Study on the Correlation between the Returns of the Shanghai and Shenzhen Stock Market Based on Copula Model

Siliang Guo

School of Economics & Management

Qilu Normal University

Ji'nan, China

E-mail:GSL200601@126.com

***Abstract*-Using the Copula function to study the correlation between the stock market of Shanghai and Shenzhen in China., selecting the sample data of the two stock market to construct the Copula-EGARCH (1,1) -t model to describe the correlation degree and correlation structure between the two stock markets.**

Keywords*: *Copula model; Returns of stock market; Correlation

I. DATA SELECTION

In order to study the correlation between Shanghai and Shenzhen stock markets, we selected the daily closing price of Shanghai Composite Index (000001) and Shenzhen Component Index (399001). The symbol $_{\{P\}}$ represented the daily closing price of the market index. The rate of return was defined as logarithmic rate of return: $R_t = \ln P_t - \ln P_{t-1}$, RSH represented the rate of return for Shanghai Composite Index, and RSZ represented the rate of return for Shenzhen Component Index. We chose the sample period from December 17, 1996 to December 31, 2013, a total of 3536 sets of valid data, which came from YAHOO finance.

We made a simple descriptive statistical analysis of the two samples data series of rate of return, and the results were shown in table I.

TABLEI. DESCRIPTIVE STATISTICAL ANALYSIS OF RATE OF RETURN

samples	min	max	aver	std_deviation	skewness	kurtosis
RSH	-0.093342	0.094008	0.000449	0.01651	-0.287232	7.848058
RSZ	-0.099352	0.095299	0.000439	0.018208	-0.195240	7.103302

Table 1showed that the two samples data series had the characteristics of the peak and the skewness, the average yield of the income tends to 0. We made a normal test for the distribution of the two samples data series, and the results were shown in table II.

TABLE II. DISTRIBUTION NORMALITY TEST OF THE DISTRIBUTION OF THE TWO SAMPLES DATA SERIES

samples	Jarque-Bera statistics	P-value
RSH	2716.046	0.000000
RSZ	1936.102	0.000000

Table II showed that the two samples of return rate series rejected the assumption that the data follow normal distribution. So we should not use normal distribution mechanically but consider the actual characteristics of data when we use the Copula model.

The Two samples of return rate series were obvious volatility clustering. In order to test the effect of the aggregative effect (ARCH effect), we made a Lagrange Multiplier test (test LM), which was in the hypothesis that the sequence had no ARCH effect, and the test results were shown in Table III.

TABLE III. THE TEST OF THE AGGREGATION EFFECT OF THE VOLATILITY OF THE RETURN SERIES

samples	Lag=1	Lag=5	Lag=10
RSH	75.8759(0.00)	153.3261(0.00)	166.7470(0.00)
RSZ	95.5103(0.00)	200.0027(0.00)	225.2831(0.00)

Table III showed that all the test results rejected the original hypothesis, so the two samples series of rate of return had ARCH effect. So we can choose the ARCH class model when fitting the marginal distribution of the yield, we can use AR (1) -EGARCH (1,1) -t model to fit the marginal distribution of the yield.

II. MODEL ESTABLISHMENT

We used AR (1) -EGARCH () (1,1) -t model to fit the two stock market returns, and the distribution of the two yield sequences are as follows:

$$r_{1t} \mid I_{t-1} \sim i.i.d \qquad t_{v1} (\mu_1 + \phi_1 r_{1,t-1}, h_{1t}) \qquad (1)$$

$$r_{2t} \mid I_{t-1} \sim i.i.d \qquad t_{v2} (\mu_2 + \phi_2 r_{2,t-1}, h_{2t}) \qquad (2)$$

Among them，$v1$ and $v2$ represent degrees of freedom of the two distributions. Assuming that the joint distribution of $(r_{1t}, r_{2t}) \mid I_{t-1}$ is F , then according to the "Sklar theorem":

$$F(r_{1t}, r_{2t} \mid I_{t-1}) = C(T_{v1}(r_{1t} \mid I_{t-1}), T_{v2}(r_{2t} \mid I_{t-1}) \mid I_{t-1}) \qquad (3)$$

Among them，T_{v1} and T_{v2} are t distribution functions in formula 1and formula 2. According to the property of Copula function, when the random variable is monotonically increasing, the corresponding Copula

function is kept constant. So we can use the Copula function to study the correlation between the stock market return sequence and the correlation between the two standard residuals.

So the process of our model construction was that: First of all, using -EGARCH (1) AR (1,1) -t model to fit the sequence of returns; then we can use the formula $z_t = \dfrac{\varepsilon_t}{\sqrt{h_t}}$ to separate the non-observed standard residuals; then we can make the probability integral transform to the standard residuals, so that the transformed sequence was distributed evenly (0,1).Secondly, selecting the appropriate Copula function to describe the correlation structure of the sequence, and to estimate the parameters of the selected Copula function, and to find out the correlation between the rank correlation coefficient and the tail correlation coefficient. There were no more authoritative methods for the selection of the appropriate Copula function, the more commonly used method was to select a large number of Copula functions to fit the sample data, then according to a certain standard to select the optimal Copula function.

III. FITTING OF CORRELATION ANALYSIS MODEL OF STOCK MARKET

We used AR (1) -EGARCH () (two) (1,1) -t model to estimate the distribution of the rate of return in Shanghai and Shenzhen stock markets, and the maximum likelihood method was used to estimate the parameters. The results were shown in Table IV.

TABLE IV. PARAMETER ESTIMATION OF YIELD SEQUENCE

sam ples	μ	ϕ_1	α_0	α_1	γ	β_1	v	K-S val ue	P-v alue
RSH	0.00 3 (0.1 2)	0.02 2 (0.2 4)	-0.4 92 (0.0 0)	0.2 51 (0. 00)	-0.0 57 (0. 0)	0.9 63 (0. 00)	4.9 62 (0. 00)	0.0 23	0.46 2
RSZ	-0.0 00 (0.7 26)	0.02 2 (0.2 53)	-0.3 84 (0.0 0)	0.2 26 (0. 00)	-0.2 78 (0.0 3)	0.9 74 (0. 00)	5.4 41 (0. 00)	0.0 13	0.97 2

In Table IV, the value of the parentheses were estimated by AR (1) -EGARCH (1,1) -t model, and the value in parentheses were adjoint probability value. From table 4 we saw that the yield of two samples were not subjected to AR (1) process, that is, the average yield of 0, which was consistent with the results of the description of Table 3. The γ values of the two models were all negative, which showed that the yield sequence had a negative leverage effect, which was consistent with our expectations. Next, we draw the Q-Q diagram between the two standard error term sequences and the corresponding t distribution to make a visual display of the fitting effect of the model.

The Q-Q diagram showed that the AR (1) -EGARCH (1,1) -t model can well fit the distribution of the two sample series of rate of return. Next, chose the appropriate Copula function to describe the correlation structure

between the two sample series of rate of return. Firstly, we gave the scatter diagram of the two standard residuals, named U and V, which had been transformed by probability integral.
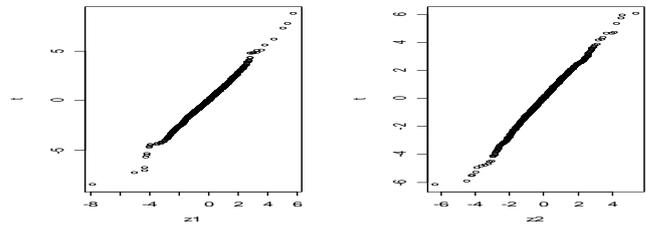

Figure 1. The Q-Q diagram between sequence of residuals and t distribution
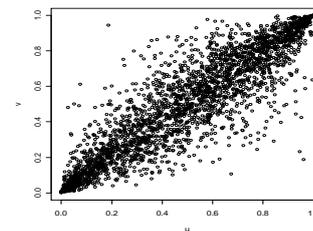

Figure 2. The scatter diagram of the two standard residuals

The scatter diagram showed that there were obvious positive correlation between the two samples of returns series, and there were also strong tail dependence in the upper and lower tail. So we choose Normal-Copula, t-Copula, Gumbel- Copula, Frank- Copula, Clayton-Copula and Galambos-Copula, which can describe different structures of the sequences , to describe the correlation structure of the sequences. The estimation results of Copula function parameters, made by the methods of two stage maximum likelihood estimation, nonparametric method proposed by Rivest and Genest, and pseudo maximum likelihood estimation, were given in table V.

TABLE V. PARAMETER ESTIMATION VALUE OF COPULA FUNCTION WITH DIFFERENT ESTIMATION METHODS

Copula	Two stage maximum likelihood estimatio（IFM）		Pseudo maximum likelihood estimatio（CML）		Genest and Rivest	
	pa ra me ter	Estimated value	par am ete r	Estimated value	pa ra me ter	Estim ated value
Normal	ρ	0.9100481	ρ	0.909193	—	—
Student-t	ρ v	0.9139808 4.0024497	ρ v	0.9134155 3.8363042	—	—
Gumbel	α	3.530776	α	3.498786	α	3.761 65
Frank	α	13.12987	α	13.05366	α	3.152 37
Clayton	α	3.699362	α	3.795428	α	5.523 30
Galambos	α	2.814481	α	2.780745	α	3.213 45

Table V showed that the difference of estimated value between the two stage maximum likelihood estimation method and the pseudo maximum likelihood estimation method was small. But the difference of estimated value between the nonparametric method of Genest and Rivest and the two methods before was obvious.

After estimating the parameters of the Copula function, we need to select the optimal Copula function for fitting the sample data. In this paper, we used the conditional distribution method and Kolmgorov-Smirnov test to select the optimal Copula function. We tested the estimation results of the three estimation methods. The conditional distribution Q-Q graphs of the estimated results of the three methods were given as follows.

From Figure3, we can conclude that the fitting effect of the methods of t-Copula, Gumbel- Copula, Frank- Copula and Galambos-Copula were good, but the fitting effect in the lower tail of Gumbel- Copula and Galambos- Copula were poor.
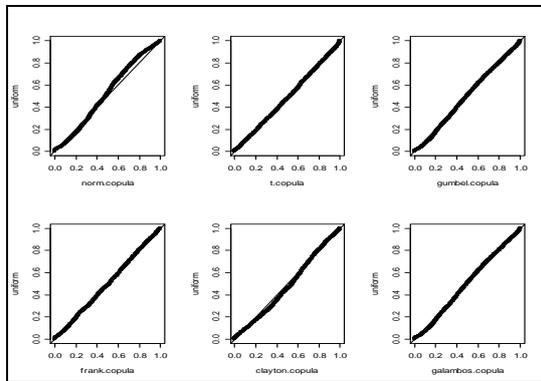

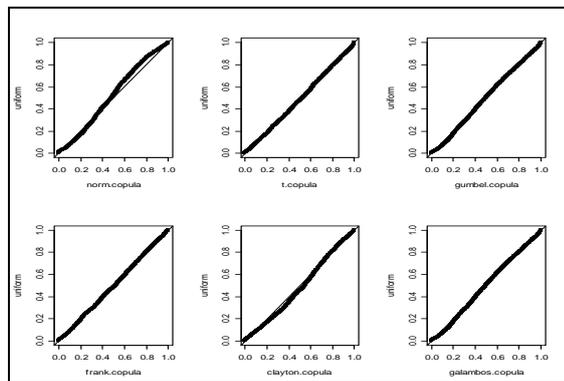Figure 3. Conditional distribution Q-Q graphs of the estimated results of IFM

Figure 4. Conditional distribution Q-Q graphs of the estimated results of CML

Figure 4 showed that the fitting effect oft-Copula, Gumbel- Copula, Frank- Copula and Galambos- Copula were good. The fitting effect of Frank-Copula and t-Copula in the whole distribution were good, but the fitting effect in the lower tail of Galambos- Copula and Gumbel- Copula were poor. The results of pseudo maximum likelihood estimation are very similar to the results of maximum likelihood estimation in the two stages, but the fitting effect of Galambos-Copula and

Gumbel- Copula were better than the two stages maximum likelihood estimation. The fitting effect of Galambos-Copula and Gumbel- Copula were good. The fitting effect of Frank- Copula was poor, which was different from the results obtained by the two methods.

By analyzing the conditional distribution Q-Q graphs of the three estimation methods, we found that different estimation methods have significant effect on the fitting effect of Copula function. In general, the performance of Galambos-Copula and Gumbel-Copula in the three estimation methods were relatively stable, Frank-Copula and t-Copula were better than the previous two methods, but the fitting effect of nonparametric estimation method of Genest and Rivest and Frank- Copula method were poor. In addition, it may not be used to determine which Copula function was the best fitting of the sample data. Therefore, we use Kolmgorov-Smirnov test to perform the goodness of fit test. The results are shown in Table VI..

TABLE VI. K-S TEST RESULTS FOR DIFFERENT PARAMETER ESTIMATION METHODS

Copula	Two stage maximum likelihood estimation （IFM）		Pseudo maximum likelihood estimation （CML）		Genest and Rivest	
	K-S value	P-value	K-S vaklue	P-value	K-S value	P-value
Normal	0.0666	0.0001	0.0673	0.000	—	—
Student-t	0.0322	0.1177	0.0315	0.134	—	—
Gumbel	0.038	0.0383	0.0395	0.028	0.0271	0.2692
Frank	0.023	0.4623	0.0227	0.483	0.0892	0.0007
Clayton	0.06	0.0001	0.06	0.000	0.0622	0.0005
Galambos	0.0406	0.0221	0.0417	0.017	0.0318	0.1255

We judged the goodness of fit of different Copula functions according to the P- value. For the two stage maximum likelihood estimation method, by comparing the P- value, we found that the fitting effect of Frank-Copula was the best in all the methods, the fitting effect of t-Copula was followed. For the pseudo maximum likelihood estimation, the results were consistent with the results of the two stage maximum likelihood estimation method, the fitting effect of Frank-Copula was the first, and the fitting effect of t-Copula was second. For the nonparametric estimation of Genest and Rivest, method, the fitting effect of Gumbel-Copula was the first, and the fitting effect of Galambos-Copula was second, which was the same as that of graphic method.

IV. RESULTS ANALYSIS AND CONCLUSIONS

After estimating the parameters value of Copula function and selecting the optimal Copula function, we can use the corresponding Copula function to study the correlation degree and correlation structure between the stock market in Shanghai and Shenzhen. We choose t-Copula, Gumbel-Copula, Frank-Copula, and

Galambos-Copula to describe the correlation between Shanghai and Shenzhen stock market. Kendall's tau and the corresponding tail correlation coefficient, which were calculated according to different estimation methods, were shown in TableVII.

TABLE VII. CORRELATION COEFFICIENTS OF SHANGHAI AND SHENZHEN STOCK MARKET

Copula	Two stage maximum likelihood estimation（IFM）			Pseudo maximum likelihood estimation			Genest and Rivest		
	τ	λ_U	λ_L	τ	λ_U	λ_L	τ	λ_U	λ_L
Student-t	0.734	0.655	0.655	0.733	0.660	0.660	—	—	—
Gumbel	0.717	0.783	0	0.714	0.781	0	0.734	0.798	0
Frank	0.752	0	0	0.750	0	0	0.868	0	0
Galambos	0.775	0.890	0.782	0.781	0.887	0.779	0.793	0.918	0.806

We can made conclusions that the rank correlation coefficients named "Kendall's tau" calculated through the three different estimation methods were roughly in the range of 0.7-0.8, which indicated that there was a strong positive correlation between the sequences of Shanghai Composite Index and Shenzhen Component Index. Besides, we believe that there was a clear tail dependence which was obvious non-symmetry between the yield sequences of Shanghai composite index and Shenzhen Component Index. When extreme events occur in the stock market, like the stock price boom or slump, the price Coordination motion of Shanghai Composite Index and Shenzhen Component Index was markedly increased, and compared with the collapse of the extreme events, skyrocketing extreme events made the two stock markets had a higher correlation. These conclusions showed that the change of the stock market index of Shanghai stock market and Shenzhen stock market had a high consistency, which was more obvious when the extreme events occur. So we should not study the single market in isolation, but we should pay attention to the correlation degree and correlation model between the Shanghai stock market and the Shenzhen stock market, and analyze the two markets comprehensively.

REFERENCES

[1] Chang Cheng. Copula-GARCH model based on extreme value theory and its application in financial risk [D]. Minzu University of China, 2013

[2] Xu Miaozhi. Copula theory and its application in the analysis of financial market dependence [D]. Electronic Science and Technology University, 2013

[3] Sklar, A Functions de repartition an dimensions et leurs marges [J] .Publication de IUniversit de Pairs, 1959, (8):229-231.

[4] Genst.C and MacKay R.L. Copules archimédiennes et familes de lois bidimensionnelles dont lesmarges sont donnés [J], The Canadian Journal of Statistics, 1986, 14:145-159.

[5] Joe H. Multivariate Models and Dependence Concepts [M], London: Champan & Hall, 1997.