Comparison of Estimators of Tail Dependence Coefficient

Shide Ou
School of Statistics,
Renmin University of China,
Beijing, 100872, China
oushide@yahoo.com.cn

Danhui Yi
School of Statistics,
Renmin University of China,
Beijing, 100872, China
Corr. author: xueyi905@yahoo.com.cn

Abstract—To find effective estimations of tail dependence, we present the estimators of upper tail dependence coefficient by using survival copula. We research to two problems by using the samples from $t$-copula. Firstly, do the estimators estimate effectively the upper tail dependence coefficients of copula? Which is the best among the estimators? Secondly, if sample isn’t from true distribution, do the estimators have model risk? Simulation shows that the estimators can estimate effectively the tail dependence coefficients of $t$-copula. By comparison of estimators, a better estimator is found out among the estimators. Therefore if sample is from true distribution, the estimators can estimate effectively upper tail dependence coefficients. However, if sample isn’t from underlying copula, the estimators have model risk.

Keywords- tail dependence; copula; survival copula

I. INTRODUCTION

Tail dependence coefficient is a measure describing the associated movement of tail extreme value of bivariate. Tail dependence of copula function can effectively capture the dependence between stocks (see for example Joe (1997)). However, how do we estimate tail-dependence coefficient (TDC)? We know some methods to TDC estimation in Frahm et al. (2005) and Dobric et al. (2005). Frahm et al. (2005) showed that the use of parametric margins instead of empirical margins bears a model risk and may lead to wrong interpretations of dependence structure. Therefore, for constructing an effective portfolio we need to estimate precisely TDC.

Tail dependence is a hot topic in dependence structure. Scholars are more and more focusing on the estimation of TDC. Embrechts et al. (2002) researched the correlation and the dependence in risk management and pointed out some pitfalls. Coles et al. (1999) studied the dependence measure of extreme value. Caperaa et al. (1997) researched the nonparametric estimation for bivariate extreme value copulas. Schmidt et al. (2006) studied the nonparametric estimation of tail dependence. To explore the estimation of TDC, Dobric and Schmid (2005) presented three estimators of TDC and compared the estimators. These estimators are unbiased. However, they didn’t find which estimator is the best. Frahm et. al. (2005) compared three estimators for upper TDC and found that the estimator $\hat{\lambda}_{CFG}$ is better than others.

Is the estimator $\hat{\lambda}_{CFG}$ better than the estimators mentioned in Dobric and Schmid (2005)? Since the upper TDC of copula is equal to the lower TDC of survival copula (see for example Frahm et. al., 2005, page 3), to find effective estimators of upper TDC we present the estimators of upper TDC. Then we compare the estimators and $\hat{\lambda}_{CFG}$. By using the samples from $t$-copula, we research two problems. Firstly, do the estimators estimate effectively the upper tail dependence coefficient of copula? Which is the best among the estimators? Secondly, if the sample is from Gumbel copula or Clayton instead of $t$-copula, is there model risk for the estimators? The problem stems from the practice that we don’t know what distribution sample is from.

This paper is organized by the structure as follows. Section 2 addresses the concepts of tail dependence and empirical copula, and states the estimators of TDC (see for example Dobric and Schmid, 2005; Frahm et. al., 2005). Section 3 presents the estimators of upper TDC by survival copula. Simulation and comparison of the estimators are also included in the section. Section 4 concludes the paper.

II. PREVIOUS ESTIMATORS OF TDC

A. The definition of TDC

We know the definition of TDC from Joe (1997). Let $X$ and $Y$ be continuous random variables with margins $F$ and $G$ and a copula function $C$. Then the upper TDC and the lower TDC are as follows

$\lambda_u = \lim_{u \to 1} P(Y > G^{-1}(u) \mid X > F^{-1}(u))$

$\lambda_l = \lim_{u \to 0} P(Y < G^{-1}(u) \mid X < F^{-1}(u))$

where $\tilde{C}(u_i, u_j) = u_i + u_j - 1 + C(1-u_i, 1-u_j)$ is a survival copula. If $\lambda_u = 0$ or $\lambda_l = 0$, then $X$ and $Y$ are independent.

For a bivariate $t$-copula,

$C(u, v; \rho, \alpha) = T_{\rho, \alpha}(T_{u \rho}^{-1}(u), T_{v \rho}^{-1}(v))$
$c(u, v; \rho, \alpha) = \rho^{-1/2} \Gamma\left(\frac{\alpha + 2}{2}\right) \Gamma\left(\frac{\alpha}{2}\right) \left[1 + \frac{\rho^2 + \frac{2}{\alpha}(1 - \rho^2)}{\alpha(1 - \rho^2)}\right]^{-\alpha/2}$

where $\rho_i = t_{\alpha_i}^{-1}(u_i), \rho_2 = t_{\alpha_2}^{-1}(v).$ Then

$\lambda_n = \lambda = \frac{\lambda}{\alpha_1} = \frac{\alpha_1}{\alpha_2}, \rho \in [-1, 1]. \ 
\tau_{\alpha_1}$ (see for example Embrechts et al., 2002) is a $t$-distribution function with the $\alpha + 1$ degree of freedom.

For a Clayton copula,

$C(u, v) = \left(\frac{u^{-\theta} + v^{-\theta} - 1}{\theta}\right)^{-1/\theta}, \theta \in (0, \infty),$

it has $\lambda_n = 0$, $\lambda = 2^{-1/\theta}$. For a Gumbel copula, $\phi(x; \theta) = (-\ln x)^{1/\theta}, \theta \in (0, 1],$

$C(u_1, \cdots, u_k) = \exp\left[-\sum_{i=1}^{k}(-\ln u_i)^{1/\theta}\right]$.

The tail correlation coefficients are $\lambda_n = 2 - 2^{-1/\theta}$ and $\lambda = 0$. It means that Gumbel copula is more sensitive to the upper tail of distribution, while isn’t sensitive to the lower tail.

Similarly to unitary empirical distribution, empirical copula (see for example Bouye et al., 2000) can be described by the following representation. Let $11{(x, y), (u, v)}$,…, $(x, Y), \,(X, y)$ be from $(X, Y), \, Y \leq i \leq X \leq X, \, Y \leq i \leq Y$ are the order statistic of $X$ and $Y$ respectively. Then the empirical copula of the sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ is

$\hat{C}_n(l, j/n) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq X(j/n), Y_i \leq Y(j/n)).$

If $i=0$ or $j=0$, then $\hat{C}_n(l, n/n) = 0$, where $I$ is an indicative function.

### B. The estimation of TDC

For the nonparametric estimation of lower TDC Dobric and Schmid (2005) introduced three estimators as follows

$\hat{\lambda}_n^{(1)} = k \cdot \hat{C}_n(k/k, n/n)$

$\hat{\lambda}_n^{(2)} = \sum_{i=1}^{k} \hat{C}_n(i/n, i/n) - \sum_{i=1}^{k} \hat{C}_n(i/n, i/n)$

$\hat{\lambda}_n^{(3)} = \sum_{i=1}^{k} \frac{\hat{C}_n(i/n, i/n) - (i/n)^2}{(i/n)^2} - \sum_{i=1}^{k} \frac{\hat{C}_n(i/n, i/n) - (i/n)^2}{(i/n)^2}$

where $k$ is determined by $n$, generally $k = \sqrt{n}, \hat{C}_n$ is an empirical copula. We also refer to (1)-(3) as the upper TDC based on empirical copula.

Estimators (1)-(3) are asymptotically unbiased, i.e.

$\lim_{n \to \infty} E(\hat{\lambda}_n^{(i)}) = \lambda.$

Whether $\hat{\lambda}_n^{(1)}$ effectively estimates the lower TDC $\lambda$ or not? Dobric and Schmid (2005) researched the problem using the samples from the following three copulas.

Setting 1: $C_\lambda(u, v) = \lambda \min(u, v) + (1 - \lambda)uv, \ 0 \leq \lambda \leq 1$. The copula has $\hat{\lambda}_n = \lambda$.

Setting 2: $C_\theta(u, v) = (u^\theta + v^\theta - 1)^{-1/\theta}, \ 0 < \theta < \infty$. The copula has $\hat{\lambda}_n = 2^{-1/\theta}$.

Setting 3: The family of Gauss copulas

$C_\rho(u, v) = \Phi\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$

where $\Phi_\rho$ is a bivariate normal distribution with correlation $\rho$. The family of Gauss copulas has $\hat{\lambda}_n = \hat{\lambda}_n = 0$.

Their research showed that the estimation biases of (1)-(3) depend on underlying copula, sample size $n$, and the true $\lambda$. However, they didn’t obtain the conclusion which is the best estimator.

For the estimation of upper TDC Frahm et al. (2005) compared the following estimators. The first (introduced in Coles et al. (1999)) is

$\hat{\lambda}_n^{LOG} = 2 - \frac{\log \hat{C}_n(m-h, m-h)}{\log (m-h)}$, $0 < h < m (4)$

where $\hat{C}_n(u, v) = \frac{1}{m} \sum_{i=1}^{m} I(R_i / m \leq u, R_j / m \leq v)$, $I$ is an indicative function, $R_i$ and $R_j$ are respectively the rank of maximum $X^+_j$ and $Y^+_j, j = 1, \cdots, m, l = n/m, h$ is a threshold.

The second (a special case in Joe et al. (1992)) is

$\hat{\lambda}_n^{SEC} = 2 - \frac{1 - \hat{C}_n(n-k, n-k)}{n - k}$, $0 < k \leq n (5)$

The third (motivated in Caperaa et al. (1997)) is

$\hat{\lambda}_n^{CVG} = 2 - 2 \sum_{i=1}^{m} \log \left[\log \left(\frac{U_i - \log V_i}{\log \frac{1}{\max(U_i, V_i)}}\right)\right] (6)$

where $\{(U_i, V_i), \cdots, (U_m, V_m)\}$ is the sample from a copula $C$.

Whether (4)-(6) estimate the upper TDC or not? Frahm et al. (2005) performed simulations to the samples from the distributions such as elliptical copula, t-copula, Archimedean copula and symmetric Gumbel copula (defined by Tawn (1998)). Comparing the three estimators, they found out that if the assumption of model isn’t true, the estimation effect isn’t good. Estimator (6) is the best among the three estimators.

### III. SIMULATION STUDY

Estimators (1)-(3) are to the estimation of lower TDC. Since the upper TDC of copula is equal to the lower TDC of...
survival copula, vice versa. Namely, \( \hat{\lambda}_U(C) = \hat{\lambda}_U(C) \), \( \hat{\lambda}_U(C) = \hat{\lambda}_U(C) \). To estimate upper TDC we change (1)-(3) into the following representations

\[
\hat{\lambda}_U^{(1)}(C) = \hat{\lambda}_U^{(1)}(C) = \hat{\lambda}_U^{(1)} = \left( \frac{k}{n} \right)^{-1} \cdot \hat{C}_n \left( \frac{k}{n} \right) \\
= \left( \frac{k}{n} \right)^{-1} \left[ 2k - 1 + C \left( 1 - \frac{k}{n} \right) \right] \\
\hat{\lambda}_U^{(2)}(C) = \hat{\lambda}_U^{(2)}(C) = \hat{\lambda}_U^{(2)} = \left( \sum_{i=1}^{k} \left( \frac{i}{n} \right) \right)^{-1} \sum_{i=1}^{k} \left( \frac{i}{n} \right) \hat{C}_n \left( \frac{i}{n} \right) \\
\hat{\lambda}_U^{(3)}(C) = \hat{\lambda}_U^{(3)}(C) = \hat{\lambda}_U^{(3)} = \sum_{i=1}^{k} \left( \frac{i}{n} \right) \left( \frac{i}{n} \right) \hat{C}_n \left( \frac{i}{n} \right) \\
= \sum_{i=1}^{k} \left( \frac{i}{n} \right) \left( \frac{i}{n} \right) \hat{C}_n \left( \frac{i}{n} \right) \left( \frac{i}{n} \right) \\
\hat{\lambda}_U^{(4)}(C) = \hat{\lambda}_U^{(4)}(C) = \hat{\lambda}_U^{(4)} = \sum_{i=1}^{k} \left( \frac{i}{n} \right) \left( \frac{i}{n} \right) \hat{C}_n \left( \frac{i}{n} \right) \left( \frac{i}{n} \right)
\]  

where \( \hat{C}_n \left( \frac{i}{n} \right) = \frac{2i - 1}{n} + C \left( 1 - \frac{i}{n} \right) \).

Whether (6)-(9) estimate effectively the upper TDC of \( t \)-copula or not? Which is the best estimator? We use R-software package (see for example Jun Yan, 2007; Ivan Kojadinovic, Jun Yan, 2010) to research the problems. Because returns of financial asset obey heavy tail distribution, we select the sample from \( t \)-copula which has the degree of freedom \( \alpha = 1.5 \), the correlation coefficient \( \rho = 0.5 \) and the size \( n = 1000 \). The TDC \( \hat{\lambda}_U = \hat{\lambda}_U = 0.4405996 \) are easily computed from the \( t \)-copula. Via designing a program of empirical copula and using R-software package, we test to the simulation that the number of replications is 500. The results are reported as follows

<table>
<thead>
<tr>
<th>estimator</th>
<th>Mean</th>
<th>Std.</th>
<th>Abs. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\lambda}_U^{(1)} )</td>
<td>0.446787879</td>
<td>0.114539995</td>
<td>0.006188279</td>
</tr>
<tr>
<td>( \hat{\lambda}_U^{(2)} )</td>
<td>0.444492936</td>
<td>0.127288191</td>
<td>0.003893336</td>
</tr>
<tr>
<td>( \hat{\lambda}_U^{(3)} )</td>
<td>0.432129546</td>
<td>0.12665904</td>
<td>0.008470054</td>
</tr>
<tr>
<td>( \hat{\lambda}_U^{(4)} )</td>
<td>0.45142159</td>
<td>0.02395559</td>
<td>0.01082199</td>
</tr>
</tbody>
</table>

We can conclude from Table I and Figures 1 to 4 that \( \hat{\lambda}_U^{(1)} \), \( \hat{\lambda}_U^{(2)} \), \( \hat{\lambda}_U^{(3)} \) and \( \hat{\lambda}_U^{(4)} \) estimate effectively the upper TDC of \( t \)-copula. In the sense of standard deviation, \( \hat{\lambda}_U^{(4)} \) is the best among the estimators, while significant difference between \( \hat{\lambda}_U^{(1)} \), \( \hat{\lambda}_U^{(2)} \) and \( \hat{\lambda}_U^{(3)} \) isn’t found out.

If true distribution isn’t \( t \)-copula, however, we use the sample from \( t \)-copula to fit other copulas. Under the
assumption, do the estimators (6)-(9) have model risk? The question stems from the practice that we don’t know what distribution sample is from. We test to the problem using the sample from $t$-copula with $\alpha = 1.5, \rho = 0.5, n=1000$. The sample is used to fit to a Gumbel copula. The number of replications is 500. We obtain the mean of the upper TDC, $\hat{U}_{\lambda} = 0.4335799$ with standard deviation 0.02495657. The result approaches almost to that of $\hat{U}_{\lambda}^{(1)}, \hat{U}_{\lambda}^{(2)}, \hat{U}_{\lambda}^{(3)}$ and $\hat{U}_{\lambda}^{CFG}$. If the same sample is used to fit to a Clayton copula, the lower TDC is obtained 0.4364681 (standard deviation 0.04086731). However, the estimations of $\hat{U}_{\lambda}^{(1)}, \hat{U}_{\lambda}^{(2)}, \hat{U}_{\lambda}^{(3)}$ and $\hat{U}_{\lambda}^{CFG}$ don’t approach to the upper TDC of Clayton copula $\hat{U}_{\lambda} = 0$. Thus if the sample isn’t from underlying copula, the estimations of $\hat{U}_{\lambda}^{(1)}, \hat{U}_{\lambda}^{(2)}, \hat{U}_{\lambda}^{(3)}$ and $\hat{U}_{\lambda}^{CFG}$ exist model risk.

IV. CONCLUSION

This paper presents the estimators $\hat{U}_{\lambda}^{(1)}, \hat{U}_{\lambda}^{(2)}, \hat{U}_{\lambda}^{(3)}$ by the relationship between copula function and survival copula function. By simulations, we find that the estimators $\hat{U}_{\lambda}^{(1)}, \hat{U}_{\lambda}^{(2)}, \hat{U}_{\lambda}^{(3)}$ and $\hat{U}_{\lambda}^{CFG}$ effectively estimate the TDC if sample is from true distribution. Although the sample is from $t$-copula, the estimators effectively estimate also the upper TDC of Gumbel copula $\hat{U}_{\lambda} = 0$. However, if the sample isn’t from underlying copula, the estimators have model risk. Via comparison of the estimators $\hat{U}_{\lambda}^{(1)}, \hat{U}_{\lambda}^{(2)}, \hat{U}_{\lambda}^{(3)}$ and $\hat{U}_{\lambda}^{CFG}$, we find out that the estimator $\hat{U}_{\lambda}^{CFG}$ is the best among these estimators. However, it is necessary to test sample distribution before using the estimators.

REFERENCES