5. Principal Component Analysis

5.1 Correlation Analysis

The correlations between the returns of different assets can give us a direct impression of the relationships among the assets. We can find out which assets have positive relationship on returns and which ones have negative relationship, as well as which ones almost have no relationship. Based on the estimated correlation values, we can also find out whether diversification will reduce risk with these assets. Now, we can look at the correlation matrix and a virtualization plot of the correlation as follows:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **AMD** | **AAPL** | **BAC** | **COKE** | **FCX** | **F** | **GE** | **MDT** | **MRO** | **PFE** | **SIRI** | **SBUX** | **X** | **VALE** | **VZ** |
| **AMD** | **1.000** | **0.311** | **0.411** | **0.063** | **0.365** | **0.456** | **0.503** | **0.452** | **0.328** | **0.148** | **0.304** | **0.482** | **0.263** | **0.239** | **0.009** | |
| **AAPL** |  | **1.000** | **0.267** | **0.160** | **0.350** | **0.270** | **0.410** | **0.260** | **0.322** | **0.099** | **0.303** | **0.354** | **0.211** | **0.249** | **0.092** | |
| **BAC** |  |  | **1.000** | **0.083** | **0.325** | **0.479** | **0.518** | **0.411** | **0.352** | **0.288** | **0.356** | **0.310** | **0.425** | **0.365** | **0.002** | |
| **COKE** |  |  |  | **1.000** | **0.115** | **0.157** | **0.194** | **0.167** | **0.080** | **0.153** | **0.023** | **0.209** | **0.099** | **0.035** | **-0.018** | |
| **FCX** |  |  |  |  | **1.000** | **0.550** | **0.512** | **0.306** | **0.664** | **0.289** | **0.365** | **0.332** | **0.576** | **0.727** | **0.320** | |
| **F** |  |  |  |  |  | **1.000** | **0.473** | **0.472** | **0.396** | **0.281** | **0.533** | **0.459** | **0.425** | **0.471** | **0.123** | |
| **GE** |  |  |  |  |  |  | **1.000** | **0.651** | **0.512** | **0.490** | **0.353** | **0.377** | **0.396** | **0.475** | **0.460** | |
| **MDT** |  |  |  |  |  |  |  | **1.000** | **0.380** | **0.616** | **0.432** | **0.406** | **0.346** | **0.149** | **0.322** | |
| **MRO** |  |  |  |  |  |  |  |  | **1.000** | **0.356** | **0.394** | **0.286** | **0.422** | **0.550** | **0.213** | |
| **PFE** |  |  |  |  |  |  |  |  |  | **1.000** | **0.356** | **0.325** | **0.289** | **0.171** | **0.476** | |
| **SIRI** |  |  |  |  |  |  |  |  |  |  | **1.000** | **0.310** | **0.218** | **0.326** | **0.009** | |
| **SBUX** |  |  |  |  |  |  |  |  |  |  |  | **1.000** | **0.273** | **0.246** | **0.109** | |
| **X** |  |  |  |  |  |  |  |  |  |  |  |  | **1.000** | **0.545** | **0.262** | |
| **VALE** |  |  |  |  |  |  |  |  |  |  |  |  |  | **1.000** | **0.361** | |
| **VZ** |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **1.000** | |

Table 5.1: Correlation Matrix of the 15 Assets

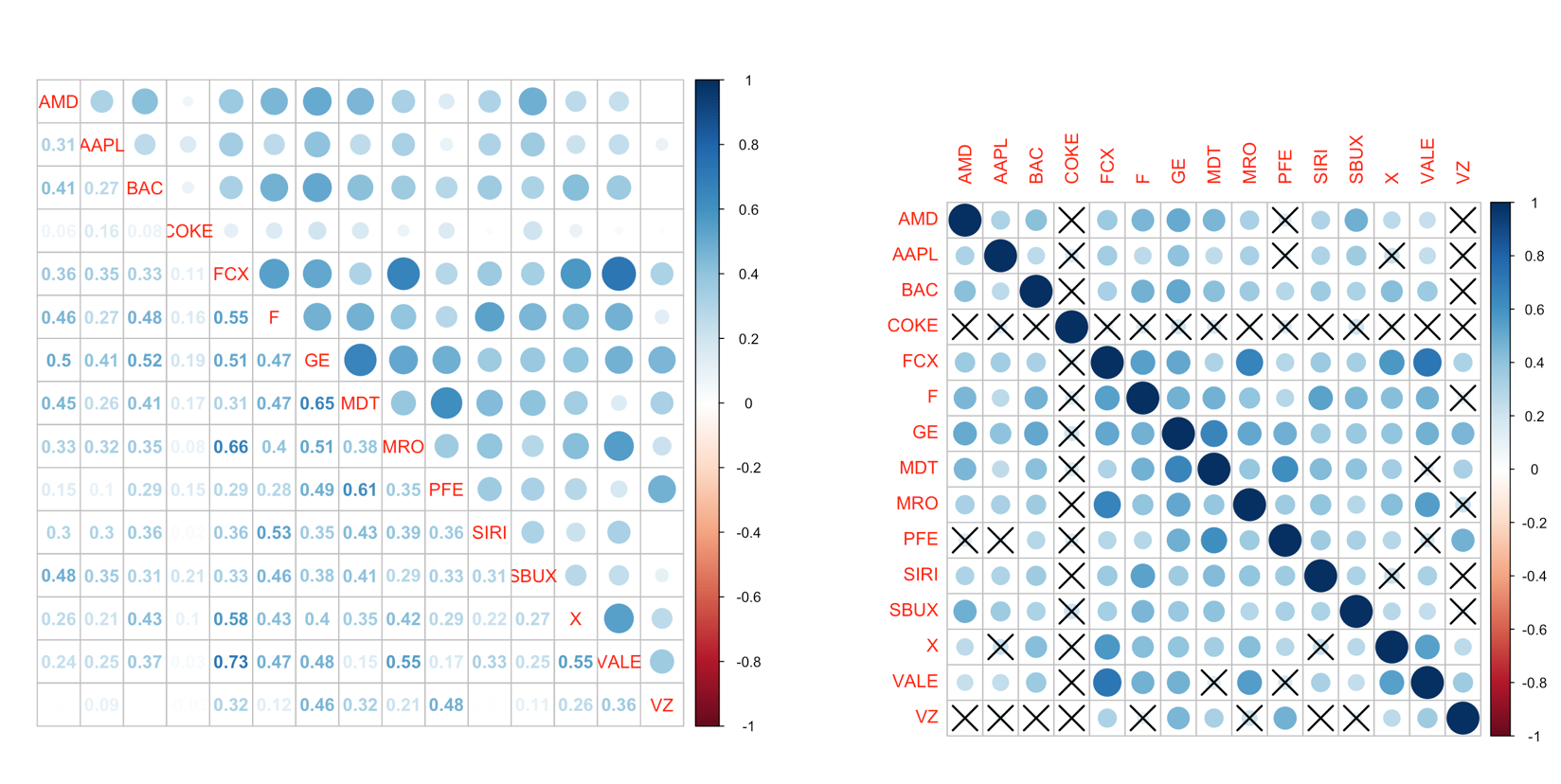


Fig 5.1: Correlation Plot with Number (left) and Correlation Plot with Significance Test (right)

Firstly, there is an obvious phenomenon that the “Cock-Cola” has no significant correlation with any other company. This can tell us that the stock price of Coca-Cola would not fluctuate with other companies, which means Coca-Cola maybe come from a different industry with other companies. Although we Coca-Cola and Apple, as well as Ford belong to Consumer Goods section together, this kind of Beverage would not be influenced by the others. It may imply that Coca-Cola has become a stable company which is necessary to us. Secondly, Verizon is another company which was not influenced by other companies. All the correlations are less than 0.5, and the highest correlation is 0.48 with Pfizer which is a Drug Manufacturers company.

Now, it’s time to focus on some high correlations. We can find that most of correlations between Freeport-McMoRan, which is a copper company, and other companies are relatively high. And the highest correlation appears between FCX and Vale which is a metals and minerals company. The FCX also has high correlations with the Marathon Oil which is an Oil & Gas company, the United States Steel which is obviously a Steel & Iron company, as well as Ford which is an Auto Manufactures company. This phenomenon indicates that the Basic Materials industry has a relatively large influence to other industries, including other metals’ price.

What’s more, the correlation between General Electric and Medtronic is 0.65. It seems that an electronic company should not have such a high relationship with a healthcare company. But one reason is that GE is a diversified machinery company, which involves in various industry including the medical equipment company. And the other reason is that if we search “GE and Medtronic” online, we can find that GE Healthcare has a five-year-of collaboration with Medtronic, this definitely can explain why the two companies have a high correlation. Another pair of high correlation is Medtronic and Pfizer, a medical appliances & equipment company and a drug manufacturers company. These two companies come from two related industry. So that it is reasonable to have a high correlation.

Based on all the analysis above, we can find that most of the correlations among these 15 companies are less than 0.5, which implies a low correlation. So that buying a portfolio of these stocks can obtain a lower risk than just buying one of them. On the other hand, since there is no negative correlation, it means that any two of these companies do not come from contrary industries. Thus, the reduction of risk will be limited, which means there exists better portfolios if we add some other stocks.

5.2 Principal Component Analysis

Now, it’s time to turn to the PCA. When we run Principal Component Analysis, our goal should be reducing the dimension of the variable. So that every principal component is a linear combination of all variables and explains a part of the variance of the whole data. Then how many PCs should we keep?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PC1 | PC2 | PC3 | PC4 | PC5 |
| AMD | -0.223 | 0.391 | 0.016 | -0.221 | -0.278 |
| AAPL | -0.062 | 0.186 | 0.245 | -0.644 | 0.263 |
| BAC | -0.284 | 0.239 | -0.027 | 0.191 | -0.313 |
| COKE | 0.332 | 0.072 | 0.167 | 0.162 | -0.232 |
| FCX | -0.388 | -0.257 | 0.151 | -0.050 | 0.033 |
| F | -0.346 | 0.221 | 0.053 | 0.319 | 0.009 |
| GE | -0.255 | 0.025 | -0.369 | -0.379 | -0.219 |
| MDT | -0.089 | 0.243 | -0.516 | 0.001 | -0.035 |
| MRO | -0.365 | -0.160 | 0.015 | -0.100 | 0.222 |
| PFE | 0.033 | -0.066 | -0.564 | 0.157 | 0.149 |
| SIRI | -0.239 | 0.254 | -0.045 | 0.286 | 0.644 |
| SBUX | -0.063 | 0.380 | 0.016 | -0.134 | -0.134 |
| X | -0.302 | -0.243 | 0.062 | -0.384 | -0.384 |
| VALE | -0.354 | -0.325 | 0.179 | -0.158 | -0.016 |
| VA | 0.031 | -0.415 | -0.356 | -0.014 | -0.014 |
| Proportion of Variance | 0.276 | 0.196 | 0.171 | 0.078 | 0.067 |
| Cumulative Proportion | 0.276 | 0.473 | 0.643 | 0.721 | 0.788 |

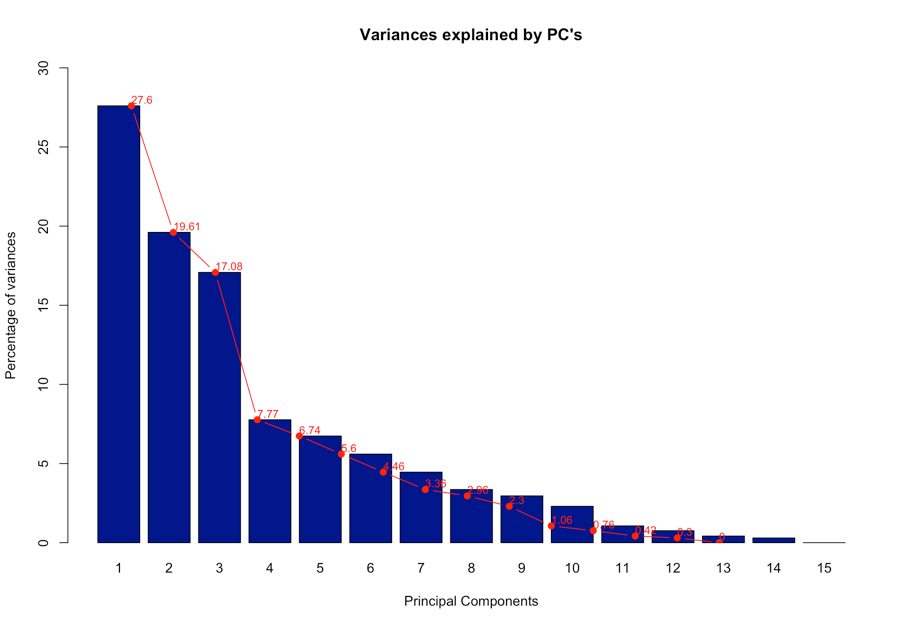


Table 5.2: PCA loadings and explained variance Fig 5.2: PCA Scree Plot

We can simply find that the first three principal components explained 64.3% variance and contributed most to the variance explanation. And variance explanation of the 4th PC decreased significantly. But containing the 4th PC may increase the total variance explanation to 72.1%. So, it will be reasonable to keep the first three or four principal components and no more principal components should be included. Now, we can do more analysis to the first three principal components.

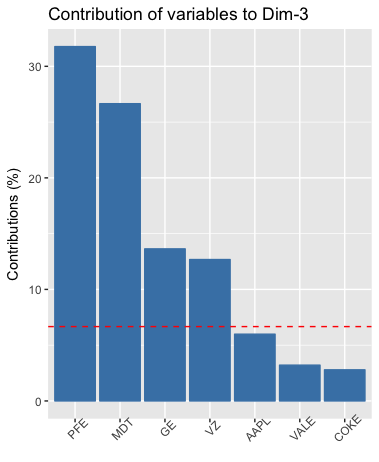
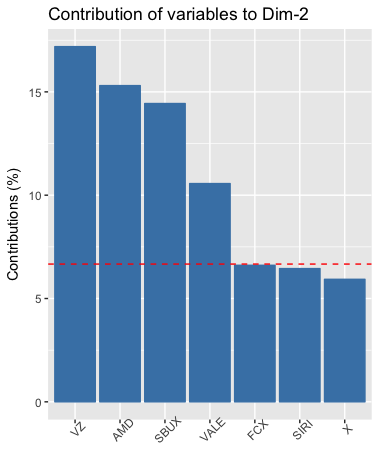
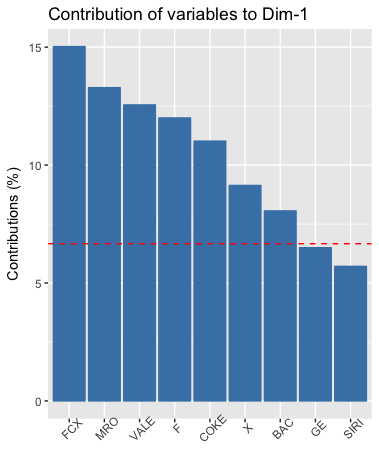
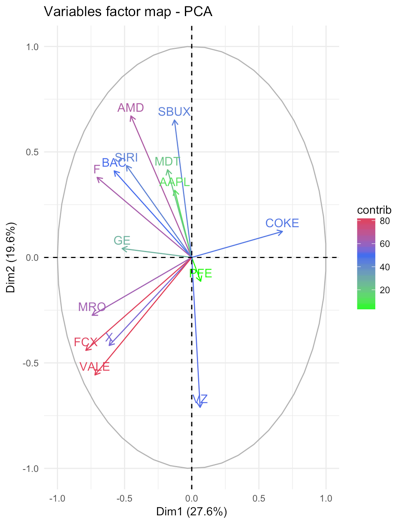


Fig 5.3: Projection of the first two PCs (1), Contribution of each variable to the PC1-PC3 (2-4)

In the first plot, we can find out which companies contribute to the first two PCs more, which means more important than other companies. It is clear that FCX and VALE contribute most to the variance of first two PCs. And Ford, AMD as well as MRO contribute a little bit less. And next are SBUX, SIRI, COKE, VZ and BAC. These phenomena can provide us a sense of the financial market. First, the fluctuation of basic materials industry may cause significant change to other industries. So, when we do research or find opportunities in financial market, the price of basic materials like copper, steel and oil may be concerned prior. Second, the fluctuation of basic materials may influence the manufacturer industry first because they use basic material in production.

On the other hand, the companies like PFE and APPLE have little contribution to the first two PCs. But as we can find from the results above, PFE contributed most to PC3 and APPLE contributes most to PC4. This situation told us that healthcare and consumer goods also have huge influence to the financial market, but they influence the market from different aspect.

5.3 Factor Analysis

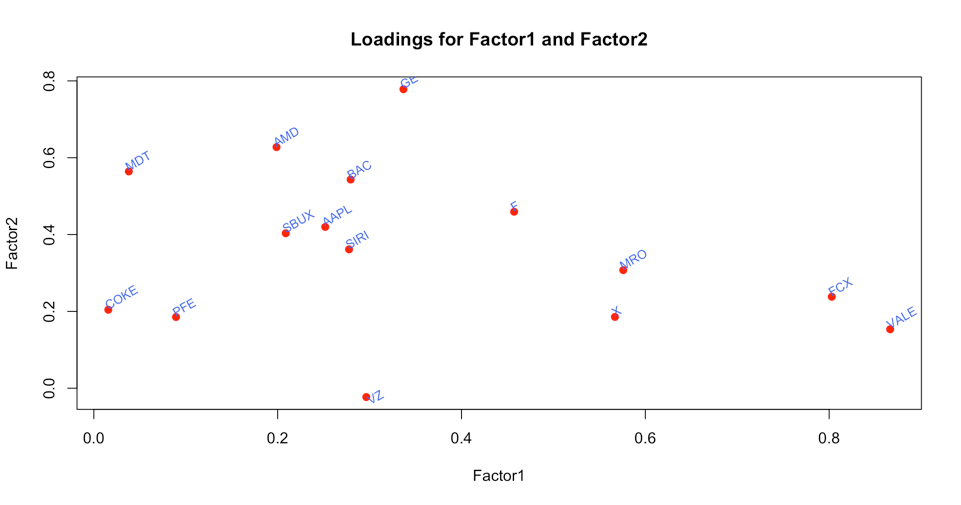
Factor analysis, in some extent, is like the PCA, but provides a more direct way to explain the results. So, for the convenience of explanation, it is important to choose how many factors to use. Thus, we did factor analysis for the number of factors equals to 2, 3, 4 and 5. Here is a brief summary of the results and the specific results will be included in the Appendix.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number of Factors | 2 | 3 | 4 | 5 |
| Cumulative Variance | 0.428 | 0.511 | 0.538 | 0.59 |
| p-value for the test of factor number sufficiency | 0.0309 | 0.58 | 0.724 | 0.748 |

Table 5.3: Summary of the results for Factor Analysis

From the table above, we can choose the number of factors. Firstly, the null hypothesis of the test is: the number of factors is sufficient. So, only when the p-value of the test exceeds 0.05, the number of factors will be reasonable. Thus, 2 factors are not sufficient in this situation. Secondly, in order to explain the results better, it is necessary to choose a number of factors that can explain more variance of the data. So that we decide to choose 4 factors as our result. This is because, 4 factors explain more variance than 3 factors and the p-value of the test increase significantly from 0.58 to 0.724. On the other hand, although 5 factors can explain more variance than 4 factors, it increased the p-value by only 0.02. Thus, 4 factors seem to be the best choice. Now, let’s take a look at the results of the factor analysis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | FA1 | FA2 | FA3 | FA4 |
| FCX | 0.803 | 0.238 | 0.149 | 0.167 |
| MRO | 0.576 | 0.307 | 0.198 | 0.189 |
| X | 0.567 | 0.186 | 0.172 | 0.218 |
| VALE | 0.866 | 0.153 | 0.119 |  |
| AMD | 0.199 | 0.628 |  | 0.209 |
| BAC | 0.279 | 0.543 |  | 0.204 |
| GE | 0.337 | 0.778 | 0.511 | -0.121 |
| MDT |  | 0.564 | 0.490 | 0.420 |
| PFE |  | 0.185 | 0.714 | 0.411 |
| VZ | 0.296 |  | 0.710 | -0.119 |
| AAPL | 0.252 | 0.420 |  |  |
| COKE |  | 0.204 |  |  |
| F | 0.457 | 0.459 |  | 0.405 |
| SIRI | 0.278 | 0.361 |  | 0.443 |
| SBUX | 0.209 | 0.403 |  | 0.370 |
| Proportion of Variance | 0.185 | 0.172 | 0.109 | 0.072 |
| Cumulative Proportion | 0.185 | 0.357 | 0.466 | 0.538 |



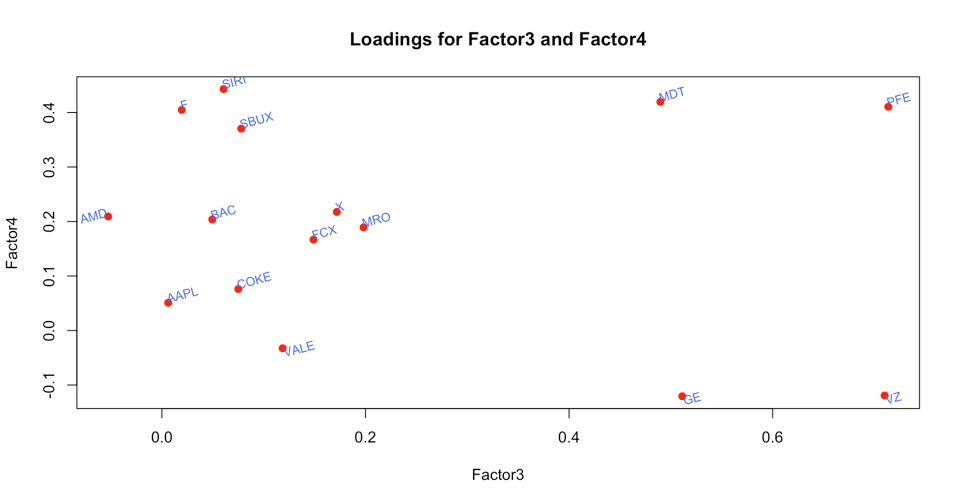


Table 5.4: Results of Factor Analysis

Fig: 5.4: Factor Loadings Plot

From the table and plot above, we can find the loadings of FCX, MRO, X and VALE are relatively high in factor 1 than any other factors. As we mentioned before, these companies belong to copper, oil, steel, and metals industries respectively, this tells us the first factor is an index of basic materials companies. The prices of the stocks of these companies influence the financial market most. At the same time, factor2 has similar influence to the financial market. And the loadings of AMD, BAC, GE, MDT, APPLE, COKE, FORD, as well as SBUX are all relatively high than any other factors. These companies involve in all kinds of industries. But the common property of these companies is that they all relate to consumption, which means the factor2 can be regarded as a consumption index.

Then, let’s take a look at factor 3, the loadings of VZ and PFE are much larger than other factors. These two companies are in the Drug Manufacturers and Telecom Services industries. Obviously, drug and telecom are two necessary things in our life, so that the factor3 can be seen as an index of rigid demands, which means people have to pay for it sometime. Finally, the factor4, the only one loading which is larger than others is the loading of SIRI, which is a broadcasting company. We cannot define it in a clear way because we have no more relevant companies. But as a guess, we may consider it as an index of entertainment which also play a role in people’s life.

7. Copulas

We compared 5 kinds of copulas: Normal copula, T copula, Clayton copula, Gumbel copula and Frank copula to find out the best model of the joint distribution of the returns. And the results can be compared by various criteria like AIC, BIC and Likelihood which are showed below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Copula | Normal | T | | Clayton | | Gumbel | Frank |
| AIC | -297 | | -300 | | -236 | -244 | -236 |
| BIC | -295 | | -296 | | -234 | -242 | -234 |
| Likelihood | 150 | | 152 | | 119 | 123 | 119 |

Table 7.1: Copulas Comparison Results

From the values of AIC, we find T copula has lowest value, and BIC also provides the same result. For AIC and BIC, lower values mean that that copula fits better. And the larger the likelihood is, the better the model is. Since the likelihood is 152 for the T copula, so we conclude that T copula fits better than others.