoin and all point to the same direction. The opposite sign of Bitcoin indicates that it has negative correlation with other assets within PC2 (figure 3 right). Within the third PC, Tesla has negative correlation with respect to other assets and Lufthansa (LHA) has the highest loading. Lufthansa and Adidas (ADS) have the equal loadings while negative correlation with each other within PC4 (figure 4).

From figure 5 we can see the portfolio returns with respect to the first and second PCs. The curve of PC1 is more volatile compared to PC2, this can be seen from the loadings of PC1 (figure 3 left) that Bitcoin explains nearly all of the variances, so the returns with respect PC1 reflect the returns of Bitcoin. As the volatility of Tesla is not that extreme compared to Bitcoin, the portfolio of PC2 is relatively stable.

The scatter plot of the outcomes of PCA with respect to PC1 and PC2 can show the cluster of different years, we visualize them in figure 6. While 2017-2019 forms a close cluster, the data with respect to 2020 differ with the others and is scattered in a wider range. This is due to the corona virus crisis.

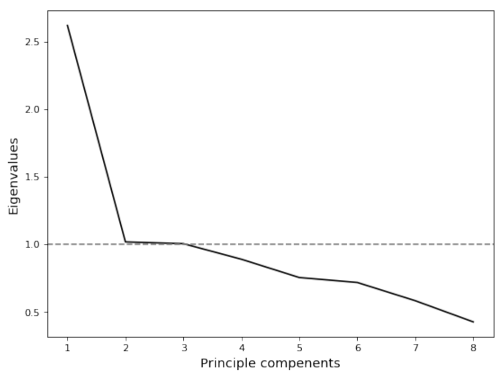



**Figure 2.** The eigenvalues with respect to each PC.

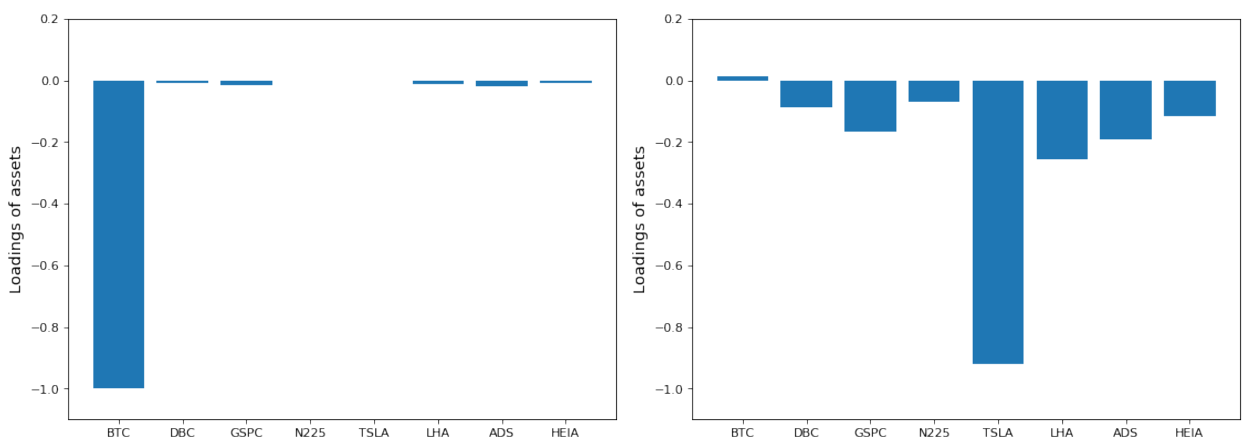
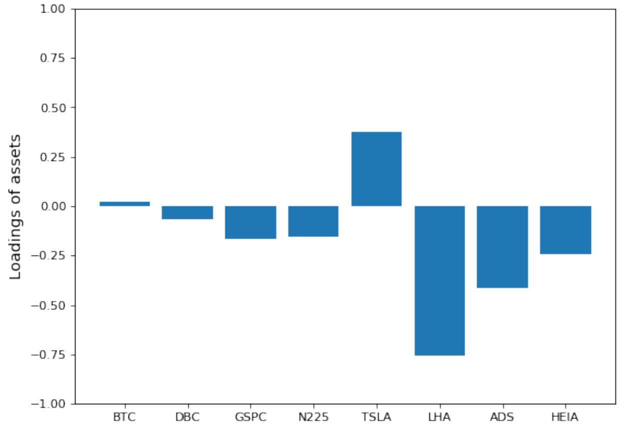



**Figure 3.** (Left) The loading of each asset with respect to the first PC. (Right) The loading of each asset with respect to the second PC.



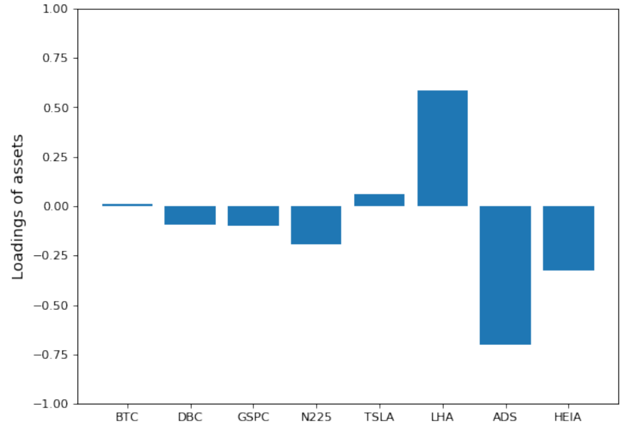




**Figure 4.** (Left) The loading of each asset with respect to the third PC. (Right) The loading of each asset with respect to the fourth PC.

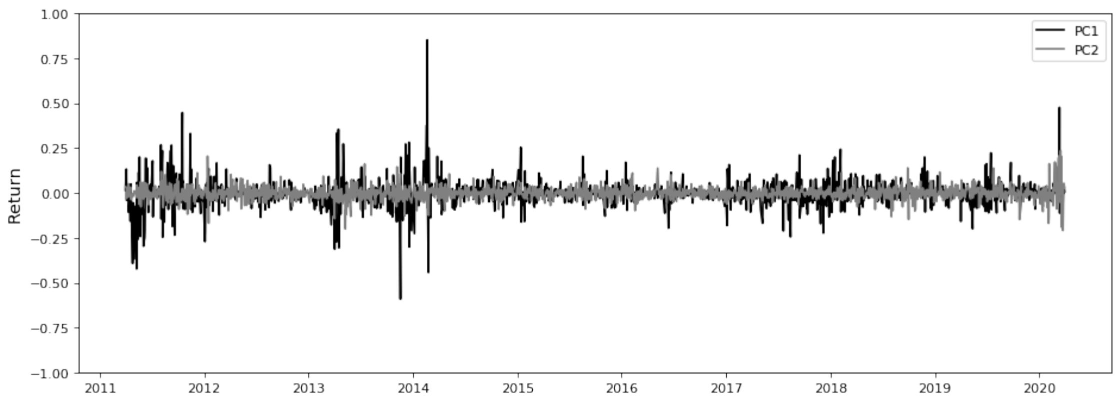



**Figure 5.** The returns of the portfolio with respect to the first and second PCs.

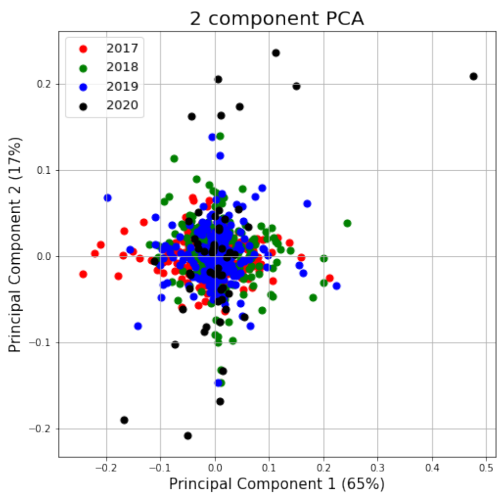



**Figure 6.** The scatter plot of the daily principle component 1 and principle component 2 within the recent three years (750 days).

## 3 Factor analysis (FA)

Different from PCA, FA assumes that there are common latent variables which explain the variances. This can be expressed as:

$$Y1 = b1 * F + u1$$
$$Y2 = b2 * F + u2$$
$$Y3 = b3 * F + u3$$
$$Y4 = b4 * F + u4$$

where Y means each variable, F the factor, b(i) the weights and u(i) the error term. Although both PCA and FA are both data reduction methods, there is difference between them.

We used varmax as the rotation method. Different from PCA, 8 factors explains only around 40% variances (figure 7 left). The scree plot indicates that the optimal number of factors are 3 or 4 as from factor 4 has eigenvalue lower than 1 (figure 7 right). The eigenvalues are the same for both PCA and FA as they are calculated with the same correlation matrix without rotation for FA. The loadings of each asset with respect to each factor are shown in figure 8 and 9. ADS and HEIA (Heiniken) have higher loadings in factor 1, DBC and GSPC have higher loadings in factor 2, GSPC and TSLA have higher loadings in factor 3 and GSPC and LHA have higher loadings in factor 4. Figure 10 shows the portfolio returns with respect to F1 and F2, while F1 is more volatile between 2014 and 2015, F2 is more volatile at the beginning (between 2011 and 2012) and the end of the period (from 2020).

From the results of both PCA and FA, we can see that there are negative correlations between assets using PCA while the correlations are all positive for FA (for 4 PC and 4 Factor). In addition, For the first and second PsC, BTC and TSLA explain the majority of the variances respectively. FA, on the other hand, shows relations more on groups. I.e. ADS and HEIA are both explained much by F1 and DBC and GSPC are explained the most by F2.

We can conclude that both PCA and FA are applicable for analyzing our portfolio and they analyze in a different way.

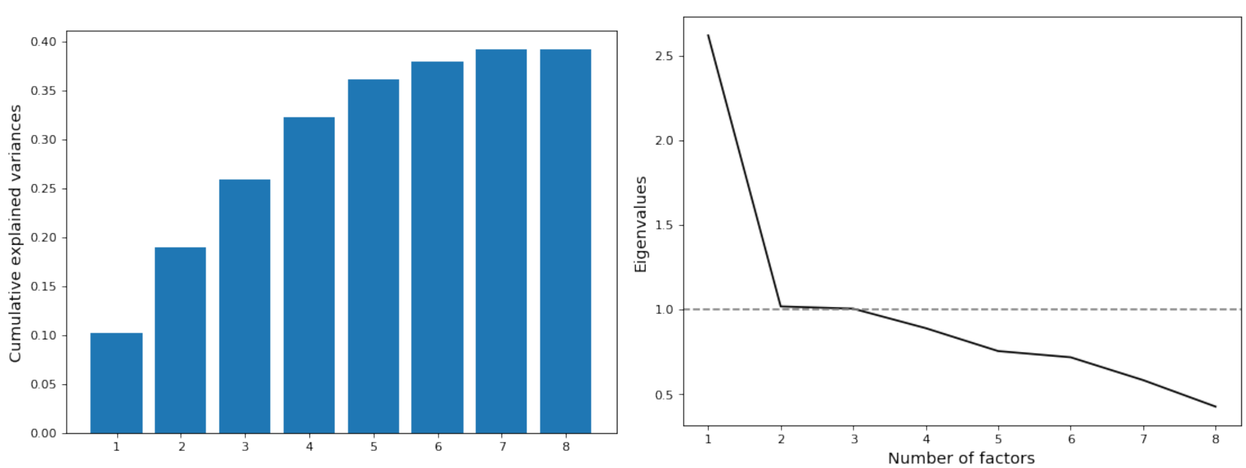



**Figure 7.** (Left) The accumulative explained variances for using eight factors. (Right) The eigenvalues of each factor.

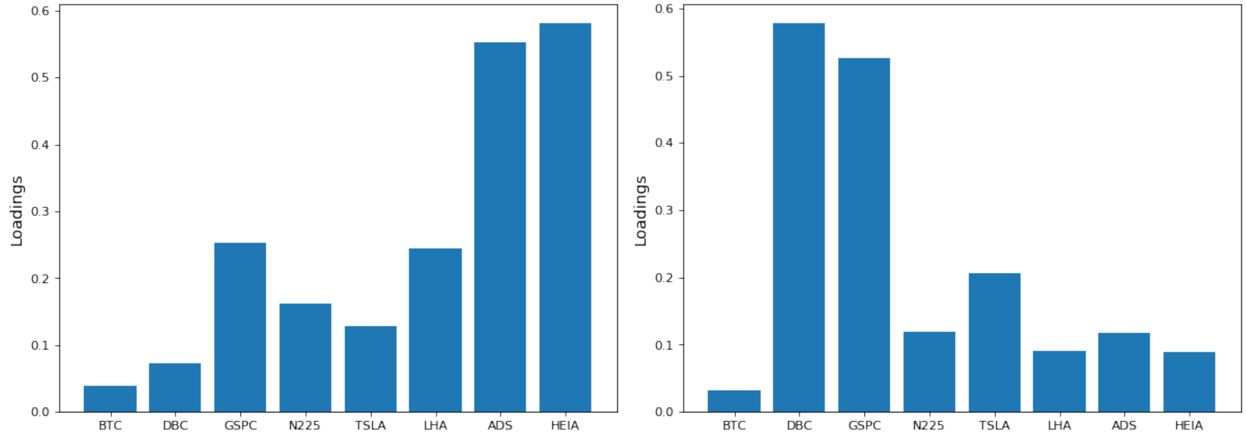



**Figure 8.** (Left) The loadings of each asset of the first factor. (Right) The loadings of each asset of the second factor.

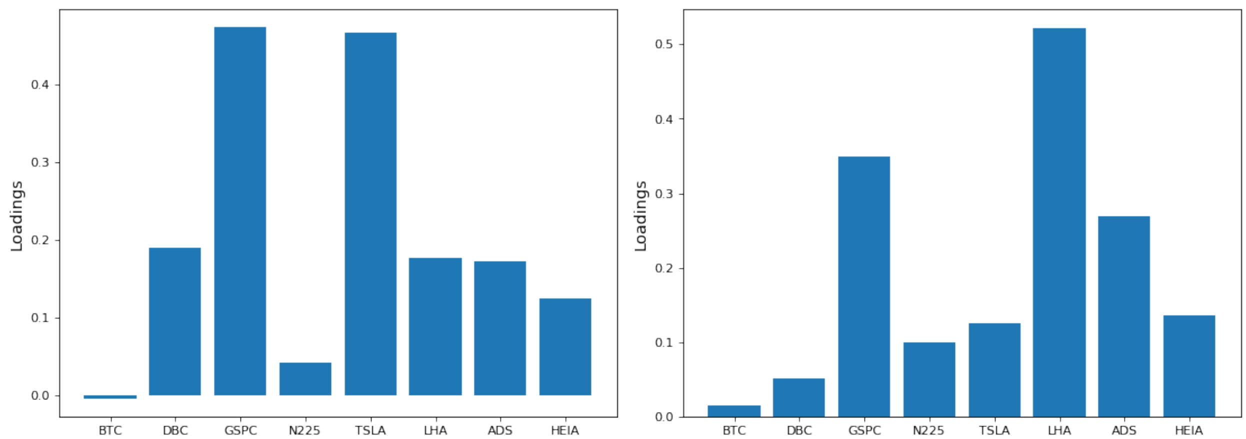



**Figure 9.** (Left) The loadings of each asset of the third factor. (Right) The loadings of each asset of the fourth factor.

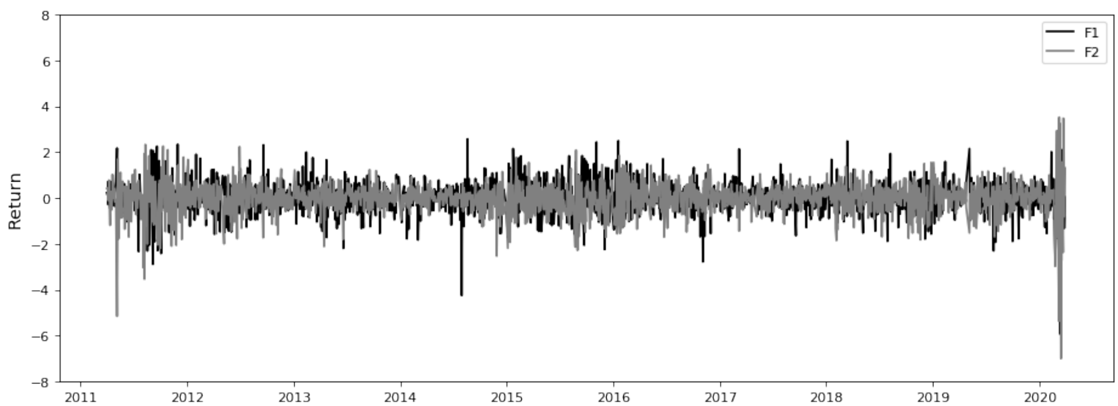



**Figure 10.** The returns of the portfolio with respect to the first and second factors.

# 4 Copulas

We fit copulas for five pairs of assets using VineCopula R package, which determines the best copula model and its parameters. The package selects student-t copula as the best fit for all five pairs. For comparison purpose, we also fit Gaussian, Gumbel and Frank copulas, estimate their parameters and perform a goodness of fit test to validate the automatically generated choice. This section is structured as follows: first, we present the copula equations and the estimated parameters, second we describe the model selection strategy based on a goodness of fit test along with results for our five pairs of assets, third we show the pseudo-samples plots, and finally we include the graphs of the fitted copulas.

## 4.1 Bivariate Copulas - Theoretical Framework

### 4.1.1 Gaussian Copula

The Gaussian copula is an implicit copula and can be expressed as an integral over the density of X (assuming X is a standardized version of Y). It describes a dependence structure between positive and negative dependence, where the $\rho$ shows the strength of the dependence.

$$C_\rho(u_1, u_2) = \int_{-\infty}^{\phi^{-1}(u_1)} \int_{-\infty}^{\phi^{-1}(u_2)} \frac{1}{(2\pi(1-\rho^2)^{1/2}} \exp\left(\frac{-(s_1^2 - 2\rho s_1 s_2 + s_2^2)}{2(1-\rho^2}\right)) ds_1 ds_2 \tag{1}$$

The normal distribution has some limitations when applied to financial data as it does not capture dependence in the tails and extreme outcomes, which are common properties of asset returns.

### 4.1.2 Student-t Copula

The student-t copula is another implicit copula, which captures better tail dependence in comparison to the Gaussian copula. Eq. 2 for the two-dimensional student-t copula is shown below:

$$C_{\nu,P}(u_1, u_2) = (t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) \tag{2}$$

### 4.1.3 Gumbel Copula

The Gumbel copula is also an implicit copula, which is restricted by the property that it cannot take any negative dependency. Since this copula captures the dependencies on the positive tail and does not allow negative values, we do not expect that the copula will perform well on our asset pairs.

$$C_\theta(u_1, u_2) = exp(-((-ln(u_1))^\theta + (-ln(u_2))^\theta)^{1/\theta}), 1 \leq \theta < \infty \tag{3}$$

### 4.1.4 Frank Copula

The Frank copula is from the Archimedian family and has no restrictions on the $\theta$ parameter, hence it can be regarded as a rather flexible copula. This means we expect that the copula will always provide a fit, but this does not necessarily imply that it will provide the most optimal fit.

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} ln[\frac{(1-e^{-\theta}) - (1-e^{-\theta u_1})(1-e^{-\theta u_2})}{2}], -\infty < \theta < \infty \tag{4}$$

### 4.2 Bivariate Copulas - Results

The parameters of the estimated copulas are presented in Tables 1 and 2. Table 1 shows the parameters of the student-t copula and Table 2 shows the parameters of the other copulas. The next section discusses the goodness of fit.

| | $\rho$ | $\nu$ |
|---|---|---|
| LHA_ADS | 0.36 | 5.60 |
| LHA_HEIA | 0.28 | 5.10 |
| LHA_N225 | 0.18 | 6.58 |
| ADS_N225 | 0.17 | 7.98 |
| ADS_TSLA | 0.23 | 10.11 |

**Table 1.** Student-t copula parameters

| | Gaussian | Gumbel | Frank |
|---|---|---|---|
| | $\rho$ | $\alpha$ | $\alpha$ |
| LHA_ADS | 0.35 | 1.28 | 2.29 |
| LHA_HEIA | 0.28 | 1.20 | 1.70 |
| LHA_N225 | 0.17 | 1.11 | 1.05 |
| ADS_N225 | 0.17 | 1.10 | 0.99 |
| ADS_TSLA | 0.22 | 1.15 | 1.32 |

**Table 2.** Gaussian, Gumbel and Frank copula parameters

### 4.3 Model Selection: Goodness of Fit

We use BiCopGofTest function in R [2] to estimate the goodness of fit for the biavariate copulas by comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis [1]. The p-values are calculated using a parametric bootstrapping procedure. P-values $< 0.05$ reject the null hypothesis and indicate that the copula is not a good fit, while p-values $> 0.05$ show that the copula is a good fit for the dependence structure of the empirical copula with the estimated parameters. The results of the goodness of fit test on the five pairs of asset for each of the tested copulas are reported in Table 3. We confirm the results from the automatic test that the student-t copula provides the best fit with p-values consistently above 0.05 for each pair of assets. None of the other copulas are a good fit.

| | Student-t | Gaussian | Gumbel | Frank |
|---|---|---|---|---|
| LHA_ADS | 0.12 | 0.00 | 0.00 | 0.00 |
| LHA_HEIA | 0.10 | 0.00 | 0.00 | 0.00 |
| LHA_N225 | 0.43 | 0.00 | 0.00 | 0.00 |
| ADS_N225 | 0.20 | 0.00 | 0.00 | 0.00 |
| ADS_TSLA | 0.26 | 0.00 | 0.00 | 0.02 |

**Table 3.** Goodness of Fit (p-values)

Goodness of fit test. P-value 0.05 shows that the copula is a good fit.

### 4.4 Pseudo-samples and density plots

Figures 11 and 12 show the samples contained in vector when the copula is built from random samples based on student-t and Gaussian copulas respectively. For brevity, we include figures only for one pair as the figures look similarly for the rest of the pairs. The similarity is due to the fact that the generated samples have identical correlation as the original data and the pairs of assets have a similar correlation in the range of 0.17-0.34. When the correlation between the assets is low, we observe rather chaotic dispersion of the generated observations across the plot. Therefore, in order to observe easier the patterns on the graph, we choose to present the pseudo-samples plots for LHA/ADS pair, which has highest correlation (0.34) (Table 4). On the student-t plot we can see slight concentrations in the left and right corner and less noticeable concentration along the diagonal, which reflects that tail dependence structure, as discussed previously. Figures 13 and 14 are the 3D density plots of the simulated data for student-t and Gaussian copulas.

| | Correlation |
|---|---|
| LHA_ADS | 0.34 |
| LHA_HEIA | 0.29 |
| LHA_N225 | 0.17 |
| ADS_N225 | 0.18 |
| ADS_TSLA | 0.21 |

**Table 4.** Asset correlations

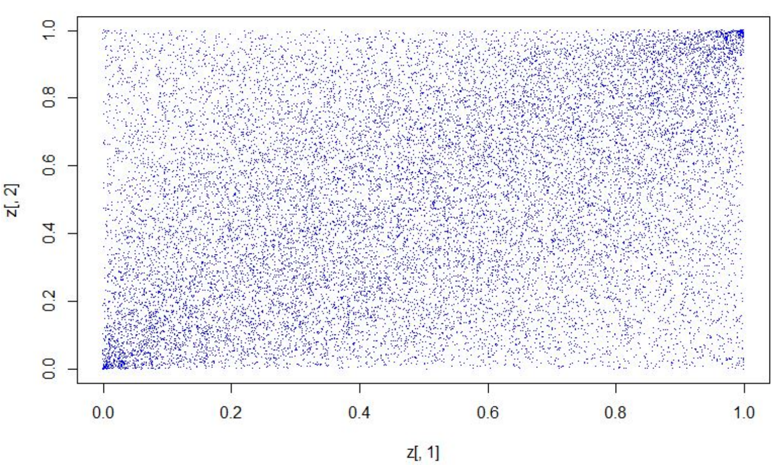



**Figure 11.** Pseudo samples generated with Student-t copula for the LHA/ADS pair

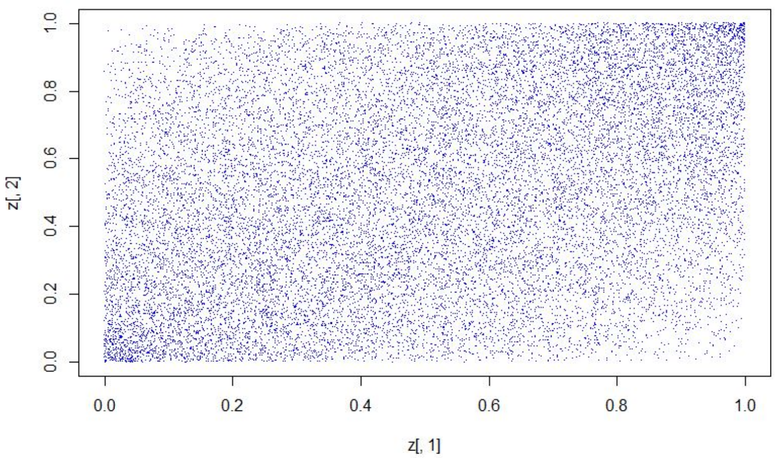



**Figure 12.** Pseudo samples generated with Gaussian copula for the LHA/ADS pair

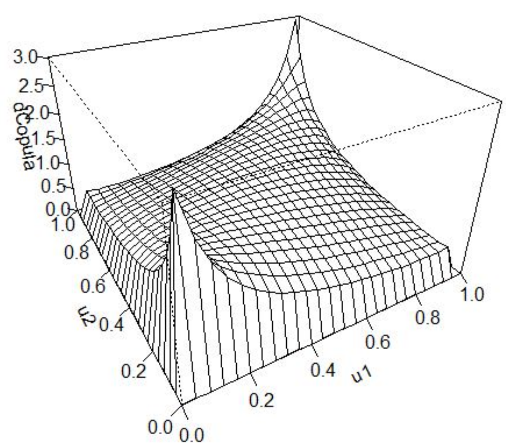



**Figure 13.** 3D plot of the simulated data with Student-t copula for the LHA/ADS pair

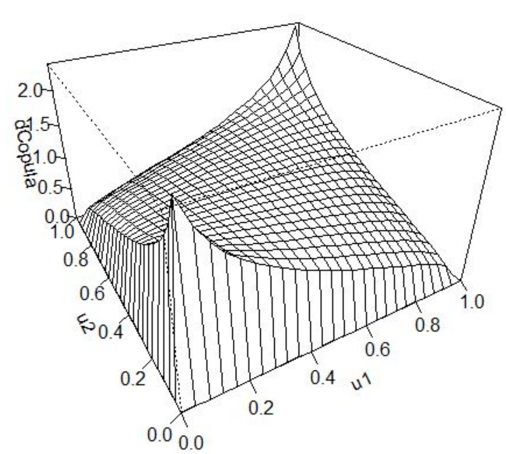



**Figure 14.** 3D plot of the simulated data with Gaussian copula for the LHA/ADS pair

### 4.5 Fitted Copulas

Finally, in figures 15 - 19 we present the scatterplots of the data for the two pairs under the assumption of normal marginals and a t-copula for the dependence structure.

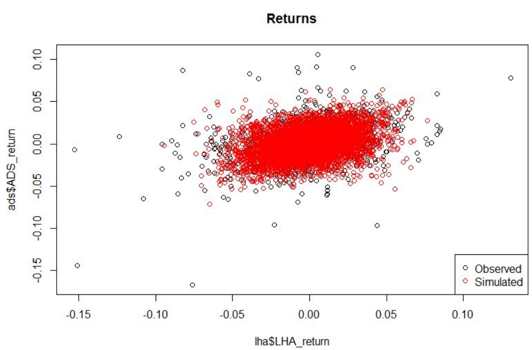



**Figure 15.** Fitted Copula LHA/ADS

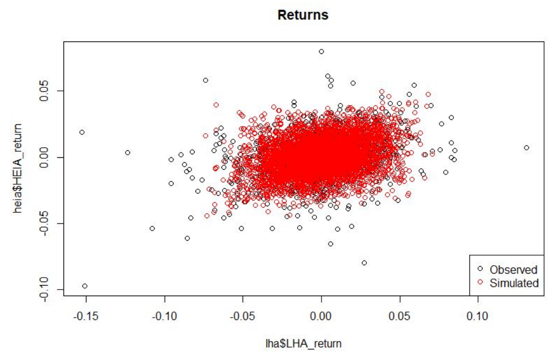



**Figure 16.** Fitted Copula LHA/HEIA

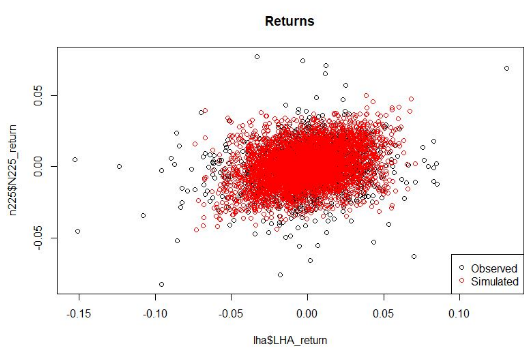



**Figure 17.** Fitted Copula LHA/N225

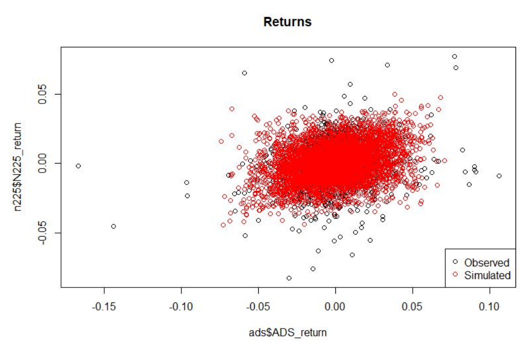



**Figure 18.** Fitted Copula ADS/N225

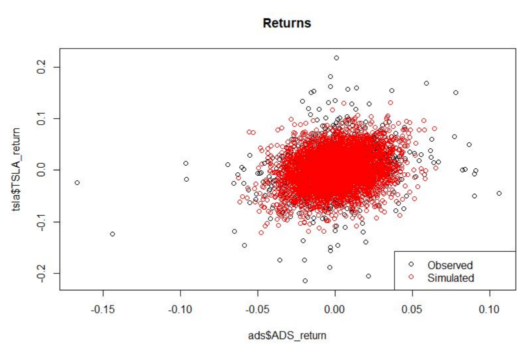



**Figure 19.** Fitted Copula ADS/TSLA

We conclude that the t-copula leads to predictions close to the observed values, but there are some returns which the model failed to capture. This is problematic for the more extreme results and should be addressed by outlier investigation or through model calibration.

## 5 Extreme Value Theory (EVT)

To determine which asset to use for analysis, we first plotted Q-Q plots for each asset with respect to normal distribution. From figure 20 we can see that both SP500 (GSPC) and Bitcoin (BTC) have heavy tails regarding the upper tail. The kurtosis, 20.44 for GSPC and 21.25 for BTC cannot tell which one has heavier tail. Therefore we did the analysis on both assets and found out GSPC has heavier tail due to higher k (xi) value.

We chose the 97.5 percentile of the returns as the threshold. The number of exceedance is 52 data points. We then fitted the generalized pareto distribution (GPD) to the difference of the tail data and the threshold (the exceedance). Although there are some noises, the exceedance follows the GPD curve, this indicates that GPD is a good fit for the exceedance values (figure 21). In addition, we can see that the distribution of GSPC is indeed heavy-tailed with some values in the tail. For a better visualization of the overall probability density fitness of the model, we plotted the empirical cumulative distribution function (CDF) with the one fitted with GDP. It clearly shows that GDP fits the tail data well (figure 22).

Using the maximum likelihood method to estimate k (shape parameter) and $\sigma$ (scale parameter), we have estimators (k, $\sigma$) = (0.4283, 0.8898) where the standard error (SE) being (0.2129, 0.2198). While k $>$ 0 indicates the heavy-tailed nature of the asset, the higher the value, the higher the tail. 0.4283 confirms that GSPC has heavier tail than BTC, where k = 0.1422. From SE we saw that accuracy of k is less than $\sigma$ . Interpretation of the standard errors is based on the assumption that, if the same fit could be repeated many times on data that came from the same source, the maximum likelihood estimates of the parameters would approximately follow a normal distribution. Therefore, we used bootstrap simulation by generating 1,000 replicate datasets by resampling from the data, fit a GP distribution to each one, and calculated all the replicate estimates. The estimates of k is close to the normal distribution while $\sigma$ is skewed to the right (figure 23). Since $\sigma$ is not symmetric, we transformed the estimates with the log scale. After applying this transformation, from figure 24 we can indeed observe that both parameters appear acceptably close to normality. By constructing the confidence interval for $\sigma$ by first computing one for log(sigma) under the assumption of normality, and then exponentiating to transform it back to the original scale for $\sigma$, we obtained a CI of (0.5483,1.4441). CI of k is (0.0110,0.8456).

**VaR and ES of three methods**

By using EVT, variance-covariance method with t-distribution with df = 3 and historical simulation method, we obtained the VaR and ES for 99 percentile shown in Table 5. We can see that both EVT and HS close VaR and ES and both are higher than the one of t-distribution. This concludes that EVT is suitable for simulating the risk measurements of the portfolio in a way that it is able to capture the extreme event factor and correct back to produce a more promising outcome.

| | VaR (%) | ES (%) |
|---|---|---|
| EVT | 3.3 | 5.6 |
| t-distribution | 2.9 | 4.75 |
| HS | 3.24 | 5.39 |

**Table 5.** VaR and ES of three simulation methods.

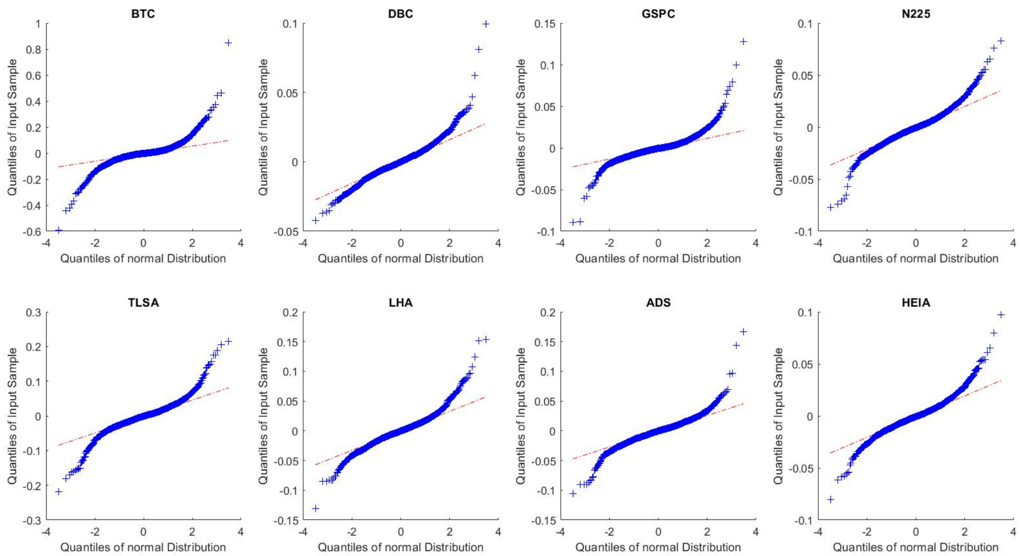



**Figure 20.** The Q-Q plots using normal distribution with respect to eight assets where GSPC has the heaviest tail.

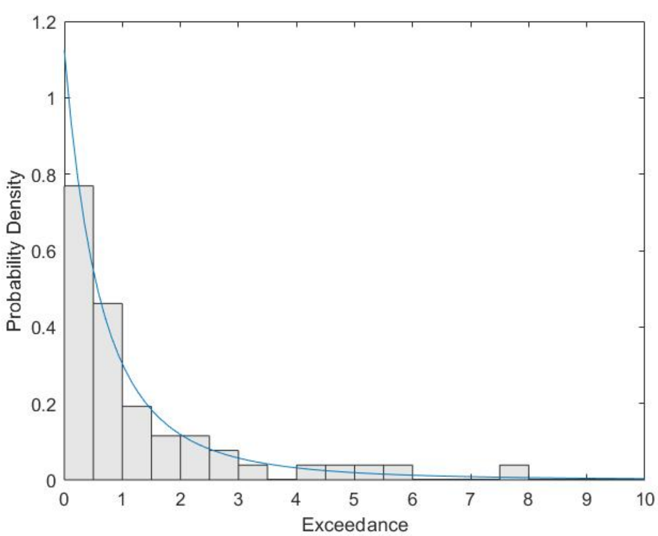



**Figure 21.** The distribution of exceedances of GSPC and the fitted generalized pareto distribution curve.

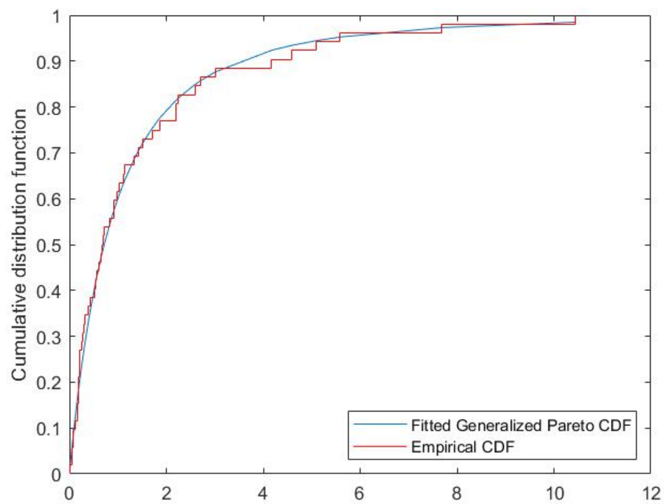



**Figure 22.** The cumulative distribution function curves of the empirical data and the simulated data from GPD.

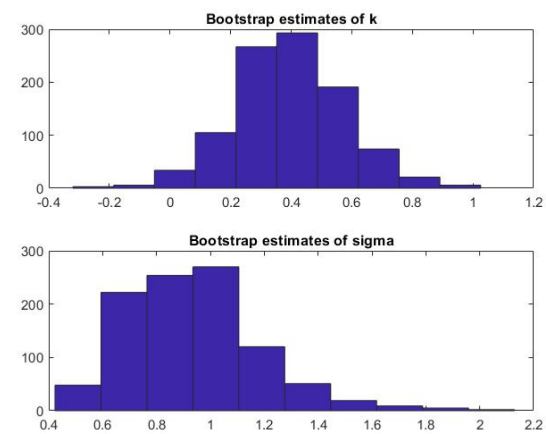



**Figure 23.** The histograms showing the distributions of both k and $\sigma$ using bootstrap simulations.

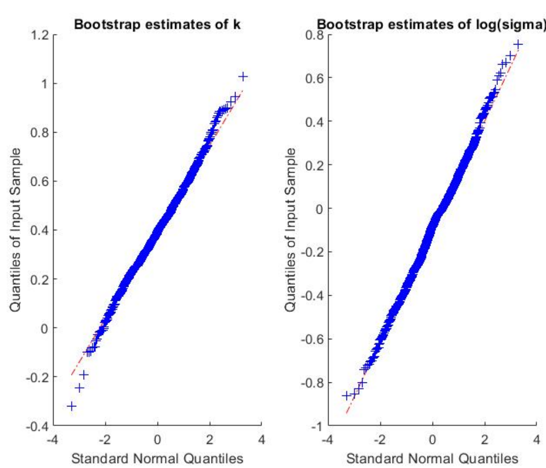



**Figure 24.** The Q-Q plots showing the normality of both parameters, k and the log-scaled $\sigma$.

## 6 Bibliography

[1] KOJADINOVIC, I., AND YAN, J. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing 21*, 1 (jul 2009), 17–30.

[2] NAGLER, T., AND SCHEPSMEIER, U. Vinecopula: Statistical inference of vine copulas, 2019.