**DATA AND EMPIRICAL RESULT**

Our steps to apply GARCH-EVT COPULA method are described as the following:

**Step1**

Having index prices CTG, MSN, VIC, VNM,

Transform them to log returns (ri) with i = 1, ..., 4

Fit the ARMA-GJR-GARCH model to each log-return of companies and then we get the standardized residuals for each company.

**Step2**

From the standardized residuals we will find the optimal marginal distributions for each of them by POT method. In particular, we fit the Generalized Pareto Distribution for the 2 tails and Gaussian Distribution for the interior.

**Step3**

From the standardized residuals, we transform them to the Uniform data using marginal distribution of each company found from the Step 2. (F(xi)=Ui)

**Step4**

Fit a Copula to the transformed margins from Step 3 and estimate the Copula parameters.

**Step5**

From the estimated Copula parameters (Clayton Copula, Student-t Copula, Clayton Copula, Gumbel Copula and Frank Copula), we generate 10,000 scenario for each company over the testing period.

Transform to the original scales of the log returns using the inverse quantile

function of the marginals, i.e F−1 i(Ui).

**Step6:**

Reintroduce the autocorrelations and heteroscedasticity observed in the original

returns using again the ARMA(p,q)-GJR\_GARCH(1,1) model to get log-return which corresponds to the

simulated returns for each corresponding marginal distribution.

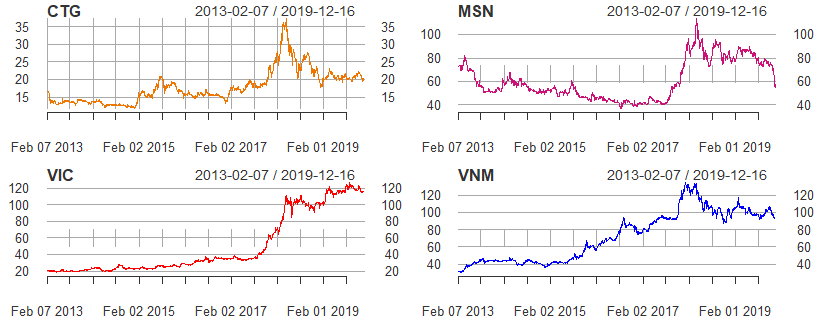
**Step7:**

Using the rolling prediction technique, we will predict the mean and sigma in the validation data.

The one day value at risk at time t with confidence level α (99%, 95% and 90% ),

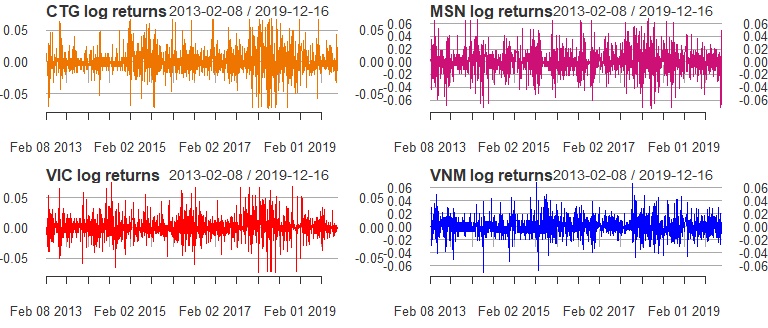
V aRt(α) is just the 1- α quantile of the distribution of the discounted log-return series.

In this article, we will demonstrate the backtesting procedure in order to compare the daily log-return and daily VaR (90%,95%, 99%). The full dataset is collected from 7 February 2013 to 17th December 2020 of all 4 stock price (CTG, MSN, VIC, VNM), we use the data from 7 February 2013 to 16th December 2019 as the train data and the rest is the validation data used for backtesting procedure, the price of the train data are illustrated as following:

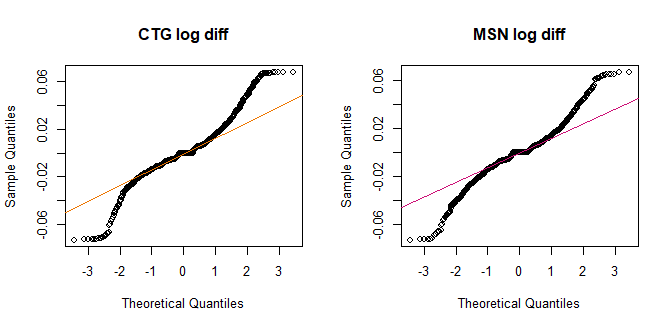


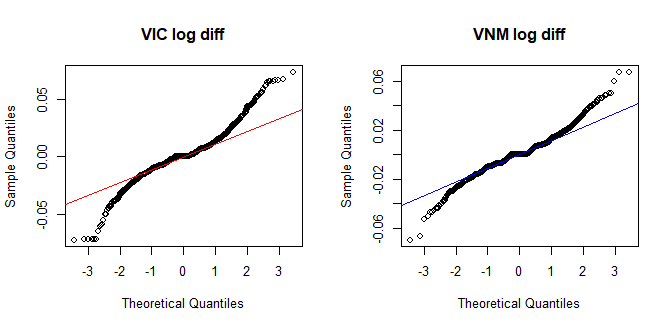
(Figure 1: the daily stock price of 4 companies: CTG, MSN, VIC, VNM)

The daily log-returns are calculated for each company. The daily log-return for each of them are shown as below:



(Figure 2: The daily log-return of 4 companies: CTG, MSN, VIC, VNM)





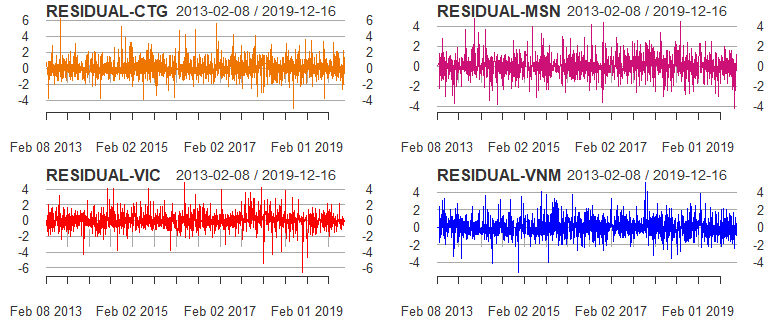
(Figure 3: the QQ-plot of 4 companies: CTG, MSN, VIC, VNM)

We subsequently fit the ARMA(p,q)-GJR\_GARCH(1,1) model for each log-returns. The parameters p and q are chosen by AIC criteria, which is shown in the following table, note that the (\*) denote the ARMA order that we choose.

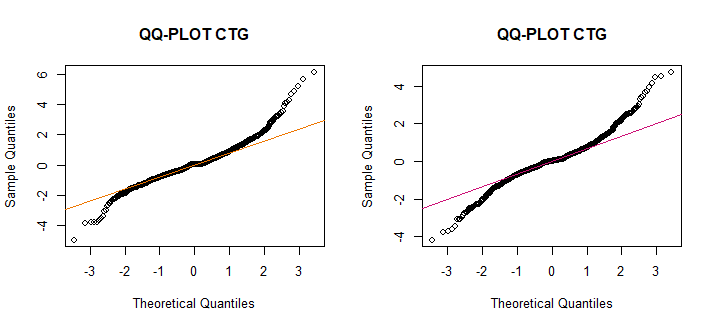
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Company** | **AR order** | **MA order** | **AIC** | **Company** | **AR order** | **MA order** | **AIC** |
| **CTG** | 0 | 0 | -8640.742 | **VIC** | 0 | 0 | -9144.221 |
| 0 | 1 | -8640.588 | 0 | 1 | -9142.429 |
| 1 | 0 | -8640.76 | 1 | 0 | -9142.419 |
| 1 | 1 | -8663.517 (\*) | 1 | 1 | -9140.424 |
| 1 | 2 | -8661.785 | 1 | 2 | -9139.631 |
| 2 | 1 | -8661.788 | 2 | 1 | -9139.768 |
| **MSN** | 0 | 0 | -8737.249 | 2 | 2 | -9146.417 (\*) |
| 0 | 1 | -8736.116 | 2 | 3 | -9144.64 |
| 1 | 0 | -8736.025 | 3 | 3 | -9142.852 |
| 1 | 1 | -8734.743 | **VNM** | 0 | 0 | -9725.689 |
| 1 | 2 | -8737.179 | 0 | 1 | -9723.746 |
| 2 | 1 | -8738.099 | 1 | 0 | -9723.746 (\*) |
| 2 | 2 | -8747.399 (\*) | 1 | 1 | -9721.746 |
| 3 | 2 | -8747.177 | 1 | 2 | -9719.75 |
| 3 | 3 | -8747.938 |

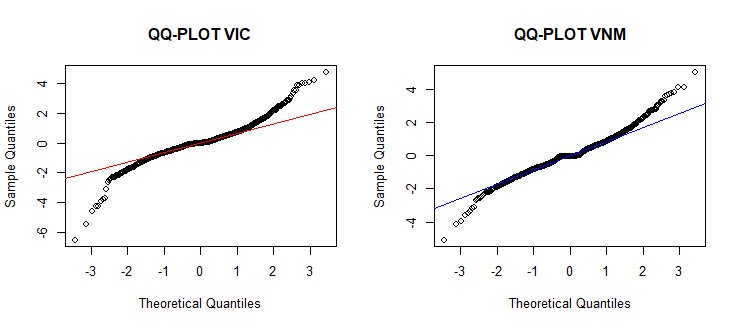
(Table 1: ARMA order of each company and their corresponding AIC.)

After fitting ARMA-GARCH model, we obtain the standardized residuals for each company (CTG, MSN, VIC, VNM). The comparison of them are shown below:

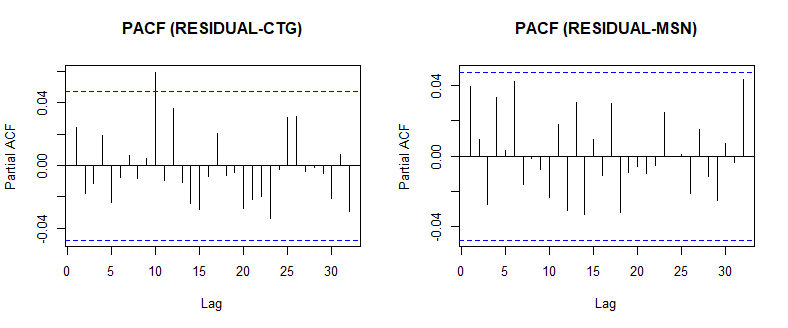
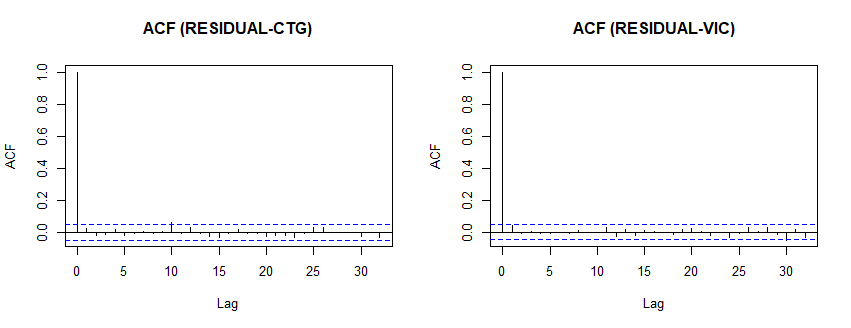


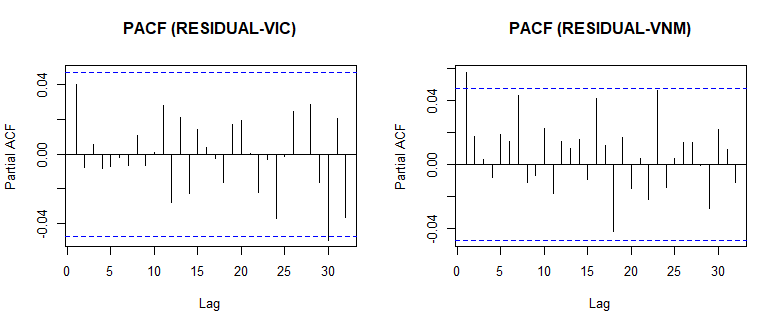
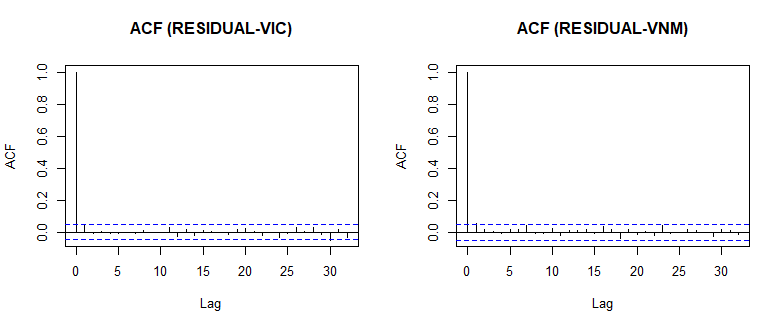
(Figure 4: The daily standardized residuals of 4 companies: CTG, MSN, VIC, VNM after fitting ARMA-GJR\_GARCH model)



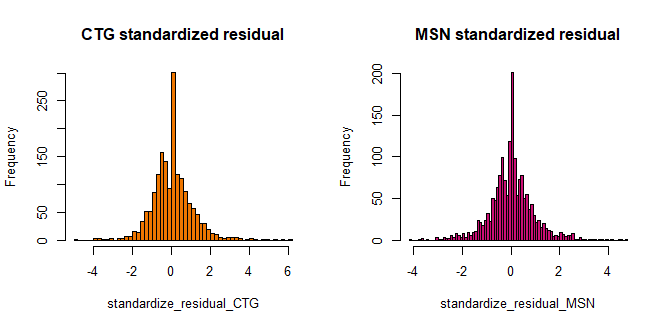


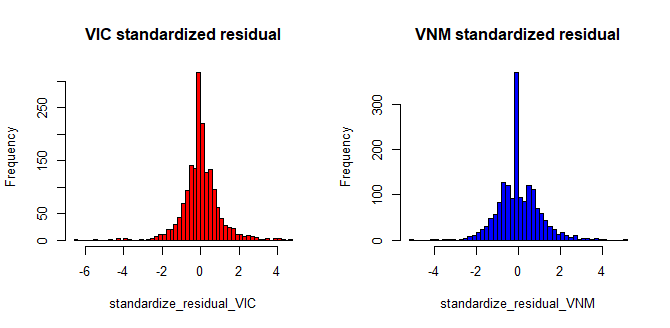
(Figure 5: The QQ-plot of 4 companies: CTG, MSN, VIC, VNM)





(Figure 6: The ACF and PACF of 4 companies: CTG, MSN, VIC, VNM)





(Figure 7: The histograms of 4 companies: CTG, MSN, VIC, VNM)

The mean, standard deviation, skewness and kurtosis of each company and its improvements are shown in the below table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CTG** | **MSN** | **VIC** | **VNM** |
| **Mean (return)** | 0.0001131556 | -0.0001543795 | 0.001002083 | 0.0006497589 |
| **SD(return)** | 0.0192109604 | 0.0186751973 | 0.016575402 | 0.0139782302 |
| **Skewness(return)** | 0.1671354803 | 0.0647890199 | 0.188440656 | 0.1985849527 |
| **Kurtosis(return)** | 2.8584628666 | 2.4148668679 | 3.446282661 | 2.4312796519 |
| **Mean (s.residual)** | 0.02377685 | 0.01849097 | 0.01777656 | 0.01196626 |
| **SD (s.residual)** | 1.010131 | 0.9325059 | 0.9139866 | 0.9843858 |
| **Skewness (s.residual)** | 0.5810241 | 0.2530943 | -0.02819653 | 0.2879513 |
| **Kurtosis (s.residual)** | 3.666771 | 3.016954 | 5.586374 | 2.249549 |

(Table 2: The mean, sd, skewness and kurtosis of 4 companies)

Since the log-returns of all 4 companies do not entirely fit the QQ-plot (Figure 4), they are all merely positive-skewed and playtykertic (kurtosis<3) , except VIC’s log-returns which is leptokurtic (Table 2), we can make an inference that these log-returns do not come from a normal distribution. However, in order to confirm our observation, we perform the Jarque-Bera Test statistic below, which are all rejected at 5% significance level.

The Portmanteau test of Ljung and Box acts as an effective tool to test whether the autocorrelation with different lags are zero, where as The Engle ARCH’s test is standard for checking autoregressive conditional heteroscedasticity. Table 3 below represents these 2 tests at 5% significance level.

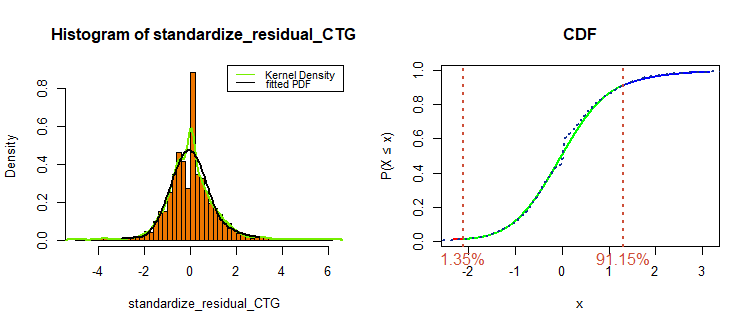
Note that in the below table, “R” represents reject and “A” represent accept.

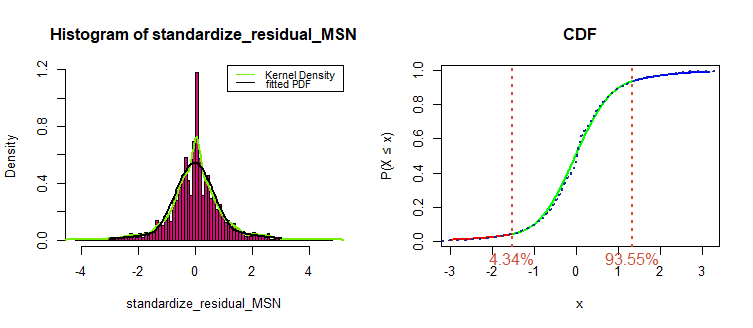
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **CTG** | **MSN** | **VIC** | **VNM** | **Test result** |
| **Jarque-Bera** | 588.7511 | 415.7218 | 854.3437 | 431.3951 | (RRRR) |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **Ljung-Box (return)** | 2.0172279 | 0.7753519 | 0.1981658 | 0.05706342 | (AAAA) |
| 0.1555228 | 0.3785665 | 0.6562054 | 0.81119901 |
| **Ljung-Box (s.residual)** | 1.0002842 | 2.6630461 | 2.80349758 | 5.65824846 | (AAAR) |
| 0.3172417 | 0.1027039 | 0.09405892 | 0.01737346 |
| **Engle’s ARCH (return)** | 9.699857e+01 | 6.083060e+01 | 4.401056e+01 | 5.992064e+01 | (RRRR) |
| 6.937740e-23 | 6.220356e-15 | 3.266093e-11 | 9.876036e-15 |
| **Engle’s ARCH (s.residual)** | 0.2759719 | 0.5593682 | 0.7976019 | 0.06221214 | (AAAA) |
| 0.5993536 | 0.4545149 | 0.3718113 | 0.80303312 |

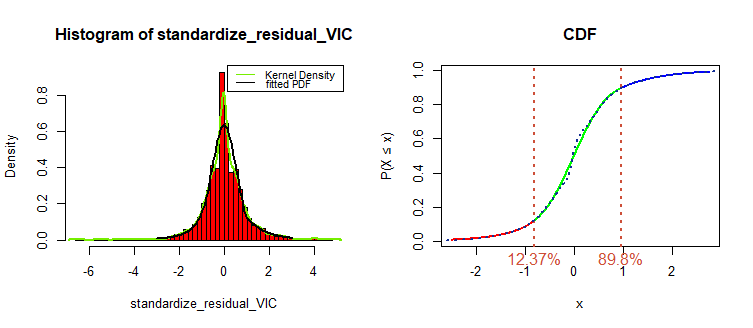
(Table 3: The statistics testing results for each company)

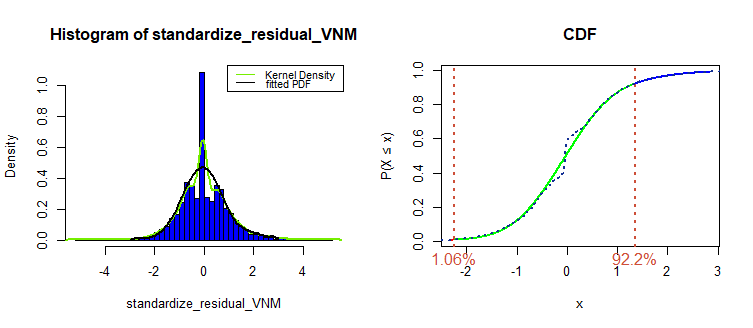
**Step 2:**

From the figure of standardized residuals we can see that all of distributions have the fat tails, therefore we use the Generalized Pareto Distribution to estimate the tails and Normal Distribution for the interior of each standardized residuals. In general, we fit the distribution called GNG (Generalized Pareto-Normal- Generalized Pareto) to estimate the standardized residuals.





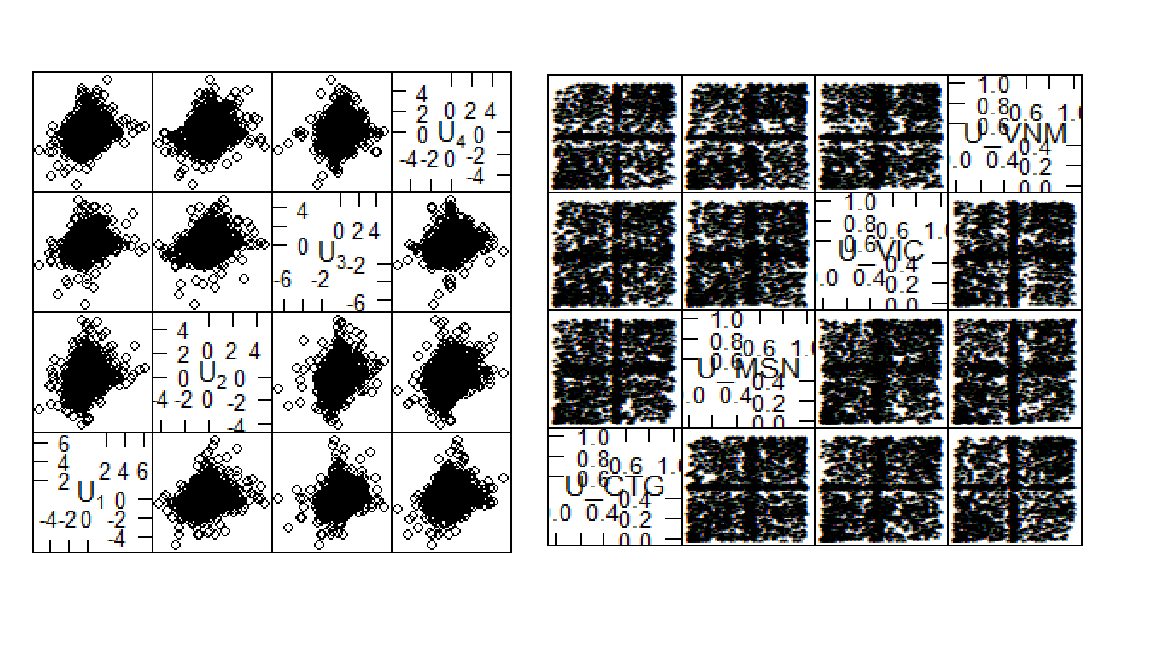




(Figure 8: the fitted density function and cumulative function of each standardized resudials)

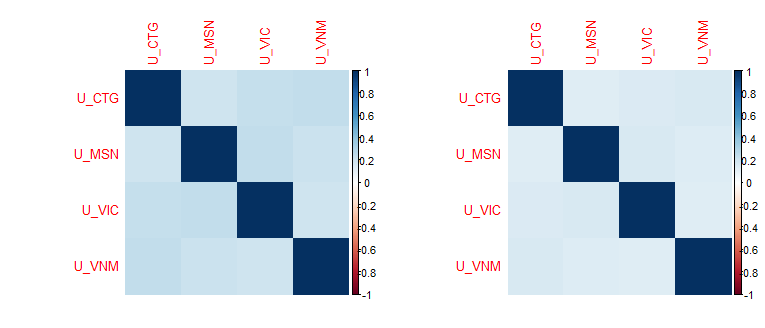
**Step 3:**

By using the marginal distributions (Fi(x)) generated from step 2, we transform the standardized residuals to the Uniform data.



(Figure 9: the standardized residuals of companies-left are transformed to to U data-right)

The dependence measure (spearman rho, Kendal tau) of portfolio is as following:



(Figure 10: The dependence measures of 4 companies using Spearman Rho- left and Kendal tau- right)

**Step 4:**

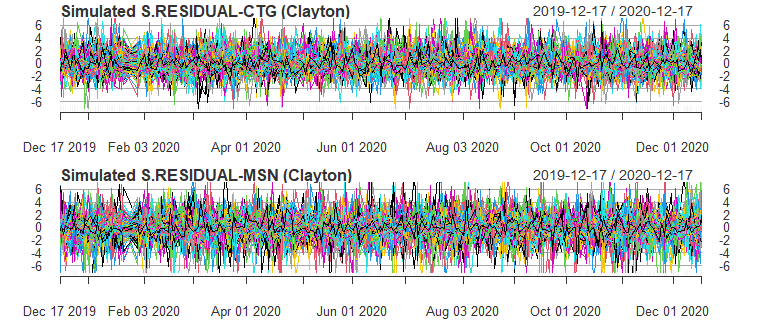
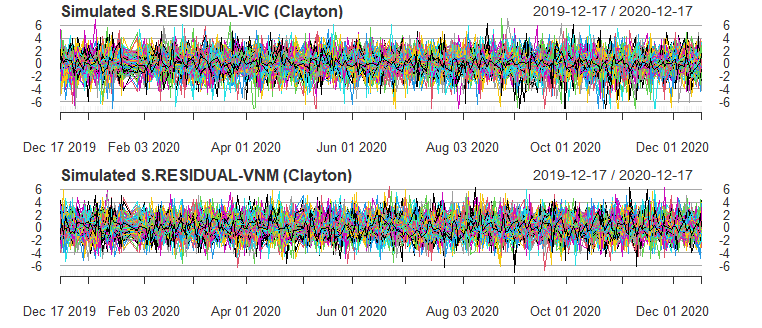
From the Uniform data, we estimate the parameters for each Copula (Gaussian, Student-t, Clayton, Gumbel, Frank) by Maximum Likelihood Estimation method (MLE). Parameters of each Copula is demonstrated in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Gaussian Copula** | **T-student Copula** | **Clayton Copula** | **Gumbel Copula** | **Frank Copula** |
| **Parameters** | 0.2462562 | 0.2448906 | 0.3553711 | 1.177686 | 1.382243 |
| 26.0066165(df) |
| **Error** | 0.012 | 0.013 | 0.026 | 0.013 | 0.005 |

(Table 4: The estimated Copula parameters for each type of Copula: Gaussian, T-student, Clayton, Gumbel, Frank)

**Step 5:**

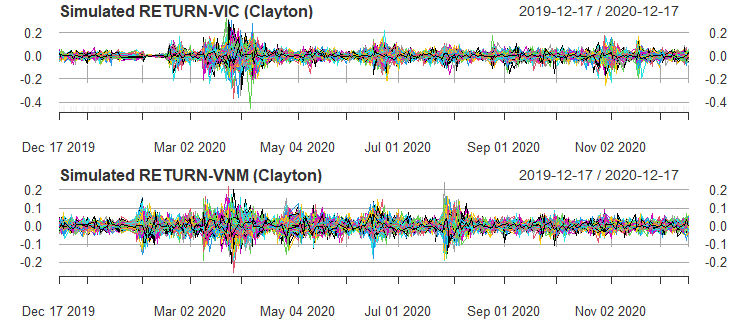
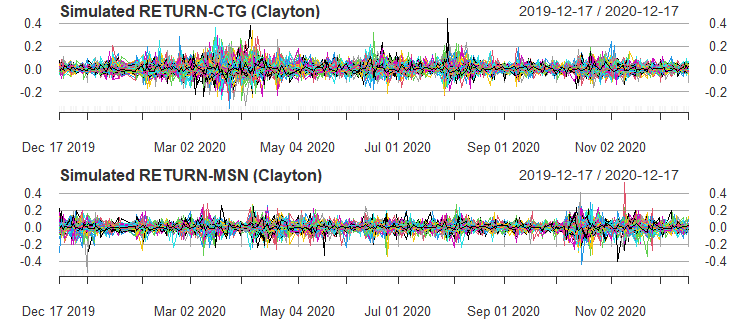
From the Copula Parameters estimated from step 4, we simulate 10,000 scenarios of each stock for each type of Copula. After that, using the Marginal distribution generated from step 2, we transform the Uniform data to simulated standardized residuals.

(Figure 11: 10,000 Simulated standardized residuals of Clayton Copula)

**Step 6:**

From mean and sigma collected from ARMA-GJR\_GARCH model, we can convert simulated standardized residuals from step 5 to simulated log-return for each company. We show here the simulated returns using Clayton Copula on validation data as an example:



(Figure 12: 10,000 Simulated return of Clayton Copula after re-introducing the ARMA-GJR\_GARCH model)

**Step 7:**

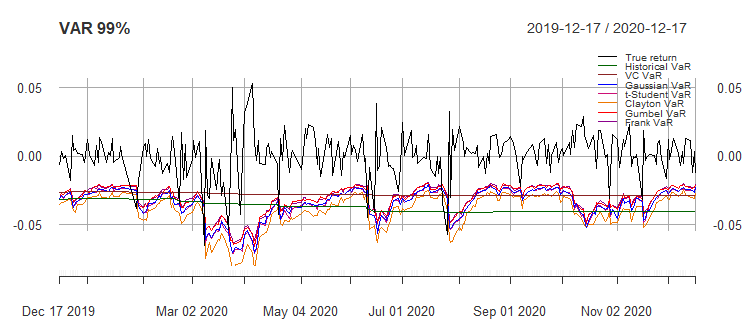
In this step, backtesting procedures are conducted by comparing daily log-return with daily VaR estimates by each type of Copula and 2 more traditional methods using Historical method and Variance and Covariance method. To more detail, we use the Kupiec’s Proportion of Failures Test, the Christof- fersen’s Independence Test and the Christoffersen’s Interval forecast for the correct number of exceedances .We use the data from 17th December 2019 to 17th December 2020 of each company as validation data.

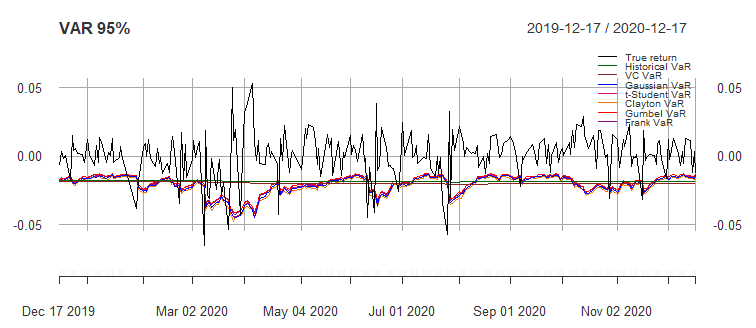
To evaluate the effectiveness of our method, we use the rolling prediction technique in validation data to predict the mean, and sigma of financial returns. The estimation sample is rolled over almost the entire data period, keeping the estimation constant, starting at the beginning of the data set. At the end of every period, a new period is added to the forecast, after 10 days, we re-fit the ARMA GARCH model to update the new parameters. This approach enables to project the future performance based on the most recent numbers and time frames. Then, as mentioned in the previous steps, from the predicted mean and sigma, we can convert it to simulated log returns.

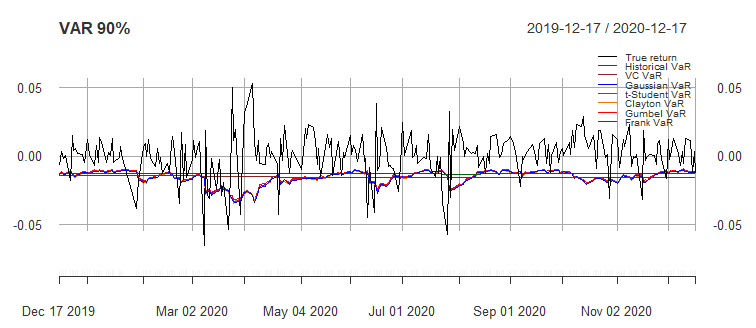
From 10,000 simulated log-return of portfolio, we can find VaR 99%,95% and 90% over testing period by choosing (1-anpha) quantile

By choosing the same weight for all assets, we calculate the true return of our portfolio

The comparison between true return of portfolio and VaR calculation can be shown as following:







(Figure 13: The estimated VaR 99%, 95% and VaR 90% using each type of Copula and the true return of portfolio over testing period)

**RESULT EVALUATION AND CONCLUSION:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **VAR 99%** | **VAR 95%** | **VAR 90%** | **Test result** |
| **Historical VaR** |  |  |  |  |
| **# of exeedance** | 9 | 22 | 30 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **RRA** |
| 0.001506472 | 0.01416477 | 0.3371425 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **RRA** |
| 0.004659949 | 0.005331087 | 0.2490315 |
| **CC** | 0.6002635 | 0.3962039 | 0.7025012 | **AAA** |
| **Variance Covariance VaR** |  |  |  |  |
| **# of exeedance** | 16 | 20 | 25 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **RRA** |
| 1.012258e-08 | 0.04974317 | 0.9497821 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **RRA** |
| 4.820793e-08 | 0.02861286 | 0.2735462 |
| **CC** | 0.2713078 | 0.3249206 | 0.4269727 | **AAA** |
| **Gaussian Copula** |  |  |  |  |
| **CC** | 7 | 15 | 30 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **RAA** |
| 0.02027658 | 0.5096227 | 0.3371425 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **RAA** |
| 0.02660305 | 0.7989569 | 0.6065338 |
| **CC** | 0.6678723 | 0.06591377 | 0.2128818 | **AAA** |
| **Student’s Copula** |  |  |  |  |
| **# of exeedance** | 6 | 16 | 30 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **AAA** |
| 0.0624619 | 0.3524215 | 0.3371425 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **AAA** |
| 0.05193667 | 0.6488839 | 0.6065338 |
| **CC** | 0.7885823 | 0.08499393 | 0.2128818 | **AAA** |
| **Clayton Copula** |  |  |  |  |
| **# of exeedance** | 6 | 14 | 30 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **AAA** |
| 0.0624619 | 0.7016122 | 0.3371425 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **AAA** |
| 0.05193667 | 0.8991027 | 0.6065338 |
| **CC** | 0.7885823 | 0.1060081 | 0.2128818 | **AAA** |
| **Gumbel Copula** |  |  |  |  |
| **# of exeedance** | 10 | 18 | 32 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **RAA** |
| 0.0003517808 | 0.1456047 | 0.1755705 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **RAA** |
| 0.001174697 | 0.3334387 | 0.3421259 |
| **CC** | 0.9949263 | 0.2154012 | 0.2068888 | **AAA** |
| **Frank Copula** |  |  |  |  |
| **# of exeedance** | 10 | 17 | 30 |  |
| **UC** | 3.841459 | 3.841459 | 3.841459 | **RAA** |
| 0.0003517808 | 0.2320721 | 0.3371425 |
| **IND** | 5.991465 | 5.991465 | 5.991465 | **AAA** |
| 0.001174697 | 0.4841837 | 0.6065338 |
| **CC** | 0.9949263 | 0.1821623 | 0.2128818 | **AAA** |

(Table 5: The test statistic and p-value of Historical, Variance Covariance, Gaussian, T-student, Clayton, Gumbel, Frank Copula using UC, IND, CC test)

UC, IND and CC represents the Kupiec’s Proportion of Failure test, the Christoffersen’s Independence test and the Christoffersen’s Interval Forecast test respectively.

The notation (A R A) in Table 5 means that the first test accept the null hypothesis that the model used for VaR estimation performs well on average, the second test rejects the null hypothesis and the last accept the null hypothesis. Note that, from Table 5 most of the number of exceptions are close to the expected exceptions (2 for the 99% confidence level, 12 for the 95% confidence level and 25 for the 90% confidence level). If the obtained number of exceptions are much larger than the expected ones, the models are said to have a poor performance in predicting the VaR. Otherwise (when they are much less) we say that the models fails to capture the information of historical observations (GARCH-EVT-Clayton Copula in our case). The results in the last Column of Table 5.are obtained by considering a significance level of 5%, and check if a test statistics’ p value is less than the significance level to reject the null hypothesis, that the model is not accurate on average. From the results we conclude that in general the GARCH- EVT-Copula approaches outperforms the commonly used Variance Covariance method and the Historical Simulation method with the exception of the GARCH-EVT-Clayton Copula case which passes only the Christoffersen’s Independence test.

In conclusion, our GARCH-EVT- COPULA method outperform the tradional way of calculating VaR (Historical VaR and Variance Covariance VaR). Especially, the T-Student and Clayton Copula pass all the test mentioned above, while VaR calculation using Gaussian VaR, Gumbel VaR are rejected only in VaR 99% (Kupiec and Christofersen’s Independence) and Frank Copula fails only in Kupiec test for VaR 99%. Notably, the VaR calculation using GARCH-EVT- COPULA method covers all the sudden changes occurred throughout the year 2020- year of Covid 19.