# Research and Analysis of Portfolio Based on Half-parameter Copula-VaR-CVaR Methods

Yifeng Qin

School of Mathematics, South China University of Technology, Guangzhou 510000, China.

## Abstract

**In this study, we discuss the portfolio based on Half-parameter Copula-VaR-CVaR Methods. In order to fit the cumulative distribution function better, we use Half-parameter Method rather than Parameter Method or Non-parameter Method. In this paper, we also consider five kinds of copula functions, including Gaussian Copula, Student T Copula, Clayton Copula, Frank Copula and Gumbel Copula, where the first two copulas are of Elliptical Type, while the last three copulas are of Archimedean Type. To discover which copula is better, we compare each copula mentioned above with Empirical Copula. Finally, we set up a portfolio model by minimizing VaR or CVaR. Some experiments are conducted in detail by using the data of Hushen 300 Index (IF300) and Zhongzheng 500 Index (IC500) to illustrate our developed models.**

## Keywords

**Half-parameter; Portfolio, Copula; Empirical Copula; VaR; CVaR.**

## 1. Introduction

All Modern portfolio theory is originated from the research work of Markowitz's mean–variance model [1]. Later, Sharpe [2] and many other researchers developed different mathematical methods in portfolio theory, most of which were based on probability and optimization theory.

In 21st century, Berger Theo and Missong Martin [3], Stulajter F. [4], T. Nagler, C. Bumann and C. Czado [5] and many other researchers introduced copula method into portfolio theory and illustrated some meaningful results.

At the same time, Lux Thibaut, and Rueschendorf Ludger [6], Madhusudan Karmakar and Samit Paul [7] use VaR(Value at Risk) and CVaR(Conditional Value at Risk) to build the portfolio model instead of varience. Manying Bai, and Lujie Sun [8] combined copula method and CvaR and applied them into portfolio models. Xiao-Li Gong, Xi-Hua Liu and Xiong Xiong [9] and Zong-Run Wang, Xiao-Hong Chen, Yan-Bo Jin and Yan-Ju Zhou [10] even introduced GARCH model into Copula-VaR-CVaR based portfolio models, and hence, many magnifique results have been made.

In summary, the primary focuses of this study are to research the portfolio theory based on Half-parameter Copula-VaR-CVaR Methods. The remainder of this paper is organized as follows. Section 2 introduces basic definitions and preliminary results related to the following models. Section 3 is the model construction. In Section 4, this study will give a numerical example, and the comparison analysis among different situation. Finally, Section 5 summarizes the conclusions and delivers future study scopes.

The major contributions and highlights of this study are: (a) Use the Half-parameter method to fit the cumulative distribution function; (b) Introduce different kinds of copula and compare each kind of copula with empirical copula; (c) Apply the VaR and CVaR models in portfolio theory, and even demand a constraint on the return, and compare their results.

## 2. Preliminaries

### 2.1 The definition

Two variables copula function is a function $C(u,v)$: $[0,1]\times[0,1]\rightarrow[0,1]$, and satisfies all the properties of a bi-various cumulative distribution function and has uniform marginal.

### 2.2 Basic theorem

#### 2.2.1 Sklar's theorem

Let H(x,y) be a joint distribution function of two variables with marginal distribution functions F(x) and G(y). Then, there exists a copula function C(u,v), such that H(x,y)=C(F(x),G(y)).

### 2.3 Some kinds of copulas

The above definition and theorem can also be extended to higher dimensional case.

Let ρ be a correlation matrix and $u_1$, $u_2$, $u_3$ in [0,1], the 3 dimensional Gaussian copula density is given by:

$$c(u_1,u_2,u_3;\rho)=\frac{1}{|\rho|^{1/2}}exp\left\{-\frac{1}{2}\xi^t\left(\rho^{-1}-\mathrm{I}\right)\xi\right\}$$

where ρ and ρ−1 are a three-dimension matrix and its inverse respectively, |ρ| is the determinant of the correlation matrix, ξ is the vector of the inverse standard univariate Gaussian cumulative distribution function, which is applied to each element $u_1$, $u_2$, $u_3$, and finally $\xi^t$ is the transposed vector of ξ. A three-dimension identity matrix I (i.e. with unit diagonal terms and zero elsewhere) is also employed.

Let ρ be a correlation matrix, ν a degree of freedom and $u_1$, $u_2$, $u_3$ in [0,1], the 3 dimensional Student T copula density is given by:

$$c(u_1,u_2,u_3;\rho,\nu)=\frac{1}{|\rho|^{1/2}}\frac{\Gamma\left(\frac{\nu+3}{2}\right)\left\{\Gamma\left(\frac{\nu}{2}\right)\right\}^3\left(1+\frac{1}{\nu}\xi^t\rho^{-1}\xi\right)^{-\frac{\nu+3}{2}}}{\left\{\Gamma\left(\frac{\nu+1}{2}\right)\right\}^3\Gamma\left(\frac{\nu}{2}\right)\prod_{n=1}^{3}\left(1+\frac{\xi_n^2}{\nu}\right)^{-\frac{\nu+1}{2}}}$$

where ρ and ρ−1 are a three-dimension matrix and its inverse respectively, |ρ| is the determinant of the correlation matrix, Γ is the Gamma function, ξ is the vector (ξ1, ξ2, ξ3) of the inverse univariate Student cumulative distribution function, which applies to each element $u_1$, $u_2$, $u_3$, and finally $\xi^t$ is the transposed vector of ξ.

Let θ >0 be a positive correlation parameter and $u_1$, $u_2$, $u_3$ in [0,1], Clayton copula is given by:

$$C(u_1,u_2,u_3;\theta)=\frac{1}{\left(u_1^{-\theta}+u_2^{-\theta}+u_3^{-\theta}-2\right)^{1/\theta}}$$

Let θ >0 be a positive correlation parameter and $u_1$, $u_2$, $u_3$ in [0,1], Frank copula is given by:

$$C(u_1,u_2,u_3;\theta)=-\frac{1}{\theta}ln\left\{1+\frac{\left(e^{-\theta u_1}-1\right)\left(e^{-\theta u_2}-1\right)\left(e^{-\theta u_3}-1\right)}{\left(e^{-\theta}-1\right)^2}\right\}$$

Let θ >1 be a positive correlation parameter and $u_1$, $u_2$, $u_3$ in [0,1], Gumbel copula is given by:

$$C(u_1,u_2,u_3;\theta)=\exp\left(-\left\{(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}+(-\ln u_3)^{\theta}\right\}^{1/\theta}\right)$$

## 3. Model construction

### 3.1 The construction of marginal distribution function

In the field of finance and insurance, the tail is often worth of attention. However, using only parameter estimates may bring the risk of model misplacement. Purely using non-parametric estimates, considering the extreme values of the tail may result in over-fitting. A compromise method is to segment the empirical data into three part, each part is estimated separately, and then combine them together. More precisely, the upper tail part and the lower tail part are estimated by parameter method, while the middle part is estimated by nom-parametric method. Then one can put these three parts together to construct the marginal distribution function. This is called half-parameter method.

In this paper, we set an upper tail threshold u, and a lower tail threshold l. Thus, the marginal distribution function has the form:

$$F(x)=\begin{cases} F1(x) & x<l \\ F2(x) & l\le x\le u \\ F3(x) & x>u \end{cases}$$

For the upper tail distribution, we use generalized pareto distribution, which is a thick-tailed distribution and is suitable to describe financial time series. It has the form:

$$F3(x)=1-(1+k\frac{x-u}{\sigma})^{-\frac{1}{k}}$$

where k, σ are estimated by maximum likelihood estimate. For the lower tail, after a symmetric transform, we can also use generalized pareto distribution.

To ensure the nondecreasing of distribution function, we use linear substitution when necessary.

For the middle part, we will apply the kernel density estimate. We will use Gaussian kernel function since it has good property. The Gaussian kernel function is defined as:

$$K(x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$$

Then, the density of the middle part has the form:

$$\hat{f}_h(x)=\sum_{i=1}^{n}\frac{1}{nh}K(\frac{x_i-x}{h})$$

where n is the length of financial time series, $x_i$ denotes the element of financial time series $(x_1,x_2,\dots,x_n)$, and h denotes the bandwidth.

To obtain the optimized h, we just need to minimize mean integrated squared error (MISE), which is defined as followed:

$$\mathrm{MISE}(h)=E\int(\hat{f}_h(x)-f(x))^2\,dx.$$

where f(x) denotes the real density function.

Finally, the estimate distribution function of the middle part is

$$F2(x)=\int_u^x \hat{f}_h(t)dt.$$

### 3.2 The construction of joint distribution function by introducing copula function

Up till now, we have obtained the marginal distribution function of each asset. Next, we introduce five kinds of copula function to obtain the joint distribution function. They are Gaussian Copula, Student T Copula, Clayton Copula, Frank Copula and Gumbel Copula. To discover which copula function fits the reality best, we define Empirical Copula as follows:

$$C(u,v)=\frac{1}{n}\sum_{i=1}^{n}I_{\{F(x_i)\le u\}}*I_{\{G(y_i)\le v\}}$$

where $F(x_i)$ and $G(y_i)$ denote the marginal distribution function of each asset, and $I_{\{A\}}$ is the characteristic function of set A.

By calculating the difference between each of the five copulas with empirical copula, and comparing the results, we can find the one which has the smallest difference, and hence the best one.

### 3.3 The construction of portfolio model based on Half-parameter Copula-VaR-CVaR Methods

Assume that the historical return of a financial asset is equal to the expected rate of return. Set the weights for investing in these two indices as $w_1$ and $w_2$. Then, define the Loss of Portfolio as below:

$$L=\max\{0, -(w_1 * r_1 + w_2 * r_2)\}$$

where $r_1$, $r_2$ denote the logarithmic return of two asset.

For a given confidence level $\alpha$, define Value at Risk:

$$VaR=F_L^{-1}(\alpha)$$

where $F_L^{-1}$ is the inverse function of distribution function of loss.

Define Conditional Value at Risk:

$$CVaR=E(L|L>=Var)$$

where E(X) denote the mathematical expectation of random variable X.

Finally, we construct the portfolio model based on Half-parameter Copula-VaR-CVaR Methods:

$$\min VaR \text{ or } CVaR$$

$$s.t. \begin{cases} w_1 + w_2 = 1 \\ w_1, w_2 > 0 \\ F(r_1, r_2) = C(F_1(r_1), F_2(r_2)) \end{cases}$$

where $r_1$, $r_2$ denote the logarithmic return of two asset, $F_1(r_1), F_2(r_2)$ denote the marginal distribution function of each asset, C(u,v) denote the copula function, and $F(r_1, r_2)$ denote the joint distribution function.

Moreover, we can construct another portfolio model which may be better by introducing a constraint on the expected return:

$$\min VaR \text{ or } CVaR$$

$$s.t. \begin{cases} w_1 + w_2 = 1 \\ w_1, w_2 > 0 \\ E \max\{0, (w_1 * r_1 + w_2 * r_2)\} \geq v \\ F(r_1, r_2) = C(F_1(r_1), F_2(r_2)) \end{cases}$$

Here, the additional constraint means that while we want to minimize the risk, we also want to maximize the return. So we add a constraint such that the return can not be too small, it has to be larger than some value. This new model provides a balance between the risk and the return.

## 4. Numerical examples

### 4.1 Data processing

In this section, we make some empirical analysis of the models developed in the last section. The data used in this section comes from the webpage https://money.163.com/. We download the daily prices of Hushen 300 Index and Zhongzheng 500 Index from 2015.01.06 to 2019.05.17. The reason why we choose these two asset is that roughly speaking, they represent the stocks of grand companies and small companies, respectively, which are of great representativeness. Then, we define the daily logarithmic return:

$$r_t = \ln(P_t) - \ln(P_{t-1})$$

where $P_t$ is the price on day t.

### 4.2 Figures and comparison

By the software R, we first give the histogram of daily logarithmic return of Hushen 300 Index and Zhongzheng 500 Index respectively, see Fig.1 and Fig.2.

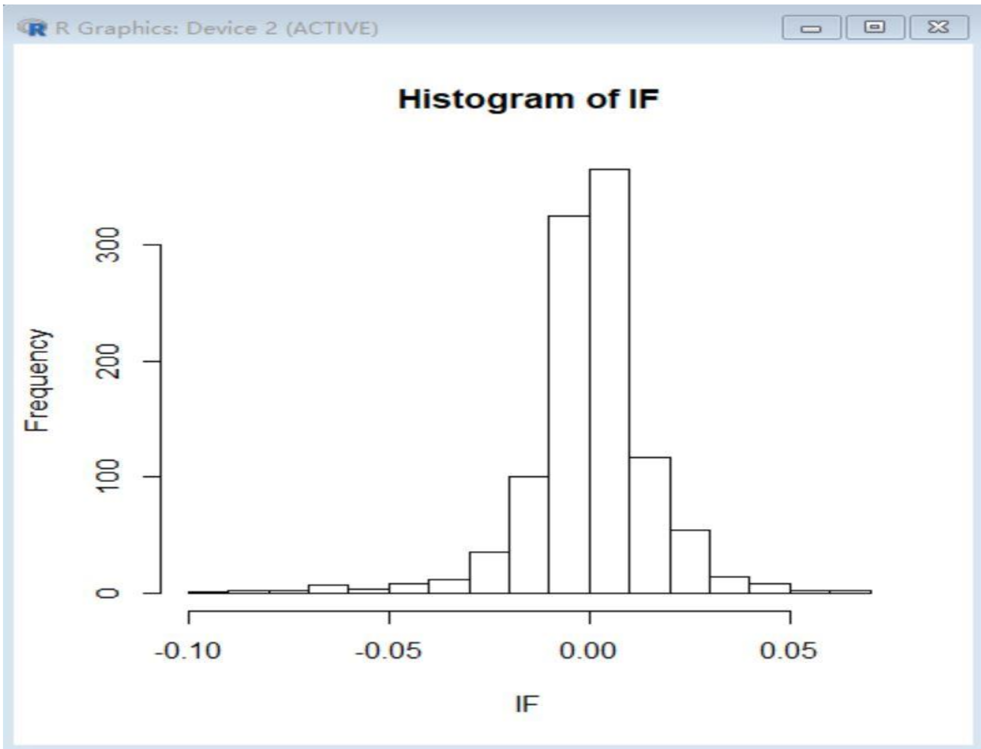


Fig. 1 Histogram of IF

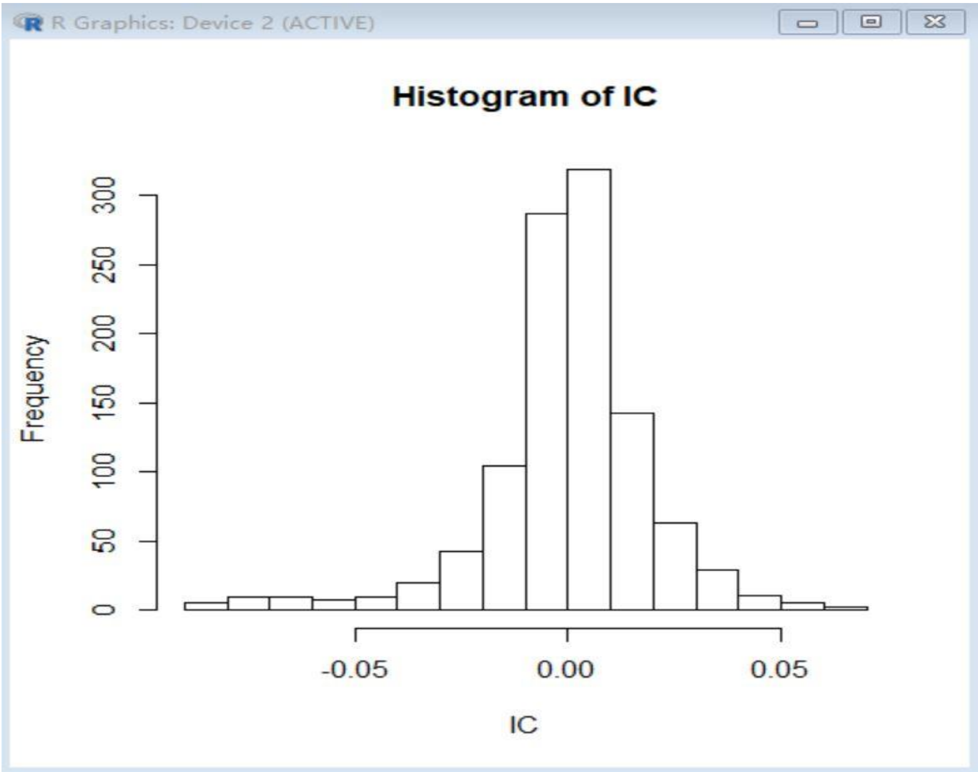


Fig. 2 Histogram of IC

From the figures above, we can see that the data are between -0.1 and 0.1, and most of the data are between -0.02 and 0.02. So we can view the data as tail extreme values as they are outside the interval [-0.02, 0.02]. Thus, applying the Half-parameter Estimation, with upper tail threshold u=0.02, and lower tail threshold l=-0.02, we get the marginal distribution function of daily logarithmic return of Hushen 300 Index and Zhongzheng 500 Index respectively, see Fig.3 and Fig.4.

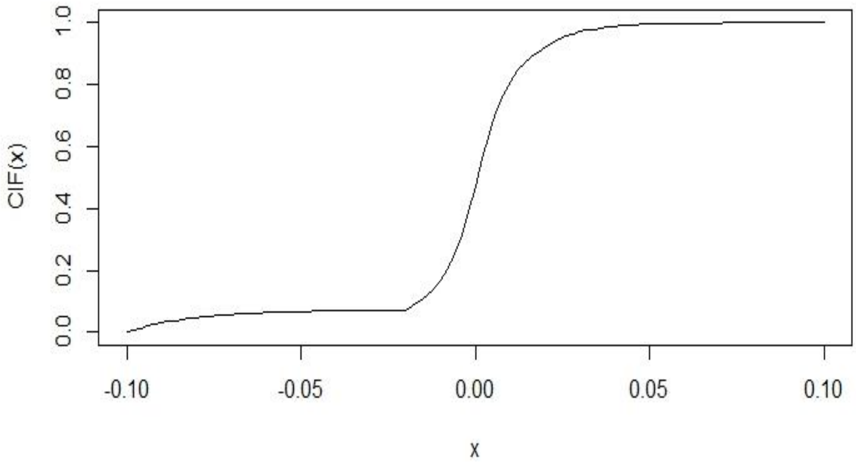


Fig. 3 Cumulated distribution function of IF

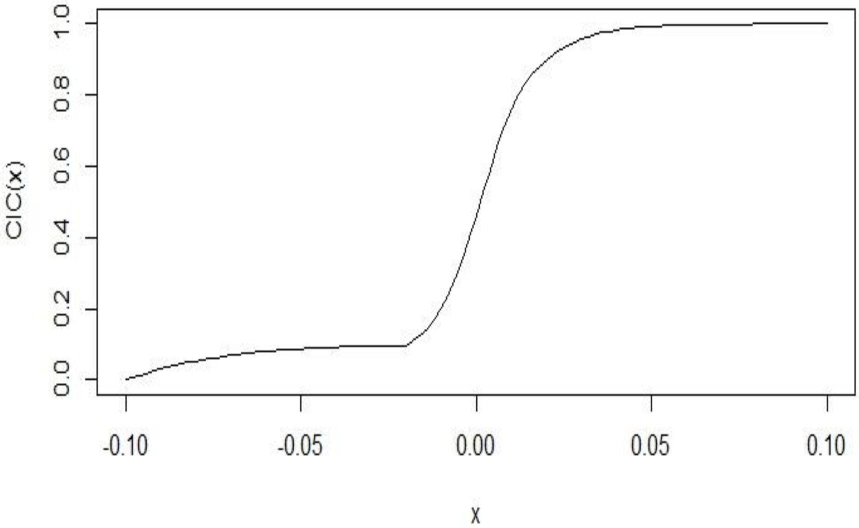


Fig. 4 Cumulated distribution function of IC

And also the density function of daily logarithmic return of Hushen 300 Index and Zhongzheng 500 Index respectively, see Fig.5 and Fig.6.

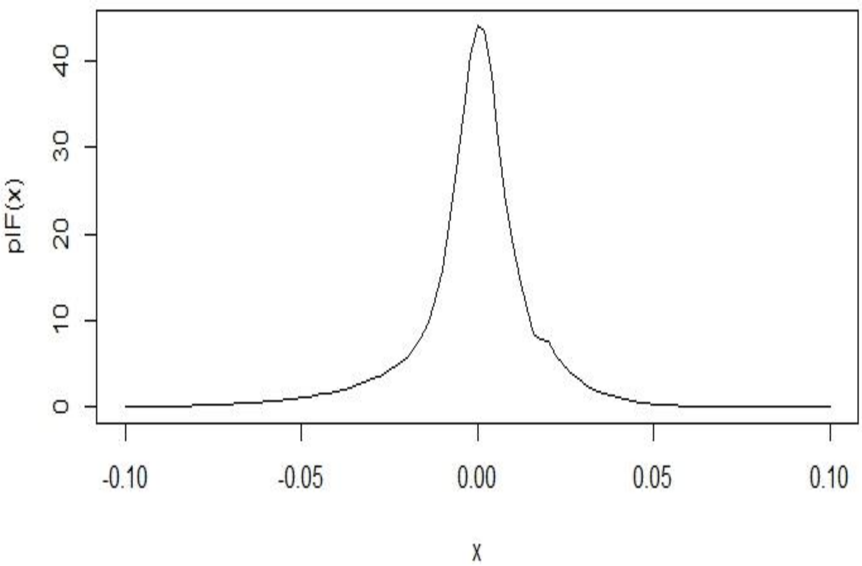


Fig. 5 Density of IF

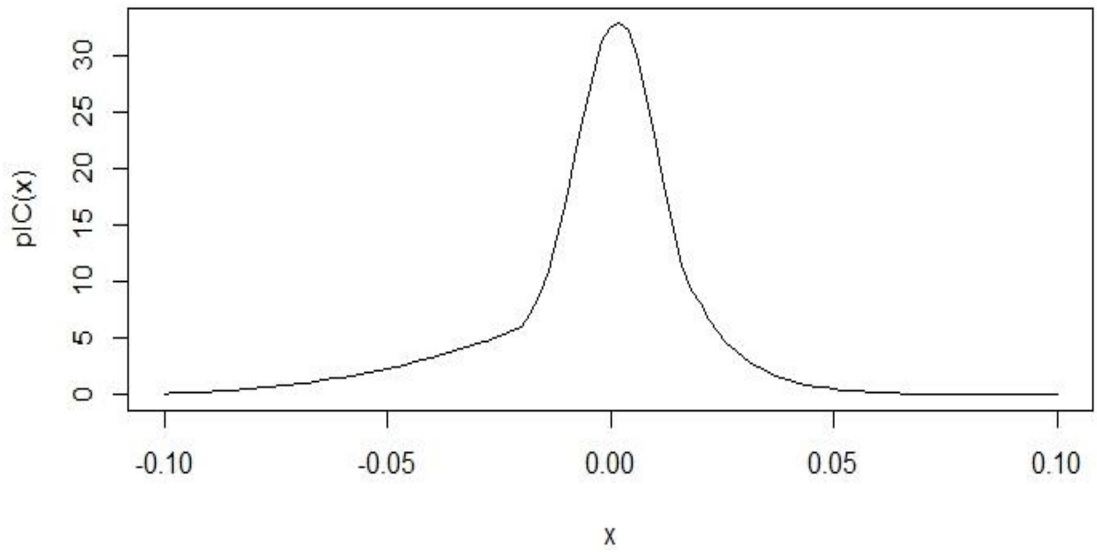


Fig. 6 Density of IC

Next, we apply the empirical copula function to obtain the joint distribution function (Let U,V be the marginal distribution function of the daily logarithmic return of Hushen 300 Index and Zhongzheng 500 Index respectively), see Fig.7. This is obtained by the software Matlab.

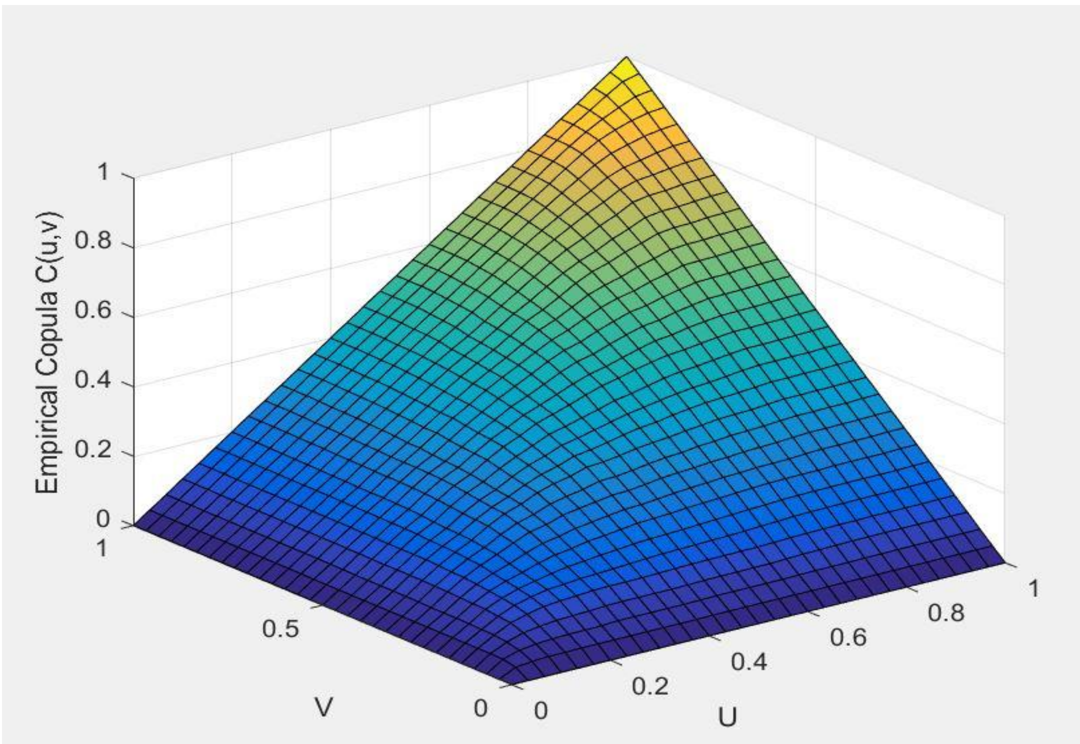


Fig. 7 Empirical Copula

By comparing the mean square difference between five kinds of copula (Gaussian Copula, Student T Copula, Clayton Copula, Frank Copula and Gumbel Copula) with empirical copula, we find that the mean square difference between Gumbel copula and empirical copula is the smallest among them, and thus, Gumbel copula is the best one to combine Hushen 300 Index and Zhongzheng 500 Index together. (Here are the mean square difference of five kinds of copula respectively: dGau2 = 0.1456, dt2 = 0.1412, dClay2 = 0.7160, dFr2 = 0.2595, dGum2 = 0.0914)

Now, we present the figure of Gumbel copula, see Fig.8.

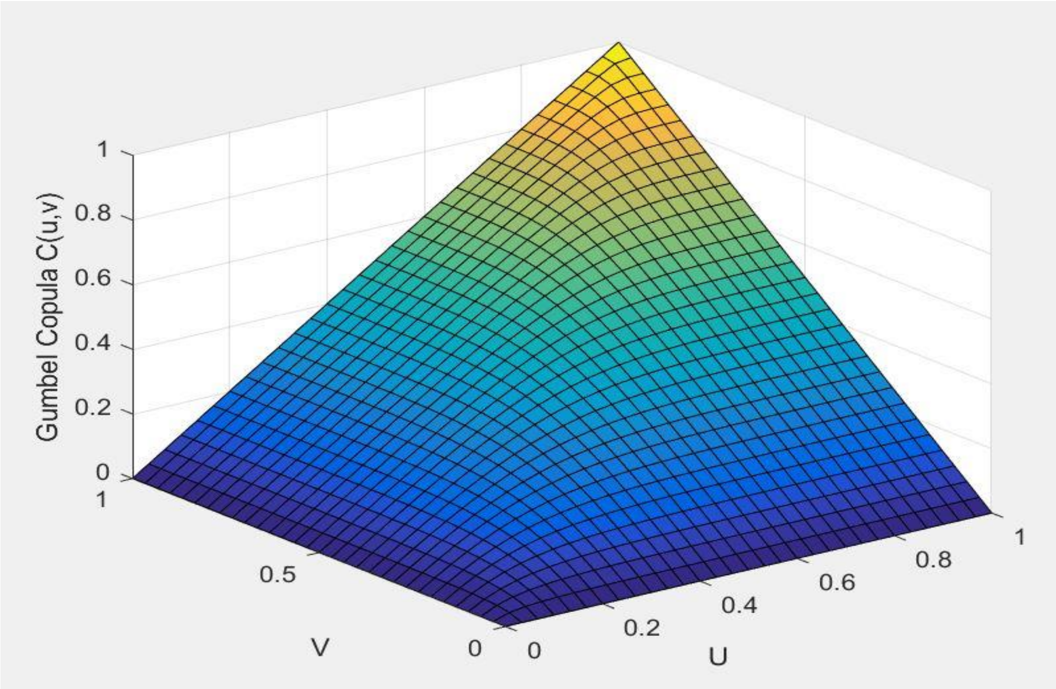


Fig. 8 Gumbel Copula

At last, we obtain the value of VaR and CVaR under the confidence level of 0.99, 0.95, 0.90.

In order to calculate VaR and CVaR, we use Monte Carlo Method to simulate the portfolio, by method of rejection, and then calculate the VaR and CVaR based on the simulated sequence. More precisely, given the density function f(x,y) of a random vector (X,Y), knowing that X⊆[a,b], Y⊆ [c,d] and f⊆[0,K], we want to simulate the random vector (X,Y). Firstly, we simulate three uniformly distributed random variables U, V and W on the interval [a,b], [c,d] and [0,K] respectively. If W is smaller than f(U,V), then we have finished, and the simulated (X,Y) is (U,V). Otherwise, we have to repeat, and to simulate the uniformly distributed random variables U, V and W again, until W is smaller than f(U,V). This is called the method of rejection, and it's very useful in our case.

If we set $w_1 = 0.99, w_2$=0.01, we get the following Table 1.

Table 1. VaR and CVaR under different confidence levels

| confidence level | VaR | CVaR |
|---|---|---|
| 0.99 | 0.0509 | 0.0610 |
| 0.95 | 0.0278 | 0.0409 |
| 0.90 | 0.0194 | 0.0322 |

From the table above, we sum up some facts. There is a 99% of confidence to say that the maximum loss of portfolio is no more than 0.0509, and there is a 90% of confidence to say that the maximum loss of portfolio is no more than 0.0194. As the confidence level goes down, both VaR and CVaR go down, which indicates that if we wish to speak more precisely, the number of maximum loss of portfolio has to be larger. On the other side, comparing Var and CVaR, we find that at each confidence level, CVar is no less than VaR, which indicates that CVaR describes the risk of portfolio more completely than VaR does, and hence it is a better index to describe the risk of portfolio.

Now we find the optimal value of VaR and CVaR as the weight of the portfolio varies. Here are the results, see Table 2.

Table 2. Optimal portfolio by minimizing VaR or CVaR

| confidence level | optimal $(w_1, w_2)$ by min VaR and optimal VaR | optimal $(w_1, w_2)$ by min CVaR and optimal CVaR |
|---|---|---|
| 0.99 | (0.89, 0.11) 0.0507 | (0.99, 0.01) 0.0610 |
| 0.95 | (0.98, 0.02) 0.0278 | (0.99, 0.01) 0.0409 |
| 0.90 | (0.96, 0.04) 0.0193 | (0.99, 0.01) 0.0322 |

From the table above, we know that if we hope to minimize the maximum loss of portfolio at the confidence level 0.99, we should divide a 89% of our money to buy Hushen 300 Index, while the rest to buy Zhongzheng 500 Index. In general, we can see that the risk of Zhongzheng 500 Index is larger than Hushen 300 Index, this is why we allocate more amount to Hushen 300 Index.

However, if we not only wish to minimize the risk but also hope to maximize the return, then we can add an additional constraint as in the model discussed at the end of section 3 with v=0.012. Therefore, we obtain the following results, see Table 3.

Table 3. Optimal portfolio by minimizing VaR or CVaR with an extra constraint

| confidence level | optimal $(w_1, w_2)$ by min VaR with an extra constraint and optimal VaR | optimal $(w_1, w_2)$ by min CVaR with an extra constraint and optimal CVaR |
|---|---|---|
| 0.99 | (0.79, 0.21) 0.0511 | (0.79, 0.21) 0.0624 |
| 0.95 | (0.77, 0.23) 0.0278 | (0.79, 0.21) 0.0416 |
| 0.90 | (0.79, 0.21) 0.0194 | (0.79, 0.21) 0.0328 |

From this table, we can conclude that after adding the extra constraint, both the VaR and CVaR increase. This is because if we want to gain more return, we will have to suffer more risk. And the optimal weight is (0.79, 0.21) at confidence level 0.99. Comparing with the former table, where the optimal weight is (0.89, 0.11), this implies that Zhongzheng 500 Index may have more return than Hushen 300 Index, while at the same time, it may have more risk.

## 5. Conclusion and future study scopes

In this paper, we discuss the portfolio based on Half-parameter Copula-VaR-CVaR Methods. We use Half-parameter Method rather than Parameter Method or Non-parameter Method to fit the cumulative distribution function better. We also consider five kinds of copula functions including Gaussian Copula, Student T Copula, Clayton Copula, Frank Copula and Gumbel Copula. By comparing each copula mentioned above with Empirical Copula, we find that Gumbel copula is the best among the five above to describe the joint distribution function. Finally, we set up a portfolio model by minimizing VaR or CVaR and find that CVaR can describe the risk of portfolio more completely than VaR.

Looking forward to the follow-up research work, some further research is worthy. (a): There is only one joint distribution function in this paper, yet one can develop a joint distribution function for each year, and this will fit the real data better. (b): In this paper, we use VaR and CVaR to develop portfolio, yet one can introduce other index such as all kinds of entropy. Some interesting results may come out!

## Acknowledgements

I would like to express my special thanks and gratitude to Professor Bing Li and Professor Xue Deng for their patient guidances and encouragement in my research. I would also like to thank my parents and my friends for their support throughout my study.

## References

[1] Markowitz H., Portfolio selection. Journal of Finance, 7 (1952): 77-91.

[2] Sharpe W.F., Portfolio theory and capital markets. McGraw-Hill, New York.

[3] Berger Theo, Missong Martin, Copulas and portfolio strategies: an applied risk management perspective. Journal of risk, 17, 2 (2014): 51-91.

[4] Stulajter F., Introduction to copula functions and their application in portfolio and risk management. Rizeni a modelovani financnich rizik, (2008): 261-270.

[5] T. Nagler, C. Bumann, C. Czado, Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. Journal of Multivariate Analysis, 172 (2019) 180–192.

[6] Lux Thibaut, Rueschendorf Ludger, Value-at-Risk bounds with two-sided dependence information. Mathematical Finance, 29,3 (2019): 967-1000.

[7] Madhusudan Karmakar, Samit Paul, Intraday portfolio risk management using VaR and CVaR: International Journal of Forecasting 35 (2019) 699–709.

[8] Manying Bai, Lujie Sun, Application of copula and Copula-CVaR in the Multivariate portfolio optimization. Lecture notes in computer science, 4614 (2007): 231.

[9] Xiao-Li Gong, Xi-Hua Liu, Xiong Xiong, Measuring tail risk with GAS time varying copula, fat tailed GARCH model and hedging for crude oil futures. Pacific-Basin Finance Journal 55 (2019) 95–109.

[10] Zong-Run Wang, Xiao-Hong Chen, Yan-Bo Jin, Yan-Ju Zhou, Estimating risk of foreign exchange portfolio: Using VaR and CVaR based on GARCH-EVT-Copula model. Physica A 389 (2010) 4918-4928.

[11] Maziar Sahamkhadam, Andreas Stephan, Ralf Östermark, Portfolio optimization based on GARCH-EVT-Copula forecasting models. International Journal of Forecasting 34 (2018) 497–506.

[12] Qiang Ji, Bing-Yue Liu, Ying Fan, Risk dependence of CoVaR and structural change between oil prices and exchange rates: A time-varying copula model. Energy Economics 77 (2019) 80–92.

[13] Kun Yang, Yu Wei, Jianmin He, Shouwei Li, Dependence and risk spillovers between mainland China and London stock markets before and after the Stock Connect programs. Physica A 526 (2019) 120883.