A Novel Application of the Copula Function to Correlation Analysis of Hushen300 Stock Index Futures and HS300 Stock Index

Fang WU * 1,2, Yu WEI 1

1. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China
2. School of Economics and Management, Southwest University, Chongqing 400715, China

Abstract — Stock index futures is grounded in hedging, with correlation between stock index futures and stock index spot is important due to its significance in asset pricing and venture management. In this paper, we apply the static and time varying Copula functions, by which correlation of HS300 stock index future and HS300 index is analyzed in detail. Our research shows that, just before and after the official launching of HS300 stock index futures, difference between the average rate of return of HS300 index is smooth, but fluctuation of the former is much obvious as compared with the latter. Furthermore, to model the correlation of HS300 stock index future and HS300 stock index, the t-Copula function is the best static function, while the low-tail correlated Rotated Gumbel Copula function is the most suitable time varying function.

Keywords - HS300 Stock index future; HS300 stock index; time-varying Copula

I. INTRODUCTION

After four years’ data simulation and testing, Chinese stock index future came to its official launching in April 16, 2010. At first, Chinese stock index future market has great shock on the spot market, and then it turned to be mature as time goes by. For the correlation between stock index future market and stock index spot market, scholars have developed three viewpoints. Harris and Damodaran take the opinion that the stock index spot market can get much information when the stock index future market is launched [1-3]. The higher information transmission speed, the higher the efficiency and the higher investor’s reaction speed. As a result, the much fluctuation of stock index spots market. Bessembinder think that stock index future market has the capability of hedging [4], falls of stock index spot market will lead to sell-off among the investors, and cause a new round of falls in the stock index spot. But this kind of vicious circle can be reduced by the existence of stock index future market, and fluctuation in stock index spot market can be limited too. Others think that stock index future market has no influence on stock index spot market, in which the positive effects and the negative effects are coexisting and cancel each other out.

To find an effective tool to scale the correlation between stock index future market and stock index spot market, Sklar defined the Copula function in 1959. Since the Copula function is not only a marginal distribution function, but also a joint distribution function that takes value belong to \([0, 1]\), it is the bridge that connects the single market marginal distribution and the multidimensional market joint distribution. In 1999 Nelson proved that a Copula function has no influence on stock index spot market, in which the positive effects and the negative effects are coexisting and cancel each other out.

Copula function family are proposed. For example, the Clayton Copula and the Gumbel Copula functions have the characteristics of tail asymmetry, much similar to the property of rate of return in financial products, and thus widely used in research. In 2006, Patton transfer a ARMA(1,10) function into a time varying Copula function by using correlation coefficient \(\rho\), and found the time varying Copula function is suitable for data fitting [6]. In China, the Copula family is applied to some research, for instance the hydrology and the finance fields, and has been proved well for Chinese financial data fitting [7-10]. However, in the research of correlation between stock index future market and stock index spot market, the GARCH analysis is mostly used, and few works are done on the Copula function. In this research, the Copula function is applied to Hushen300 stock index, and focuses on two topics: (1) Fluctuation about Hushen300 stock index during the period before and after its launching; (2) Correlation between the HS300 stock index future market and the market of Hushen300 stock index.

II. BACKGROUND AND PRELIMINARY OF COPULA FUNCTION

A. Copula Function

Copula is the function that connects single variable marginal distribution and multivariable joint distribution. Here, the Sklar theorem is defined as below [11].

The Sklar theorem: For a joint distribution function \(F(x_1, \ldots, x_n)\), there is a Copula function satisfies Eq. (1):

\[
F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))
\]  \(1\)
Where \( x_1, x_2, \ldots, x_N \) are the variables, \( F_i(x_i) \) is the \( i^{th} \) marginal distribution, \( C \) is the abbreviation of Copula function. By using the property of Copula function, we can get the following transfer equation:

\[
C\left(F_1(x_1), \ldots, F_n(x_n), \ldots, F_N(x_N)\right) = f(x_1, \ldots, x_N) \prod_{i=1}^{n} f_i(x_i)
\]  

(2)

Eq. (2) describes the scaling property of Copula function clearly. For a Copula function, if the involved variables are independent to each other, the Copula function is independent, which means the joint distribution is the product of the marginal distributions. On the other hand, if the variables are dependent to each other in some degree, the Copula function describes inner relationship among the variables. Therefore, the Copula function can be regarded as marginal distribution and joint distribution under different conditions, and thus making itself a suitable model in the analysis of financial research.

B. Parameter Estimation Method

Presently, three methods are widely employed in the research of market correlation.

1. The maximum likelihood estimation (MLE)[13]. In this method, the marginal distribution \( F_j(x_j) \) and the Copula function are combined to construct a likelihood function. By using the principle of least logarithm, the parameters are estimated.

2. The non-parametric estimation. This method is also called as the Genest and Rivest estimation method [12], and is mostly cooperated with the Archimedean Copula function. In this estimation, one coefficient, i.e. Spearman \( \rho \) or Kendall \( \tau \), should be calculated in advance. Consequently, the coefficient and the Archimedean Copula function are used to calculate the Copula parameters.

3. Semi-parameter estimation. There are two steps in this method. Firstly, the marginal distribution is calculated by using the kernel density estimation method. Then, the MLE method is adopted to estimate the Copula parameters.

C. Idea of This Paper

The paper will estimate the relevant structure by three steps: firstly, we determine the marginal distribution of HS300 stock index and HS300 stock index future respectively by GARCH (1, 1)-t model [14-16], which is fit for the data of finance. Then two residuals of above models have a probability integral transform into obedience to the value in the range of 0 to 1 of the uniform distribution and become sequence of U and V; secondly, the paper estimates the parameters of static Copula ;in the end ,the parameter of time-varying Copula is also estimated.

1. Confirm the marginal distribution. It is a very important to determine the marginal distribution of two variables, and usually the marginal distribution is assumed a appropriate marginal distribution. The yield of stock index is time-varying, skewness, high-peak and fat-tail, which characters is described well by t-distribution, and GARCH model is able to capture the volatility characters of financial time series, so it is suitable to select GARCH(1,1)-t model for describing the marginal distribution of HS300 stock index future and HS300 stock index.

\[
\begin{align*}
R_t &= \mu + \epsilon_t \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\sqrt{\frac{\hat{\sigma}_t}{\hat{\beta}}} |\psi_{t-1}| &\sim t(v)
\end{align*}
\]

(3)

where \( R_t \) is yield of HS300 stock index for \( t \) , \( \mu \) is the unconditional mean, it assumes that the residuals \( \epsilon_t \) obey \( t \) distribution which degree of freedom is \( v \), \( \omega, \alpha, \beta \) are invariable parameters to be estimated , \( \psi_{t-1} \) is all information at t-1 and t is representative of time. The last outcome of equation 3 is residual series which should not be used to Copula function directly, because of the demand that variable of Copula must obey uniform distribution range 0 to 1, so the probability integral transform is necessary.

2. Select the optimal static Copula function. The candidate functions in this paper are nine Copula functions, which include elliptic Copula (normal Copula and t-Copula), Archimedean Copula and SJC Copula. There is a parameter \( \tau \) in elliptic Copula which has a equation with Kendall \( \tau = \rho = \sin (\pi / 2 \rho) \), we can estimate the parameters (include \( v \) in t-Copula) of normal--Copula function and t-Copula by maximum likelihood estimation (MLE). For the parameter estimation of Archimedean Copula, Genest et al (1995) discussed the method of estimation parameters for Archimedean Copula by Kendall \( \tau \), there are three relationships between Kendall \( \tau \) and three parameters in Archimedean Copula at following table:

<table>
<thead>
<tr>
<th>Function</th>
<th>Clayton</th>
<th>Gumbel</th>
<th>Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kendall ( \tau )</td>
<td>( \alpha / (\alpha + 2) )</td>
<td>( 1 - 1 / \alpha )</td>
<td>( 1 - \frac{2}{\alpha} )</td>
</tr>
</tbody>
</table>

Remark: \( D_i(\alpha) = \int_0^\alpha t \exp(-\alpha t) dt \). The paper selected the best fitting effect between Copula function and empirical Copula by calculating the maximum gap between Copula function and empirical Copula with Kolmogorov-Smirnov test (K-S test).

(3) Parameter estimation for time-varying Copula function. We assume that parameter is constant in the above analysis; next the parameter is allowed to change. Patton proposed time-varying normal-Copula, time-varying t-Copula, time-varying Rotated Gumbel Copula (time-varying RG-Copula) and time-varying Summarized Joe-Clayton Copula (time-varying SJC Copula). He selected a procedure similar to ARMA (1, 1) as trail of change for \( \rho \). The paper
adopts time-varying normal Copula to estimate parameter, as follow is the correlation for time-varying Normal Copula evolution equation:

$$
\rho_t = \frac{1 - e^{-t\times\sum_{i=1}^{10} \Phi^{-1}(u_{t-i})\Phi^{-1}(v_{t-i})}}{1 + e^{-t\times\sum_{i=1}^{10} \Phi^{-1}(u_{t-i})\Phi^{-1}(v_{t-i})}}
$$

Where $\lambda_t = 1 - e^{-t\times\sum_{i=1}^{10} \Phi^{-1}(u_{t-i})\Phi^{-1}(v_{t-i})}$ is a modified logistic transform which is to ensure $\rho_t$ is always in the $[-1, 1]$, and $\Phi^{-1}(\cdot)$ indicate the inverse distribution of standard normal distribution.

III. DATA AND RESULT

A. Data

To examine the correlation structure between stock index future and stock index, we estimate a series of Copula function using daily data from April 16, 2008 to October 23, 2014. The dataset comes primarily from Online Data Section of WIND's website and includes HS300 stock index future and HS300 stock index. The data are arranged into four sequences, they are HS300 stock index from April 16, 2008 to October 23, 2014, HS300 stock index from April 16, 2008 to April 15, 2010, HS300 stock index from April 16, 2010 to October 23, 2014 and HS300 stock index future from April 16, 2010 to October 23, 2014, the length of time for HS300 stock index Future is shorter because it is birth at April 16, 2010. Now the data we gathered is Price, however, the data in model is yield, so a transform is essential.

$$r_t = 100\times((\log(P_t) - \log(P_{t-1}))$$

, $r_t$ and $P_t$ are logarithmic yield and closing price at t time separately.

B. Descriptive Statistics

We can find that the distribution of HS300 stock index is similar to shaped distribution with high-peak and fat-tail from Figure 1, therefore it is assumed as $t$ distribution. Table II reports the statistical results of four series, According to table II we can get the following conclusion: firstly, there is a no significant difference between before April 16, 2010 and after that time on the yield of HS300 stock index; next, there are many similarity such as negative mean, high-peak, fat-tail, non-normal and steady about HS300 stock index future and HS300 stock index, the Pearson correlation $\rho$ on them is up to 0.9448, which represents that they had high correlation.

C. Marginal Distribution

We can find that it is good idea to apply GARCH(1,1)-t (equation (3)) to the marginal distribution of HS300 stock index and HS300 stock index future, and we get table III by MLE on the above two series.

<table>
<thead>
<tr>
<th>time</th>
<th>name</th>
<th>mean</th>
<th>Standard deviation</th>
<th>skewness</th>
<th>kurtosis</th>
<th>JB</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-2014</td>
<td>HS300</td>
<td>-0.015</td>
<td>0.880</td>
<td>-0.153</td>
<td>5.156</td>
<td>185.6***</td>
<td>-30.14***</td>
</tr>
<tr>
<td>2008-2010</td>
<td>HS300</td>
<td>-0.014</td>
<td>1.043</td>
<td>-0.124</td>
<td>4.356</td>
<td>39.62***</td>
<td>-21.69***</td>
</tr>
<tr>
<td>2010-2014</td>
<td>HS300</td>
<td>-0.025</td>
<td>0.653</td>
<td>-0.305</td>
<td>4.482</td>
<td>44.63***</td>
<td>-21.88***</td>
</tr>
<tr>
<td>2010-2014</td>
<td>HS300 Future</td>
<td>-0.027</td>
<td>0.677</td>
<td>-0.117</td>
<td>4.976</td>
<td>74.10***</td>
<td>-23.07***</td>
</tr>
</tbody>
</table>

Note: (***) indicates 1% confidence level significantly

<table>
<thead>
<tr>
<th>sample</th>
<th>$\mu$</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\nu$</th>
<th>LLR</th>
<th>K-S</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS300</td>
<td>-0.0329 (0.0268)</td>
<td>0.00769 (0.0085)</td>
<td>0.0393 (0.0258)</td>
<td>0.9496 (0.0319)</td>
<td>3.757 (1.039)</td>
<td>-435.89</td>
<td>0.022</td>
<td>0.82</td>
</tr>
<tr>
<td>HS300</td>
<td>-0.0183 (0.0283)</td>
<td>0.0067 (0.0102)</td>
<td>0.0163 (0.0158)</td>
<td>0.9676 (0.0358)</td>
<td>6.0917 (2.042)</td>
<td>-433.27</td>
<td>0.036</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note: $\mu$ is mean, $\omega, \alpha, \beta$ are constant, $\nu$ is the degree of freedom for $t$ distribution, LLR is likelihood Rate; K-S is Kolmogorov-Simonov test.
TABLE IV  ESTIMATION FOR STATIC COPULA

<table>
<thead>
<tr>
<th>type</th>
<th>LL</th>
<th>AIC</th>
<th>kappa</th>
<th>[tauL, tauU]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal-Copula</td>
<td>-491.75</td>
<td>-983.49</td>
<td>0.94243</td>
<td>[0, 0]</td>
</tr>
<tr>
<td>Clayton-Copula</td>
<td>-436.73</td>
<td>-873.46</td>
<td>5.3939</td>
<td>[0.8794, 0]</td>
</tr>
<tr>
<td>Rotated-clayton-Copula</td>
<td>-398.87</td>
<td>-797.13</td>
<td>4.8254</td>
<td>[0, 0.86619]</td>
</tr>
<tr>
<td>Frank-Copula</td>
<td>-431.51</td>
<td>-863.01</td>
<td>23.539</td>
<td>[0, 0]</td>
</tr>
<tr>
<td>Gumbel-Copula</td>
<td>-482.6</td>
<td>-965.19</td>
<td>4.5941</td>
<td>[0, 0.8371]</td>
</tr>
<tr>
<td>Rotated-gumbel-Copula</td>
<td>-505.11</td>
<td>-1010.2</td>
<td>4.8178</td>
<td>[0.8453, 0]</td>
</tr>
<tr>
<td>t-Copula</td>
<td>-519.72</td>
<td>-1039.1</td>
<td>3.179</td>
<td>[0.7567, 0.7567]</td>
</tr>
<tr>
<td>SJC Copula</td>
<td>-501.23</td>
<td>-1002.5</td>
<td>4.8254</td>
<td>[0, 0.8552]</td>
</tr>
</tbody>
</table>

Note: LL is the log likelihood, AIC is Akaike statistic, kappa is parameters in model, [tauL, tauU] are respectively the lower and upper tail dependence coefficient.

In general, GARCH model is very suit for describing the cluster effect of volatility, mean equation and variance equation are both estimated well. The yields of HS300 stock index and HS300 stock index future are negative, which indicate that major investors are deficit in this period. From the table III, the degree of freedom for HS300 stock index future is 6.0917, which is bigger than that of HS300 stock index who is 3.757, through the result; we can consider that HS300 stock index future has thicker tail than HS300 stock index, so it gets more extreme events and risk. K-S statistics and P-value in table III is based on probability integral transform to marginal distribution, the result of Likelihood Rate(LLR) and K-S statistics indicates that GARCH(1,1)-t is appropriate for both HS300 stock index and HS300 stock index future, the only divergence is the degree of freedom.

D. Copula Estimation

In order to study the relationship between HS300 stock index and HS300 stock index future, the paper selects nine static Copula as the index of describing dependency structure for two series, the result is in table IV. From table 4, we find that t-Copula is the best fitting one among nine different Copula function, and $\rho$ is estimated to 0.9489, that is a small difference with real correlation which is 0.9448. It illustrate that fitting is effective and strong correlation between HS300 stock index and HS300 stock index future; and there is a thicker symmetry tail.

In recent years scholars point out that the relationship of different market is changing with different related structures in different while, that is because of dynamic and time-varying financial market. This paper select the two most common time-varying Copula function (time-varying normal Copula and the time-varying Rotated Gumbel Copula function) as representative of time-varying ellipse Copula and Archimedean, the results are as follows:

TABLE V  ESTIMATION FOR TIME-VARYING COPULA

<table>
<thead>
<tr>
<th>Type</th>
<th>$\omega_\rho$</th>
<th>$\beta_\rho$</th>
<th>$\alpha_\rho$</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-Vary Normal Copula</td>
<td>5</td>
<td>-0.7739</td>
<td>-0.6995</td>
<td>-502.13</td>
</tr>
<tr>
<td>Tim-Varying RG Copula</td>
<td>2.3421</td>
<td>-0.0018</td>
<td>-4.9999</td>
<td>-512.27</td>
</tr>
</tbody>
</table>

In Table V, the LL value of time-varying Copula is less than that of static Copula, which indicates that time-varying Copula function is more suitable for the data. The time-varying Rotated Gumbel Copula is superior to time-varying normal Copula, time-varying Copula has further comparative fitting effect corresponding static Copula, the concrete results as below.

The horizontal line in figure 2 refers to parameter $\rho$ of static Copula and wave line is $\rho$ of time-varying Copula, we can observe that the fluctuations in figure (b) are more...
violent than that of figure (a). Therefore, we can consider there is significantly right-tail influence in the data.

IV. CONCLUSION

Just before and after HS300 stock index future’s official launching, fluctuation of the former is much obvious as compared with the latter, but the difference is not remarkable, so volatility clustering of HS300 stock index future will not owe to the official launching of HS300 Stock Index Future. In nine static Copula, t-Copula is the most suitable for the data, which has a high and symmetrical tail dependence. Time-varying Copula gets better fitting effect comparing to static Copula, especially for time-varying Rotated Gumbel Copula, which has log likelihood superior to other Copula functions, consequently, we will use this method to more.

REFERENCES