Comparison of Two Methods to Check Copula Fitting

Yan Fang, Lisa Madsen, and Ling Liu

Abstract—**The novelty of this paper is how we quickly and accurately choose the true copula if we know the true copula exists and is one from a given copula families. Goodness-of-fit test approaches are the most popular procedure to select the best copula for a particular data set. However, the goodness-of- fit tests just check whether a parametric copula is rejected or is not rejected for a specific data. This paper investigates the feasibility of using Akaike Information Criterion (AIC) to choose a copula model from a series of candidate copula models. If the true copula is among the candidate families, the copula with the least AIC should be the true one; otherwise, the copula with the least AIC will be the best one. We do simulations to show that AIC method is generally faster and more precise than the multiplier goodness-of-fit test method.**

Index Terms—**AIC, empirical copula, goodness-of-fit, pseudo-likelihood, pseudo-observations, semi-parametric.**

I. Introduction

Copula have proved to be a very useful tool in the analysis of dependency structures in modern finance and insurance. It allows us to model the multi-dependence without specifying the marginal distributions. The concept of copula was based on Sklar's theorem (Sklar [50]): any multivariate distribution can be decomposed into a copula and its marginal; if the marginal distributions are continuous, then the copula is unique; otherwise, it is uniquely determined on its corresponding range. One attractive property of copula is their invariance under strictly increasing transformations of the margins. For a thorough literature review of copula, see Nelsen [42]. Copula was first used in financial applications by Embrechts et. al. [15]. Since then the application on copula theory in finance and economics has grown tremendously. Moreover, practical applications of this modeling approach are found in fields such as finance (Nikoloulopoulos et. al. [43]; Fang and Madsen [16]), hydrology (Genest et. al. [23]), public health and medical (Winkelmann [52]) and actuarial science (Frees and Valdez [19]; Otani and Imai [44]).

Although copula are becoming more and more popular among academics, and many copula families have been suggested, however, selecting the functional form for copula is an open question in the literature. Our question is how to quickly and precisely choose a true parametric copula from a series of given candidate families where the true copula is included. Many authors recently proposed for using the goodness-of-fit (GOF) test (Genest and Rivest [20] and Fermanian [17], etc.) for choosing the copula models. Most of them have been made to test the null hypothesis, e.g., Genest and Rivest [20], Shih and Louis [49], Breymann et. al. [8], Fermanian [17], Genest et. al. [22], Dobrić and Schmid [12], Genest and Rémillard [25], Genest et. al. [26], Kojadinovic and Yan [36], Kojadinovic et. al. [35], and so on. Among the existing approaches, some approaches are full multivariate approaches while several approaches are dimension reduction approaches which reduce the multivariate problem to a univariate problem, and then apply some univariate test. The former one leads to computationally exhaustive for high dimensional problems, while the latter one leads to numerically efficient approaches even for high dimensional problems.

Among the dimension reduction approaches, it is common to apply standard univariate statistics such as Kolmogorov- or Cramér-von Mises type statistics. According to the large scale simulations carried out in Genest et. al. [26], the most powerful version of test based on the Cramér-von Mises statistic is the parametric bootstrap-based GOF tests. An approximate p-value for this test is obtained by means of a parametric bootstrap whose validity was recently shown by Genest and Rémillard [25]. Since each parametric bootstrap iteration requires both random number generation from the fitted copula and estimation of the copula parameter, hence the essential inconvenience of this approach is its high computational cost. Moreover, the larger the sample size, the more restrictive the application of parametric bootstrap-based GOF tests becomes. In order to circumvent this very high computational cost, Kojadinovic and Yan [36] proposed a fast large-sample testing procedure based on multiplier central limit theorems. From now on, we will call this testing procedure as the multiplier GOF test method. However, the problem with this approach is that in order to get a valid estimation of the parameter the multiplier GOF test requires the sample size of at least 300. Another issue with the multiplier GOF test is that it can not provide the best copula from a series of candidate families.

The essence of GOF test is checking whether the unknown copula actually belongs to the chosen parametric copula family or not. It is only used to check whether we should reject or fail to reject the chosen copula, thus we cannot use the GOF test to chose the true parametric copula from a series of copula model, even when the true copula is among the given series. The contribution of the present paper is to accurately and precisely choose an exact parametric copula for a given data when the true copula exists and is among a candidate families. And this paper will be based on the assumption that the candidate families contain the true copula. Here,

Manuscript received Oct 7, 2013; revised Dec 06, 2013. This work was supported by Shanghai University of International Business and Economics under Grant Z085YYJJ13063 and Grant YC-XK-13108, Leading Academic Discipline Project of Shanghai Municipal Education Commission with Project Number: J51201, and Humanities and Social Sciences Research Project of Ministry of Education with Grant Number: 10YJC790171.

Y. Fang is with the Financial School, Shanghai University of International Business and Economics, Shanghai, 201620, China. E-mail: yiffanyfang@163.com.

L. Madsen, Statistics Department, Oregon State University, OR, 97331 USA. E-mail: madsenl@onid.orst.edu.

L. Liu, Financial School, Shanghai University of International Business and Economics, Shanghai, 201620, China. E-mail: liuling600600@gmail.com.

we introduce Akaike Information Criterion(AIC) developed by Akaike [2], to chose a true parametric copula model. Although AIC does not provide us with any understanding of the power of the decision rule employed, it provides the comparison of the fitting of different copula, therefore it can be used as a tool for copula selection. One of the benefits of AIC is that it takes less time than the multiplier GOF test method. Furthermore, the smaller the AIC, the better the copula is.

The purpose of this paper is to present a critical review of AIC as a tool for choosing copula from a series of candidates and to compare the relative effectiveness of AIC method with the multiplier GOF test method through simulation study, which involves a large number of copula alternatives and dependence conditions. We provide some basic theory regarding copulas in Section II. Both multiplier GOF test and AIC are reviewed in Section III. The simulation results from five one-parameter copula families, specifically, Clayton, Frank, Gumbel, normal and t copulas, are presented in Section IV. A brief conclusion is given at the end.

II. Basic theory

A. Copula basics

The definition of a p-dimensional copula is a multivariate distribution C with uniform $(0,1)$ margins. According to Sklar's theorem (Sklar [50]), any multivariate distribution function H with the marginal cumulative distribution functions (cdfs) $\mathrm{F}_1,\ldots,\mathrm{F}_p$ can be written as

$$\mathrm{H}(\mathbf{x}) = \mathrm{C}\left(\mathrm{F}_1(x_1),\ldots,\mathrm{F}_p(x_p)\right), \mathbf{x}\in\mathbb{R}^p, \tag{1}$$

for some copula C. If all the margins are continuous, then C is unique. Therefore, the copula of the joint distribution function for a random vector $\mathbf{X}=(X_1,\ldots,X_p)^T$ may be derived from Equation (1), i.e.,

$$\mathrm{C}(u_1,\ldots,u_p) = \mathrm{H}\left(\mathrm{F}_1^{-1}(u_1),\ldots,\mathrm{F}_p^{-1}(u_p)\right), \tag{2}$$

where $\mathrm{F}_i^{-1}(\cdot)$ is the inverse of the marginal cdfs and $u_i \in [0,1], \forall i=1,\ldots,p$. The copula density is given by

$$\mathrm{c}(u_1,\ldots,u_p) = \frac{\mathrm{h}(\mathrm{F}_1^{-1}(u_1),\ldots,\mathrm{F}_p^{-1}(u_p))}{\prod_{i=1}^{p}\mathrm{f}_i(\mathrm{F}_i^{-1}(u_i))}, \tag{3}$$

where $\mathrm{f}_i(\cdot)$ is the probability density function (pdf) for the variable $\mathrm{F}_i^{-1}(u_i)$ and $\mathrm{h}(\cdot)$ is the joint pdf for multivariate $(\mathrm{F}_1^{-1}(u_1),\ldots,\mathrm{F}_p^{-1}(u_p))$.

B. Estimation

When considering the problem of estimating the parametric multivariate density models, usually there are two broad approaches to estimate the dependence parameter θ. They differ mainly in the assumption about the parametric margins or the non-parametric margins. Copula is mostly based on a parametric copula and the non-parametric marginal distributions, i.e., the semi-parametric copula. For copula selection, we are only interested in the fit of the copula. We do not wish to introduce any distributional assumptions for the margins. Hence, we will use the semi-parametric copula or use empirical margins to transform the observed data set into the observed copula.

Assume that the unknown copula C belongs to an absolutely continuous parametric family $\mathcal{C}_0=\{\mathrm{C}_\theta : \theta\in\mathcal{O}\}$, where $\mathcal{O}$ is an open subset of $\mathbb{R}^q$ for some $q=\{1,2,3,\ldots\}$ and the vector of copula parameters $\theta=(\theta_1,\ldots,\theta_q)$ is estimated from the random sample $(\mathbf{X}_1,\ldots,\mathbf{X}_n)$. When estimating the parameters for a semi-parametric copula, a natural estimation method is the pseudo-likelihood approach introduced in Genest et. al. [21] and Shih and Louis [49]. It consists of maximizing the log pseudo-likelihood function $l(\theta)$, namely,

$$\begin{aligned}\theta_n &= \arg\max_{\theta} l(\theta)\\ &= \arg\max_{\theta}\sum_{i=1}^{n}\log \mathrm{c}_\theta\{\hat{U}_{i1},\ldots,\hat{U}_{ip}\},\end{aligned} \tag{4}$$

where c_θ is the density function of the parametric copula $\mathrm{C}_\theta\in\mathcal{C}_0$, and the $\hat{\mathbf{U}}_i=(\hat{U}_{i1},\ldots,\hat{U}_{ip})$ are the pseudo-observations or the re-scaled empirical distribution of $\mathbf{X}_i=(X_{i1},\ldots,X_{ip})$, namely,

$$\hat{U}_{ij} = \frac{R_{ij}}{n+1}, \tag{5}$$

where R_{ij} is the rank of X_{ij} among $(X_{1j},\ldots,X_{nj})$.

C. Empirical Distribution

According to Deheuvels [9], a consistent estimation of the underlying copula is possible via the empirical copula, which also can be described as the distribution function of the sample of the normalized ranks. The empirical copula for n pseudo-observations $\mathbf{U}_1,\ldots,\mathbf{U}_n$ is given by

$$\mathrm{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(U_{i1}\le u_1,\ldots,U_{ip}\le u_p), \tag{6}$$

where $\mathbf{u}=(u_1,\ldots,u_p)\in[0,1]^p$, $\mathrm{I}(\cdot)$ is the indicator function, taking the value 1 if $U_{ij}\le u_j$ and 0 otherwise.

III. The multiplier GOF test method and the AIC method

We now briefly introduce the procedures for both the multiplier GOF test method and the AIC method, respectively.

A. Multiplier GOF test method

This subsection describes the multiplier GOF test method introduced by Kojadinovic et. al. [35]. The multiplier GOF test method is based on the empirical copula (Deheuvels [9] and Deheuvels [11]), which is a consistent estimator of the unknown copula C. The cardinal principle of this test has been studied by Deheuvels [10] and Genest and Rémillard [24]. The idea is to compare the empirical copula $\mathrm{C}_n(\mathbf{u})$ defined in Equation (6) with the parametric copula $\mathrm{C}_\theta(\mathbf{u})$'s estimator, $\mathrm{C}_{\theta_n}(\mathbf{u})$, which is obtained by assuming that $\mathcal{H}_0 : \mathrm{C}\in\mathcal{C}_0=\{\mathrm{C}_\theta : \theta\in\mathcal{O}\}$. θ_n defined in Equation (4) is an estimator of θ. The natural way is to consider the distance between empirical and null hypothesis distribution functions. That is, under suitable regularity conditions, the empirical copula process is

$$\mathbb{C}_n(\mathbf{u}) = \sqrt{n}\{\mathrm{C}_n(\mathbf{u}) - \mathrm{C}_{\theta_n}(\mathbf{u})\}, \tag{7}$$

where $\mathbf{u} = (u_1, \ldots, u_p) \in [0,1]^p$. The test statistic, the Cramér-von Mises Statistic, is defined as

$$\begin{aligned} \mathrm{S}_n &= \int_{[0,1]^p} \mathbb{C}_n^2(\mathbf{u}) d\mathrm{C}_n(\mathbf{u}) \\ &= \sum_{i=1}^{n} \{\mathrm{C}_n(\hat{\mathbf{U}}_i) - \mathrm{C}_{\theta_n}(\hat{\mathbf{U}}_i)\}^2. \end{aligned} \tag{8}$$

The asymptotic distribution of test statistic S_n derived from the process $\mathbb{C}_n(\mathbf{u})$ depends on the unknown distribution $\mathrm{C}_{\theta_n}(\mathbf{u})$, hence, Kojadinovic et. al. [35] proposed the multiplier GOF test which is a fast large-sample testing procedure based on the multiplier central limit theorems inspired by Rémillard and Scaillet [46].

Let's define a sequence of the distributed processes $\{\mathbb{J}_i\}_{i=1}^n$, that is,

$$\begin{aligned} \mathbb{J}_i(\mathbf{u}) &= \alpha_\theta(\mathbf{u}) - \Theta \times \dot{\mathrm{C}}_\theta(\mathbf{u}) \\ &\quad - \sum_{j=1}^{p} \mathrm{C}_\theta^{[j]}(\mathbf{u}) \alpha_\theta(1, \ldots, 1, u_j, 1, \ldots, 1) \end{aligned}$$

with

$$\alpha_\theta(\mathbf{u}) = I(\mathbf{U}_i \le \mathbf{u}) - \mathrm{C}_\theta(\mathbf{u}),$$

$$\dot{\mathrm{C}}_\theta(\mathbf{u}) = \frac{\partial \mathrm{C}_\theta(\mathbf{u})}{\partial \theta},$$

$$\begin{aligned} \mathrm{C}_\theta^{[j]}(\mathbf{u}) &= \frac{\mathrm{C}_\theta(u_1, \ldots, u_j + n^{-1/2}, \ldots, u_p)}{2n^{-1/2}} \\ &\quad - \frac{\mathrm{C}_\theta(u_1, \ldots, u_j - n^{-1/2}, \ldots, u_p)}{2n^{-1/2}}, \end{aligned}$$

and

$$\begin{aligned} \Theta &= \left[E_{c_\theta} \left\{ \frac{\dot{c}_\theta(\mathbf{u}) \dot{c}_\theta^T(\mathbf{u})}{c_\theta^2(\mathbf{u})} \right\} \right]^{-1} \times \\ &\left[\frac{\dot{c}_\theta(\mathbf{U}_i)}{c_\theta(\mathbf{U}_i)} - \sum_{j=1}^{p} \int_{[0,1]^p} I_j \times \frac{c_\theta^{(j)}(\mathbf{u})}{c_\theta(\mathbf{u})} \frac{\dot{c}_\theta(\mathbf{u})}{c_\theta(\mathbf{u})} d\mathrm{C}_\theta(\mathbf{u}) \right] \end{aligned}$$

where $I_j = I(U_{ij} \le u_j) - u_j$, $\dot{c}_\theta(\mathbf{u}) = \partial c_\theta(\mathbf{u})/\partial\theta$ and $c_\theta^{(j)}(\mathbf{u}) = \partial c_\theta(\mathbf{u})/\partial u_j$. Then $\mathbb{J}_1, \ldots, \mathbb{J}_n$ are independent and identically distributed (i.i.d.) processes whose form depends on the estimate of θ as well as the hypothesized copula family C_θ. Let $\mathbb{N}$ be a large integer and let $Z_i^{(k)}, i \in \{1, \ldots, n\}, k \in \{1, \ldots, \mathbb{N}\}$, be i.i.d. random variables with mean 0 and variance 1 and be independent of the data $\mathbf{X}$. Under suitable regularity conditions, the GOF process $\mathbb{C}_n, \mathbb{C}_\theta^{(1)}, \ldots, \mathbb{C}_\theta^{(\mathbb{N})}$ defined in (7) converges weakly to $(\{\sum_{i=1}^n \mathbb{J}_i(\mathbf{u})\}/\sqrt{n}$, $\{\sum_{i=1}^n Z_i^{(1)} \mathbb{J}_i(\mathbf{u})\}/\sqrt{n}$, …, $\{\sum_{i=1}^n Z_i^{(\mathbb{N})} \mathbb{J}_i(\mathbf{u})\}/\sqrt{n})$ (refer to [34]).

Let $\hat{\mathbb{J}}_i(\mathbf{u})$ be the estimation version of $\mathbb{J}_i(\mathbf{u})$ in which all the unknown quantities are replaced by their estimates, the approximate p-value for the test based on the multiplier method can be obtained by means of the following procedure (see Kojadinovic and Yan [33] and Kojadinovic et. al. [35], for more details):

***Algorithm** 3.1:* **The multiplier GOF test procedure**

STEP 1: Compute pseudo-observations $\hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_n$ from the observed data $\boldsymbol{X}$ as in Equation (5), then use $\hat{\mathbf{U}}_i$ to get $\mathrm{C}_n(\boldsymbol{u})$ by using Equation (6); Estimate the dependence parameter θ by using Equation (4);

STEP 2: Compute the test statistic S_n defined in Equation (8);

STEP 3: Set N to be a large integer, and repeat the following steps for every $k \in \{1, \ldots, \mathbb{N}\}$:

3.1 Generate n i.i.d random variables $Z_1^{(k)}, \ldots, Z_n^{(k)}$ from the standard normal distribution with mean 0 and variance 1;

3.2 As in [34], form an approximate independent realization of the test statistic under $\mathcal{H}_0$ by

$$\mathbb{C}_n^{(k)}(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i^{(k)} \hat{\mathbb{J}}_i(\mathbf{u});$$

3.3 Compute an approximate independent realization of S_n under $\mathcal{H}_0$ by

$$\begin{aligned} \mathrm{S}_n^{(k)} &= \int_{[0,1]^p} \left\{ \mathbb{C}_n^{(k)}(\mathbf{u}) \right\}^2 d\mathrm{C}_n(\mathbf{u}) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{C}_n^{(k)}(\hat{\mathbf{U}}_i) \right\}^2; \end{aligned}$$

STEP 4: An approximate p-value for the test is given by $\frac{1}{\mathbb{N}} \sum_{k=1}^{\mathbb{N}} \mathbf{1}(\mathrm{S}_n^{(k)} \ge \mathrm{S}_n)$.

Since $\hat{\mathbb{J}}_i(\mathbf{u})$ only needs to be computed once, the multiplier GOF test procedure is faster than the traditional parametric bootstrap method. However, the derivation and the computation of terms $\hat{\mathbb{J}}_i(\mathbf{u})$ are complicated, since they involve partial derivatives of cdf and pdf of the hypothesized copula not only with respect to variables $\mathbf{u}_i$ but also with respect to parameters (see Kojadinovic and Yan [33] and Kojadinovic et. al. [35]). Therefore, we propose a much easier way, i.e., the AIC method, to use choose a correct copula.

B. AIC approach

This section introduces the development and application of AIC (Akaike [2], [3]) in copula selection. AIC derived in Akaike [3] is defined as

$$\begin{aligned} \mathrm{AIC} &= -2(\text{ maximum log likelihood}) \\ &\quad + 2(\text{number of free parameters}) \\ &= -2l(\theta_n) + 2q, \end{aligned} \tag{9}$$

where $l(\theta_n)$ is the maximized value of the log pseudo-likelihood function $l(\theta)$ defined in (4), and q is the number of free parameter.

The AIC is a measure of the relative GOF of a model. We are trying to reduce the distance between the true copula and the approximate copula. Lehmann and Casella [39] suggested using the Kullback-Leibler information as a measure of the distance between the true model and the null hypothesized model. Assume that the pdf of the true unknown copula and the approximate copula model under null hypothesis for a data set are $c(\mathbf{u})$ and $c_\theta(\mathbf{u})$ respectively, then the Kullback-Leibler information measure between $c(\mathbf{u})$ and $c_\theta(\mathbf{u})$ is defined as

$$\begin{aligned} \mathrm{K}\left(c(\mathbf{u}), c_\theta(\mathbf{u})\right) &= \int \left\{ \log \left[\frac{c(\mathbf{u})}{c_\theta(\mathbf{u})} \right] \right\} c(\mathbf{u}) d\mathbf{u} \\ &= \mathrm{E}_{\mathbf{u}} \left\{ \log \left[\frac{c(\mathbf{u})}{c_\theta(\mathbf{u})} \right] \right\} \\ &= \mathrm{E}_{\mathbf{u}} \left[\log c(\mathbf{u}) \right] - \mathrm{E}_{\mathbf{u}} \left[\log c_\theta(\mathbf{u}) \right], \end{aligned} \tag{10}$$

where $\mathrm{E}_{\mathbf{u}}[\cdot]$ denotes the expected value with respect to variable $\mathbf{u}$. The Kullback-Leibler information $\mathrm{K}\left(c(\mathbf{u}), c_\theta(\mathbf{u})\right) \ge$

0, or equivalently, $\mathrm{E}_{\mathbf{u}}[\log c(\mathbf{u})] \geq \mathrm{E}_{\mathbf{u}}[\log c_\theta(\mathbf{u})]$, and it equals to 0 if and only if $c(\mathbf{u}) = c_\theta(\mathbf{u})$ happens almost surely (see Chapter 2 from Kullback [38]). The smaller the Kullback-Leibler information, the closer the approximate copula to the true copula. Alternatively, the larger the quantity $\mathrm{E}_{\mathbf{u}}[\log c_\theta(\mathbf{u})]$, the closer the function $c_\theta(\mathbf{u})$ is to $c(\mathbf{u})$.

Bozdogan [7] suggested that the AIC is an unbiased estimator of $-2\mathrm{E}_{\mathbf{u}}[\log c_\theta(\mathbf{u})]$, viz, AIC $= -2\mathrm{E}_{\mathbf{u}}[\log c_\theta(\mathbf{u})]$. Substituting AIC into Equation (10), the Kullback-Leibler information can be furtherly expressed as

$$\mathrm{K}(c(\mathbf{u}), c_\theta(\mathbf{u})) = \mathrm{E}_{\mathbf{u}}[\log c(\mathbf{u})] + \frac{1}{2}\mathrm{AIC}. \qquad (11)$$

Since AIC definitely is part of Kullback-Leibler information, the value of AIC itself is not meaningful. However, we can minimize AIC to minimize the Kullback-Leibler information. In other words, the smaller the AIC, the closer the function $c_\theta(\mathbf{u})$ is to the true copula pdf $c(\mathbf{u})$. In addition, AIC provides a versatile procedure for statistical model identification and is free from the ambiguities inherent in the application of the conventional hypothesis testing procedures.

When there are several competing copulas, we want to know which copula fits the data best. The chosen copula model should be the one that minimizes the Kullback-Leibler information between the copula model and the true unknown copula. We can calculate the AIC for each model with the same data, then the "best" model is the one with the least AIC value. AIC is more computationally efficient than other copula selection methods. Though it can't do a formal GOF hypothesis test, it can be used to select the best copula from a group of copula families. The practical advantage of AIC in copula analysis will be demonstrated in next section by a simulation study through the comparison between it and the multiplier GOF test method.

IV. Simulation and Results

To assess the performance of the AIC model-selection method, we conduct a simulation study comparing the AIC method with the multiplier GOF test method. The feature of the multiplier GOF test are its ability to maintain its nominal level, which is arbitrarily fixed at 5% throughout the study, and its power under a variety of alternatives. While the feature of AIC is to check the ability of the true copula providing the least AIC, namely, the proportion of the least AIC from the true copula. We simulate $N = 1,000$ times to estimate the level or power of each of the two methods under consideration. In each times, the multiplier GOF test method, which is based on $\mathbb{N} = 1,000$ bootstrap samples, is used to find out whether the copula under the null hypothesis is rejected or not rejected; while the AIC method is used to find out which copula provides the least AIC. After repeating 1,000 times, we will obtain the rejection proportions for the multiplier GOF test as well as the proportions of the least AIC for copula under interest. The procedure for calculating the proportions of the least AIC is given in Algorithm (4.1).

Algorithm 4.1: **AIC procedure**

STEP 1: Set $N = 1000$ and repeat the following steps for every $k \in \{1, \dots, N\}$:

1.1 From a given copula family we generate the random variables $\mathbf{X}^{(k)} = (\mathbf{X}_1^{(k)}, \dots, \mathbf{X}_n^{(k)})$, where n is the number of observations ;

1.2 Compute pseudo-observations $\hat{\mathbf{U}}_1^{(k)}, \dots, \hat{\mathbf{U}}_n^{(k)}$ from the data set $\mathbf{X}^{(k)}$ by using Equation (5), then get the AIC values defined in Equation (9) for all candidate copula families and find out the copula with the least AIC;

STEP 2: For all the candidate families, calculate the proportions (out of N) that each families achieve the least AIC.

In this experimental design of the study, five one-parameter copula families were considered: Clayton, Frank, Gumbel-Hougarrd, Normal, and t with $\nu = 5$. They are abbreviated as C, G, F, N, and t_5, respectively, in the forthcoming tables. Each copula families serve as the true copula or the generating copula; For each copula family, four dependence levels (0.2, 0.4, 0.6 and 0.8) corresponding to Kendall's τ are considered; Three different sample sizes (n=100, 300 and 500) are used; Three dimension sizes (2-variate, 3-variate and 4-variate) are tested.

Therefore, there are $4 \times 3 \times 3 = 36$ scenarios. For each scenario, we will

1. conduct the hypothesis test (namely, $\mathcal{H}_0$ copula (5 choices: Clayton, Frank, Gumbel, Normal, and t with $\nu = 5$) and $\mathcal{H}_1$ copula (5 choices: Clayton, Frank, Gumbel, Normal, and t with $\nu = 5$)) for each of the five copula families (i.e., Clayton, Frank, Gumbel, Normal, and t with $\nu = 5$) by using the multiplier GOF test method and
2. compute the AIC's for each family (i.e, Clayton, Frank, Gumbel, Normal, and t with $\nu = 5$).

In this simulation, one of the interest is to check whether the true copula used to generate the random sample gives the highest proportion of the least AIC. In the following, we will call this proportion the "correct rate" (i.e. the proportion that the true copula gives the least AIC in $N = 1,000$ repetitions). We are also trying to explore the connection between the correct rate from the AIC method and the empirical level (i.e., the proportion of rejections for the true null hypothesis under $N = 1,000$ repetitions) from the multiplier GOF test method. If the AIC method performs well, then the "correct rate" is the highest and close to 1. Conversely, if the multiple GOF test performs well, then the empirical level is small and close to 0.

Tables I, II and III compare the performance of the AIC method with the multiplier GOF test method for sample size $n = 100$, $n = 300$ and $n = 500$, respectively. The AIC method globally agrees with the multiplier GOF test method. In other words, a large correct rate corresponds to a small empirical level. When both the sample size and the dimension are fixed, as the dependence measure τ increases, the correct rate from the AIC method generally increase. Or loosely speaking, the higher the dependence, the more accurate the AIC method is. For example, for the Gumbel copula with $n = 300$ and $d = 2$, the correct rates are 81.6%, 93.0%, 95.1% and 96.4% for $\tau = 0.2$, 0.4, 0.6 and 0.8, respectively.

Moreover, we notice that the smaller the sample size, the larger the correct rate. For example, for Gumbel copula with $\tau = 0.2$ and $d = 2$ the correct rate are 60.1%, 81.6% and 90.3% for $n = 100$, 300 and 500, respectively. Similarly, when both τ and sample size are fixed, as the dimension increases, the correct rate correspondingly increases. For

TABLE I: The proportions of rejection (at 5% significant level) of the null hypothesis and the proportions of the least AIC obtained from $1,000$ repetitions when $d = 2$

n	True	Kendall's	Proportions of the least AIC					Proportions of rejection under $\mathcal{H}_0$				
	Copula	τ	G	C	F	N	t	G	C	F	N	t_5
100	G	0.2	60.1	3.3	9.3	11.5	15.8	3.0	66.5	17.7	9.5	14.2
		0.4	73.6	0.3	4.1	10.4	11.6	2.9	97.5	44.0	22.5	27.0
		0.6	79	0	2.5	8.4	10.1	2.6	100	61.8	29.7	31.7
		0.8	78.3	0	1.7	6.8	13.2	3.1	100	77.4	38.5	38.4
	C	0.2	1.5	75.3	7.3	7.5	8.4	49.3	8.2	19.4	12.6	16.5
		0.4	0	93.9	1.5	2.5	2.1	91.5	8.6	58.0	41.8	47.3
		0.6	0	97.4	0.9	0.8	0.9	98.9	9.5	77.6	69.2	65.4
		0.8	0	98.5	0.9	0.1	0.5	99.8	10.2	57.4	87.0	67.6
	F	0.2	14.5	13.6	44.9	19.0	8.0	18.6	52.8	5.2	9.4	12.9
		0.4	10.4	3.5	59.8	19.0	7.3	41.2	93.7	6.6	20.3	37.3
		0.6	4.8	0.9	77.9	11.3	5.1	62.6	99.5	5.8	40.9	64.1
		0.8	1.5	0	91.4	4.3	2.8	81.4	99.9	5.9	76.2	82.8
	N	0.2	21.3	16.5	22.2	30.7	9.3	12.7	45.2	7.6	4.2	10.4
		0.4	16.0	7.2	11.7	52.2	12.9	20.9	83.3	13.1	3.8	12.4
		0.6	11.3	2.5	7.2	64.3	14.7	18.6	97.3	27.9	4.6	11.9
		0.8	10.1	0.3	3.9	64.3	21.4	9.9	99.7	39.4	4.0	12.1
	t	0.2	14.8	11.1	3.5	5.0	65.6	8.9	33.1	9.4	8.9	3.8
		0.4	16.2	4.6	4.4	10.0	64.8	12.1	79.2	19.3	6.0	3.9
		0.6	14.7	3.2	2.5	13.0	66.6	14.0	96.6	34.3	4.8	3.2
		0.8	11.7	0.9	2.3	14.8	70.3	10.4	99.3	47.1	8.2	4.7
300	G	0.2	81.6	0.2	3.9	8.4	5.9	3.5	96.3	41.6	24.3	33.0
		0.4	93.0	0	0.5	2.7	3.8	4.2	100	81.8	51.2	57.3
		0.6	95.1	0	0.1	1.5	3.3	2.8	100	97.3	64.5	68.4
		0.8	96.4	0	0.1	0.7	2.8	3.0	100	99.7	69.3	68.2
	C	0.2	0	93.8	1.9	2.7	1.6	94.0	5.0	60.0	44.0	58.0
		0.4	0	99.7	0.1	0.1	0.1	100	6.0	99.6	97.4	98.5
		0.6	0	100	0	0	0	100	5.1	100	100	100
		0.8	0	100	0	0	0	100	6.0	100	100	100
	F	0.2	6.2	4.2	68.0	18.6	3.0	55.6	87.4	4.9	17.7	43.2
		0.4	0.5	0	90.5	8.3	0.7	94.5	100	3.4	58.0	90.5
		0.6	0.3	0	98.2	1.4	0.1	99.8	100	4.8	94.1	99.4
		0.8	0	0	100	0	0	100	100	3.2	99.8	100
	N	0.2	1.26	7.9	18.6	58.5	2.4	32.0	70.5	9.1	4.9	25.5
		0.4	4.8	0.8	4.2	87.1	3.1	57.9	99.7	30.0	4.9	36.8
		0.6	1.8	0	1.5	90.4	6.2	62.3	100	66.2	3.7	26.5
		0.8	1.0	0	0.1	92.9	6.0	53.8	100	94.5	3.7	19.4
	t	0.4	5.5	4.2	1.6	1.3	87.4	21.7	61.0	23.8	11.4	4.4
		0.6	4.1	0.2	0.6	1.7	93.4	39.1	99.5	58.8	9.0	4.4
		0.8	2.0	0	0.2	2.3	95.5	50.7	100	83.8	7.7	4.4
		0.2	1.6	0	0.1	3.5	94.8	45.0	100	96.8	8.1	3.7
500	G	0.2	93.0	0	1.5	3.5	2.0	3.2	99.6	59.1	34.6	50.1
		0.4	97.7	0	0.2	1.1	1.0	4.3	100	97.0	75.0	80.6
		0.6	98.8	0	0	0.5	0.7	3.2	100	100	87.8	88.4
		0.8	99.2	0	0	0.2	0.6	2.3	100	100	89.0	89.1
	C	0.2	0	97.9	0.5	1.2	0.4	99.3	5.0	84.7	71.9	87.8
		0.4	0	99.9	0	0.1	0	100	4.2	100	99.9	100
		0.6	0	100	0	0	0	100	5.8	100	100	100
		0.8	0	100	0	0	0	100	5.7	100	100	100
	F	0.2	1.6	1.5	80.6	16.0	0.3	84.6	97.3	3.7	28.9	77.1
		0.4	0	0	97.1	2.9	0	99.9	100	4.1	84.1	99.3
		0.6	0	0	99.8	0.1	0.1	100	100	5.7	99.8	99.9
		0.8	0	0	100	0	0	100	100	2.5	100	100
	N	0.2	6.5	3.6	13.6	75.5	0.6	51.7	88.2	10.2	3.4	45.2
		0.4	1.2	0.2	1.5	96.2	0.9	81.2	100	48.1	4.4	58.7
		0.6	0.1	0	0	98.6	1.3	88.5	100	88.2	4.2	51.1
		0.8	0.1	0	0.2	98.1	1.6	85.4	100	99.9	2.8	30.5
	t	0.2	2.5	0.3	0.2	0.3	96.7	33.5	81.9	46.3	16.2	4.8
		0.4	1.3	0	0	1.4	97.3	66.5	100	85.0	13.0	4.5
		0.6	1.1	0	0	0.4	98.5	75.9	100	98.9	9.6	3.8
		0.8	0	0	0	1.1	98.9	73.6	100	100	8.2	3.3

example, for Clayton copula with $\tau = 0.2$ and $n = 300$ the correct rate are 93.6%, 99.4% and 99.8%.

For the multiplier GOF test method, the empirical levels globally agree with 5% nominal level well. However, not only does the empirical level agrees with 5% nominal level, but also the rejection proportion from other copula family agrees with 5% nominal level. Therefore, it gives more than one copula families which provide the small proportion of rejection for the null hypothesis under consideration. Sometimes, the rejection proportion from other copula instead of the true copula is even less than the empirical level. For example, when the generating copula is Frank, $\tau = 0.2$, $d = 4$, and $n = 300$, the proportions of rejection are 4.9% and 3.2% for Frank and Normal, respectively. The Normal copula provides less proportion of rejection than the Frank copula. But if we use the AIC method, only the Frank copula provides the highest proportion (98.2%) of the least AIC.

As a reference, we will compare the correct rates and the empirical levels. The simulation results are illustrated in Figure 1. Each sub-figure gives the simulation results from the different copula. In each sub-figure, the three columns from left to right give the combination of multivariate dimension and τ for n=100, 300 and 500 respectively. And in each column, the four dots on each line correspond to the simulation results with respect to the association parameter τ (from left to right, $\tau = 0.2, 0.4, 0.6, 0.8$). The three bottom lines give the empirical levels for the multiplier GOF test method, while the three upper lines give the correct rate by using the AIC method.

Figure 1 visually helps us to discern the relationship between the correct rates and sample size, dimension and τ. Obviously, all the correct rates are far more than 50%, which means the AIC method provides only one copula family that fits the data set well. For example, in figure 1 (e), all the correct rates are greater than 50% and approaches to

TABLE II: The proportions of rejection (at 5% significant level) of the null hypothesis and the proportions of the least AIC obtained from $1,000$ repetitions when $d = 3$

n	True Copula	Kendall's τ	Proportions of the least AIC					Proportions of rejection under $\mathcal{H}_0$				
			G	C	F	N	t	G	C	F	N	t_5
100	G	0.2	81.1	0.8	0.7	6.9	4.2	4.9	71.0	21.5	9.1	7.0
		0.4	88.4	0	4.5	2.8	4.3	2.0	98.3	49.1	22.2	17.8
		0.6	91.0	0	2.4	3.1	1.7	1.7	100	63.0	21.7	23.6
		0.8	89.0	0	1.3	2.4	0	0	100	49.3	9.9	10.2
	C	0.2	4.0	87.5	4.5	4.2	3.4	59.9	6.7	34.1	16.9	30.9
		0.4	0	97.8	1.4	0.4	0.4	97.7	6.5	74.3	54.1	62.6
		0.6	0	99.1	0.5	0.2	0.2	99.4	6.0	82.0	72.4	64.5
		0.8	0	99.3	0.6	0	0.1	99.6	2.0	21.9	64.1	14.3
	F	0.2	12.4	7.7	59.4	16.0	4.5	19.6	52.9	7.1	2.7	10.5
		0.4	6.1	1.8	81.0	8.5	2.6	41.9	95.6	4.9	5.1	30.1
		0.6	2.4	0	94.0	3.5	0.1	53.9	100	4.0	23.0	53.0
		0.8	0.2	0	98.3	1.0	0.5	56.4	100	0.6	48.3	47.2
	N	0.2	11.2	10.9	11.2	60.0	6.7	37.6	62.3	27.0	2.7	20.1
		0.4	6.5	3.1	3.2	78.7	8.5	57.9	97.0	39.3	2.2	22.5
		0.6	3.6	0.5	1.8	82.4	11.7	44.3	99.5	54.4	2.0	11.6
		0.8	3.5	0.1	1.0	75.3	20.1	6.0	99.8	34.5	0.5	2.0
	t	0.2	3.8	4.1	1.0	2.1	89.0	30.3	49.0	28.1	7.5	2.5
		0.4	3.5	1.5	0.5	3.7	90.8	43.3	93.0	43.7	4.6	2.9
		0.6	1.9	0.3	0.3	4.7	92.8	42.0	99.6	57.1	2.2	1.0
		0.8	3.3	0.9	0.5	5.8	90.4	11.3	100	38.6	0.9	0
300	G	0.2	96.8	0	2.1	1.0	0.1	2.8	98.5	46.7	36.3	18.4
		0.4	99.7	0	0.2	0.1	0	3.5	100	93.3	68.8	50.9
		0.6	99.5	0	0.2	0.2	0.1	2.6	100	99.5	70.6	62.8
		0.8	99.4	0	0	0.1	0.5	1.1	100	100	59.2	55.6
	C	0.2	0	99.4	0.2	0.3	0.1	98.9	4.9	86.3	7.26	89.8
		0.4	0	100	0	0	0	100	6.3	100	99.5	99.9
		0.6	0	100	0	0	0	100	4.6	100	100	100
		0.8	0	100	0	0	0	100	2.9	100	100	100
	F	0.2	1.6	0.5	92.1	5.7	0.1	58.0	93.1	5.3	7.2	46.8
		0.4	0.1	0	99.5	0.4	0	97.2	100	5.3	42.5	93.7
		0.6	0	0	100	0	0	99.9	100	3.4	91.0	99.7
		0.8	0	0	100	0	0	100	100	2.1	99.0	100
	N	0.2	2.2	2.3	4.7	90.7	0.1	85.3	95.0	46.9	3.4	63.8
		0.4	0.1	0	0.3	99.3	0.3	98.9	100	84.3	3.4	75.0
		0.6	0.1	0	0	99.6	0.3	99.4	100	96.1	3.4	59.6
		0.8	0.1	0	0	98.1	1.8	94.8	100	99.8	2.0	29.0
	t	0.2	0.2	0.2	0	0	99.6	63.2	89.4	51.7	11.6	4.1
		0.4	0.1	0.1	0	0.2	99.6	91.6	100	89.5	8.8	3.8
		0.6	0	0	0	0.2	99.8	95.6	100	99.0	6.1	3.0
		0.8	0	0	0	0.3	99.7	90.5	100	99.9	4.0	1.6
500	G	0.2	99.7	0	0.2	0.1	0	3.5	99.8	73.5	66.3	38.4
		0.4	100	0	0	0	0	2.9	100	99.6	91.8	79.7
		0.6	100	0	0	0	0	4.2	100	100	94.5	90.5
		0.8	100	0	0	0	0	1.8	100	100	87.0	83.7
	C	0.2	0	100	0	0	0.1	100	3.5	98.7	94.2	99.6
		0.4	0	100	0	0	0.1	100	5.2	100	100	100
		0.6	0	100	0	0	0.1	100	4.6	100	100	100
		0.8	0	100	0	0	0.1	100	3.3	100	100	100
	F	0.2	0.4	0	96.9	2.7	0	83.2	99.0	4.5	1.27	75.4
		0.4	0	0	100	0	0	100	100	4.6	71.0	99.8
		0.6	0	0	100	0.1	0	100	100	2.5	99.4	100
		0.8	0	0	100	0	0	100	100	2.1	100	100
	N	0.2	0.6	0.3	1.5	97.6	0	97.6	99.7	58.5	4.3	90.2
		0.4	0	0	0	99.8	0.1	100	100	97.9	3.1	96.3
		0.6	0	0	0	99.9	0.1	100	100	99.9	2.8	88.3
		0.8	0	0	0	100	0	99.7	100	100	2.1	54.5
	t	0.2	0	0	0	0	100	83.5	98.0	76.1	21.2	4.2
		0.4	0	0	0	0	100	99.3	100	99.3	11.8	3.2
		0.6	0	0	0	0	100	99.8	100	100	9.6	3.7
		0.8	0	0	0	0.1	99.9	99.5	100	100	6.4	2.1

1. Certainly, the AIC results agrees with the result from the multiplier GOF test method. From figure 1 (e), it is clear that all the bottom lines are around 5%. Hence, we fail to reject the t copula. Meanwhile, according to the correct rates, the t-copula fits the data much better than any other copula.

In addition to do the comparison of the simulation results from the multiplier GOF test method and the AIC method, the computational aspect also deserves attention, therefore we also need to compare the time spent on doing the multiplier GOF test method and the AIC method. From section III, we notice that the multiplier GOF test method includes two steps, namely, the fitting and the test, while the AIC method only needs the fitting. Since bivariate Normal copula is very common in the real life, here we just compare the run time of the AIC method and the multiplier GOF test method when using the Normal copula as the generating copula and fixing dimension to two. Table IV provides the run times in seconds performed on one 3.20 GHz processor. These are based on the R implementation of the tests available in the copula R package. The last row gives the ratios of the run time spent on the multiplier GOF test method to the run time spent on the AIC method. As one can notice, the use of the AIC method results in a very large computational gain. When the sample size is 500, the run times spent on the multiplier GOF test method is at least 6.14 times as the run times spent on the AIC method. Even when the sample is small, such as 100, the AIC method still same more time than the multiplier GOF test method. Thus, the AIC method provides more computational efficiency than the multiplier GOF test method. For the rest copula families, the ratio results are almost the same patterns.

V. Conclusion

An overview of the AIC method was given, along with the multiplier GOF test method. A large Monte Carlo study was presented, examining the proportion of the least AIC fixed

TABLE III: The proportions of rejection (at 5% significant level) of the null hypothesis and the proportions of the least AIC obtained from $1,000$ repetitions when $d = 4$

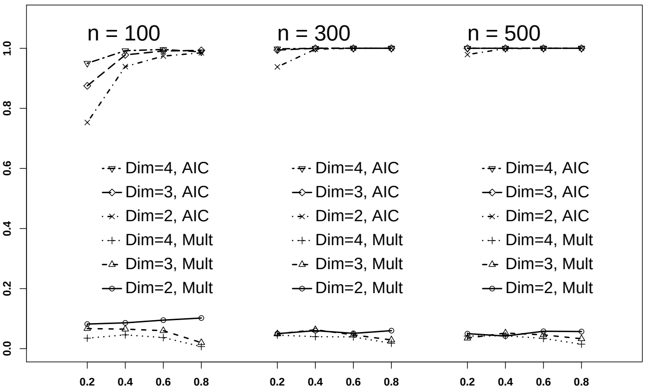
n	True Copula	Kendall's τ	Proportions of the least AIC G	C	F	N	t	Proportions of rejection under $\mathcal{H}_0$ G	C	F	N	t_5
100	G	0.2	88.0	1.0	6.2	3.0	1.8	2.8	70.0	20.7	8.8	1.8
		0.4	94.1	0	4.0	1.4	0.5	2.6	99.1	52.9	24.4	9.3
		0.6	96.0	0	1.2	1.1	1.7	1.2	100	68.9	17.6	15.8
		0.8	93.2	0	1.1	1.1	4.6	0	100	25.8	1.4	1.9
	C	0.2	2.0	95.0	1.6	1.8	1.4	62.8	3.5	36.4	17.6	35.7
		0.4	0	99.2	0.6	0.2	0	97.1	4.6	76.2	56.6	56.4
		0.6	0	99.6	0.3	0.1	0	99.5	3.7	79.1	67.0	48.9
		0.8	0	98.8	1.2	0	0	98.5	0.7	4.2	27.5	0.5
	F	0.2	9.7	5.1	74.7	9.1	14.0	12.7	53.5	6.8	2.7	10.5
		0.4	4.2	0.6	91.5	3.2	0.5	31.9	96.7	5.2	5.1	30.1
		0.6	1.7	0	97.0	1.2	0.1	42.1	99.9	1.8	23.0	53.0
		0.8	0.5	0	98.8	0.4	0.3	0.3	100	0.4	48.3	47.2
	N	0.2	5.9	7.4	6.0	77.8	2.9	59.0	73.8	52.0	1.1	28.0
		0.4	1.4	0.9	1.2	91.4	5.1	80.2	99.3	70.5	1.3	28.7
		0.6	1.8	0.1	0.4	89.5	8.2	61.9	100	71.1	1.2	11.3
		0.8	1.4	0	0.2	78.0	20.4	4.3	100	25.8	0	0.2
	t	0.2	0.5	1.1	0.5	1.1	96.8	45.9	54.0	41.8	5.4	2.1
		0.4	0.8	0.7	0.1	1.3	97.1	68.9	96.7	64.9	3.8	2.2
		0.6	0.8	0.1	0.1	1.5	97.5	63.8	100	73.8	0.6	0.5
		0.8	0.7	0	0	2.1	97.2	9.6	100	28.9	0.3	0
300	G	0.2	99.3	0	0.5	0.2	0	4.6	99.4	54.7	56.3	7.5
		0.4	99.9	0	0	0.1	0	2.9	100	96.4	84.6	31.0
		0.6	100	0	0	0	0	1.8	100	99.6	79.4	55.2
		0.8	100	0	0	0	0	0.6	100	100	53.1	46.8
	C	0.2	0	99.8	0	0.2	0	99.6	4.5	93.5	88.1	96.3
		0.4	0	100	0	0	0	100	4.0	99.9	99.8	100
		0.6	0	100	0	0	0	100	3.9	100	100	100
		0.8	0	100	0	0	0	100	1.8	100	100	99.9
	F	0.2	0.8	0	98.2	1.0	0	52.9	96.0	4.9	3.2	38.6
		0.4	0	0	100	0	0	96.2	100	4.3	21.7	91.0
		0.6	0	0	100	0	0	100	100	3.3	82.0	99.5
		0.8	0	0	100	0	0	100	100	1.9	99.5	99.6
	N	0.2	0.3	0.3	1.3	98.1	0	98.7	99.7	90.2	2.1	86.3
		0.4	0	0	0.3	99.9	0.1	100	100	99.8	3.2	92.4
		0.6	0	0	0	99.6	0.4	100	100	100	2.5	74.2
		0.8	0	0	0	99.3	0.7	99.6	100	100	0.7	27.2
	t	0.2	0	0	0	0	100	90.3	95.5	85.6	13.8	4.2
		0.4	0	0	0	0	100	99.5	100	98.7	7.8	3.4
		0.6	0	0	0	0	100	99.9	100	99.9	6.9	2.9
		0.8	0	0	0	0.2	99.8	98.7	100	100	2.8	0.5
500	G	0.2	99.7	0	0.2	0.1	0	4.0	100	77.9	86.0	15.9
		0.4	100	0	0	0	0	3.8	100	99.8	98.2	59.7
		0.6	100	0	0	0	0	2.3	100	100	97.7	85.6
		0.8	100	0	0	0	0	1.5	100	100	85.6	79.0
	C	0.2	0	100	0	0	0	100	4.2	99.3	98.5	99.9
		0.4	0	100	0	0	0	100	4.3	100	100	100
		0.6	0	100	0	0	0	100	3.4	100	100	100
		0.8	0	100	0	0	0	100	1.5	100	100	100
	F	0.2	0	0	100	0	0	78.5	99.7	5.5	6.6	73.1
		0.4	0	0	100	0	0	100	100	3.2	50.0	99.8
		0.6	0	0	100	0	0	100	100	3.2	98.6	100
		0.8	0	0	100	0	0	100	100	2.1	100	100
	N	0.2	0	0	0.1	99.9	0	100	100	97.6	3.0	99.1
		0.4	0	0	0	100	0	100	100	100	3.3	99.5
		0.6	0	0	0	100	0	100	100	100	3.6	96.8
		0.8	0	0	0	100	0	100	100	100	1.3	69.6
	t	0.2	0	0	0	0	100	98.2	99.8	95.8	21.3	2.6
		0.4	0	0	0	0	100	100	100	100	13.8	3.2
		0.6	0	0	0	0	100	100	100	100	9.0	3.1
		0.8	0	0	0	0	100	100	100	100	4.8	0.8

TABLE IV: The times (seconds) spent on doing the multiplier GOF test method and the AIC method: Normal and $d = 2$

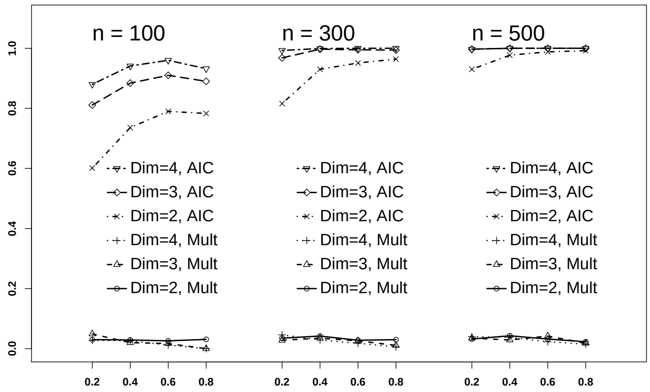
	$n = 100$				$n = 300$				$n = 500$			
τ	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
AIC	354.37	467.54	591.14	675.99	703.61	1039.36	1068.94	1237.30	1332.81	1523.44	1769.86	1814.36
Multiplier	692.6	811.18	921.86	1018.45	2228.83	2543.45	2767.10	2798.98	10504.80	11192.98	10865.67	11389.22
	1.95	1.73	1.56	1.51	3.17	2.45	2.59	2.26	7.88	7.35	6.14	6.28

at the generating copula family under several combinations of problem dimension, sample size and dependence level.

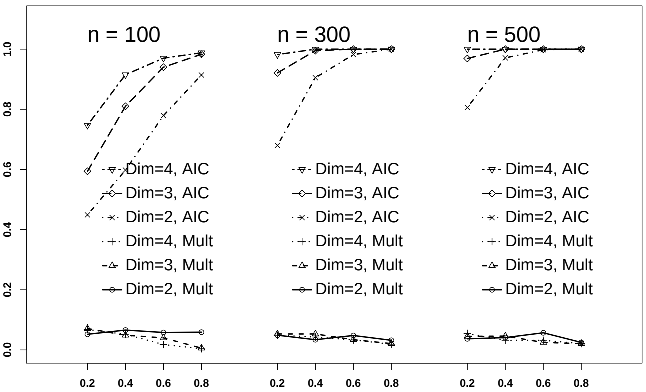
Under the assumption that the true copula exists and is among the series of candidate copula family, we have found that the AIC method achieved better performance by comparing to the multiplier GOF test method. It should also be very easy to use the AIC method to choose the real copula. Furthermore, using the AIC method is faster than using the multiplier GOF test method to choose a copula, especially for large sample sizes. Our study supports the use of AIC to choose the best copula under the certain assumption. However, without the aforementioned assumption the AIC method cannot tell whether the copula with the least AIC is suitable for the particular case, because the AIC is a measure of the relative GOF of a statistical model and the AIC method does not perform a formal GOF hypothesis test. When the true unknown copula in not among the given candidate, the copula with the least AIC will lead to the poor fitting. In this paper, we propose using AIC criteria in choosing copula models. As such, the AIC method is compared against the



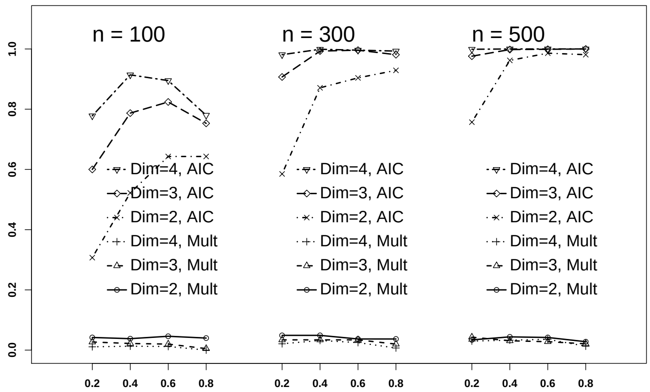
(a) Clayton



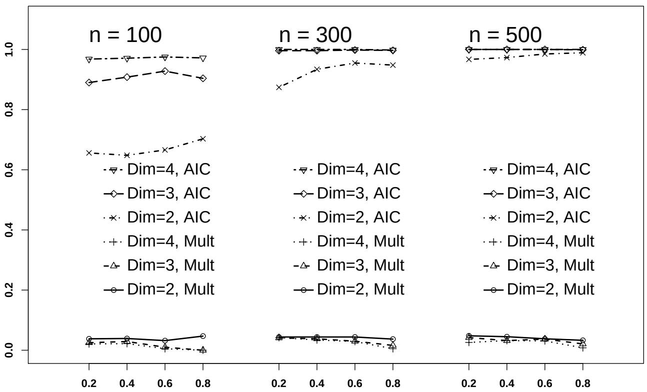
(b) Gumbel



(c) Frank



(d) Normal



(e) t (v=5)

Fig. 1: Plot for both the empirical levels and the correct rates for each copula families with $d = (2, 3, 4)$, n=(100, 300, 500), and $\tau = (0.2, 0.4, 0.6, 0.8)$. The horizontal axis gives the τ levels, and the vertical axis gives the proportions. The upper three lines give us the correct rates from AIC, while the lower three lines give us the empirical levels (at 5% significance level) of the multiplier GOF test method.

popular multiplier GOF test method.

REFERENCES

[1] B. Abdous, C. Genest and B. Rémillard, "Dependence Properties of Meta-elliptical Distributions," *In: Statistical Modelling and Analysis for Complex Data Problems*. Kluwer, Dordrecht, The Netherlands, pp. 1-15, (P. Duchesne & Rémillard, eds.), 2005.

[2] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974.

[3] H. Akaike, "Likelihood of a Model and Information Criteria," *Journal of Econometrics*, vol. 16, pp. 3-14, 1981.

[4] H. Akaike, "Factor Analysis and AIC," *Psychometrika*, vol. 52,. pp. 317-332, 1987.

[5] D. Berg, "Copula Goodness-of-fit Testing: An Overview and Power Comparison," *The European Journal of Finance*, vol. 15, pp. 675–701, 2009.

[6] D., Berg and J. F. Quessy, "Local Sensitivity Analyses of Goodness-of-fit Tests for Copulas," *Scandinavian Journal of Statistics*, vol. 36, pp. 389-412, 2009.

[7] H., Bozdogan, "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, vol. 52, pp. 345-370, 1987.

[8] W. Breymann, A. Dias and P. Embrechts, "Dependence Structures for Multivariate High-frequency Data in Finance," *Quantitative Finance*, vol. 3, pp. 1-14, 2003.

[9] P. Deheuvels, "La fonction de dépendance empirique et ses propriétés Un test non paramétrique d'indépendance." *Académie Royale de Belgique Bulletin de la Classe des Sciences 5e Série* vol. 65, pp. 274-292, 1979.

[10] P. Deheuvels, "An Asymptotic Decomposition for Multivariate Distribution-free Tests of Independence," *Journal Multivariate Analysis*, vol. 11, pp. 102-113, 1981.

[11] P. Deheuvels, "A Non Parametric Test for Independence," *Publication*

of Institute of Statistics of the University of Paris, vol. 26, pp. 29-50, 1981.
[12] J. Dobrić and F. Schmid, "A Goodness of Fit Test for Copulas Based on Rosenblatt's Transformation," *Computational Statistics & Data Analysis*, vol. 51, pp. 4633-4642, 2007.
[13] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1-26, 1979.
[14] B. Efron and R. J. Tibshirani, *An introduction to the Bootstrap*, New York: Chapman and Hall, 1993.
[15] P. Embrechts, A. McNeil and D. Straumann, "Correlation and Dependence in Risk Management: Properties and Pitfalls," *Risk management: value at risk and beyond*, pp. 176-223, 2002.
[16] Y. Fang and L. Madsen, "Modified Gaussian Pseudo-Copula: Applications in Insurance and Finance," *Insurance Mathematics and Economics*, vol. 53, pp. 292-301, 2013.
[17] J. D. Fermanian, "Goodness-of-fit Tests for Copulas," *Journal Multivariate Analysis*, vol. 95, pp. 119-152, 2005.
[18] M. Fréchet, "Sur Les Tableaux De Corrélation Dont Les Marges Son Données," *Annales de lInstitut University Lyon, Section A*, vol. 9, pp. 53-77, 1951.
[19] E. W. Frees and E. A. Valdez, "Understanding Relationships Using Copulas," *North American Actuarial Journal*, vol. 2, pp. 1-25, 1998.
[20] C. Genest and L. P. Rivest, "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, pp. 1034–1043, 1993.
[21] C. Genest, K. Ghoudi and L. P. Rivest, "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, vol. 82, pp. 543-552, 1995.
[22] C. Genest, J. F. Quessy and B. Rémillard, "Goodness-of-fit Procedures for Copula Models Based on the Integral Probability Transformation," *Scandinavian Journal of Statistics*, vol. 33, pp. 337-366, 2006.
[23] C. Genest, A. C. Favre, J. Béliveau and C. Jacques, "Metaelliptical Copulas and Their Use in Frequency Analysis of Multivariate Hydrological Data," *Water Resources Research*, vol. 43, no. 9, 2007.
[24] Genest, C., and Rémillard, B. (2004), "Tests of Independence and Randomness Based on the Empirical Copula Process" *Test* 13:335–369.
[25] C. Genest and B. Rémillard, "Validity of the Parametric Bootstrap for Goodness-of-fit Testing in Semiparametric Models," *Annales de lInstitut Henri Poincaré*, vol. 44, pp. 1096-1127, 2008.
[26] C. Genest,B. Remillard and D. Beaudoin, "Goodness-of-fit Tests for Copulas: A Review and a Power Study ," *Insurance: Mathematics and Economics*, vol. 44, pp. 199-214, 2009.
[27] N. E. Heckman and J. O. Ramsay, "Penalized Regression with Model-based Penalties," *Canadian Journal of Statistics*, vol. 28, no. 2, pp. 241-258, 2000.
[28] W. Hoeffding, "Masstabinvariante Korrelationstheorie," *Schriften Des Matematischen Instituts und des Instituts für Angewandte Matematik de Universität Berlin*, vol. 5, pp. 179-233, 1940.
[29] W. Hoeffding, "Masstabinvariante Korrelationsmasse für diskontinuierliche Verteilunge," *Arkiv fr Matematischen Wirtschaften und Sozialforschung* vol. 7, pp. 49-70, 1941.
[30] M. E. Johnson, *Multivariate Statistical Simulation*, John Wiley & Sons, Inc., 1987.
[31] J. M. Kim, E. A. Sungur, T. Choi and T. Y. Heo, "Generalized Bivariate Copulas and Their Properties," *Model Assisted Statistics and Applications*, vol. 6, no. 2, pp. 127-136, 2011.
[32] S. A. Klugman and R. Parsa "Fitting Bivariate Loss Distributions with Copulas," *Insurance Mathematicsand Economics*, vol. 24, pp. 139-148, 1999.
[33] I. Kojadinovic and J. Yan, "Comparison of Three Semiparametric Methods for Estimating Dependence Parameters in Copula Models," *Insurance: Mathematics and Economics*, vol. 47, no. 1, pp. 52-63, 2010.
[34] I. Kojadinovic and J. Yan, "Modeling Multivariate Distributions with Continuous Margins Using the Copula R Package," *Journal of Statistical Software*, vol. 34, no. 9, pp. 1-20, 2010.
[35] I. Kojadinovic, J. Yan and M. Homes, "Fast Large Sample Goodness-of-fit Tests for Copulas," *Statistica Sinica*, vol. 21, no. 2, pp. 841, 2011.
[36] I. Kojadinovic and J. Yan, "A Goodness-of-fit Test for Multivariate Multiparameter Copulas Based on Multiplier Central Limit Theorems," *Statistics and Computing*, vol. 21, pp. 17-30, 2011.
[37] W. H. Kruskal, "Ordinal Measures of Association," *Journal of the American Statistical Association*, vol. 53, pp. 814-861, 1958.
[38] S. Kullback, *Information Theory and Statistics,* New York : Dover Books on Mathematics. 2nd edition, 1968.
[39] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer: ISBN 0-387-98502-6, 2nd edition, pp. 47, 1998.
[40] R. Mashal and A. Zeevi, *Beyond Correlation: Extreme Co-movements Between Financial Assets*, Working Paper, Columbia University, 2002.
[41] A. J. McNeil and J. Něslehová, "Multivariate Archimedean Copulas, d-Monotone Functions and l_1-norm Symmetric Distributions," *Annals of Statistics*, vol. 37, pp. 3059-3097, 2009.
[42] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed., Springer, New York, 2006.
[43] A. K. Nikoloulopoulos, H. Joe and H. Li. "Vine copulas with asymmetric tail dependence and applications to financial return data." *Computational Statistics & Data Analysis*, vol. 56, no. 11, pp. 3659-3673, 2012.
[44] Y. Otani and J. Imai, "Pricing Portfolio Credit Derivatives with Stochastic Recovery and Systematic Factor." *IAENG International Journal of Applied Mathematics*, vol. 43, no. 4, pp. 176-184, 2013.
[45] R Development Core Team, *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.* Vienna, Austria, ISBN 3–900051–07–0. http://www.R-project.org, 2009
[46] B. Rémillard and O. Scaillet, "Testing for Equality Between Two Copulas," *Journal of Multivariate Analysis*, vol. 100, pp. 377-386, 2009.
[47] S. M. Ross, *Simulation*, 4th edition, Elsevier Inc. 2006.
[48] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461-464, 1978.
[49] J. Shih and T. Louis, "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data," *Biometrics*, vol. 51, pp. 1384-1399, 1995.
[50] M. Sklar, "Fonctions de Répartition à n Dimensions et Leurs Marges," *Publications de l'Institut de Statistique de l'Université de Paris*, vol. 8, pp. 229-231, 1959.
[51] W. Stute, W.G. Manteiga and M. P. Quindimil,"Bootstrap Based Goodness-of-fit Tests," *Metrika*, vol. 40, no. 1, pp. 243-256, 1993.
[52] R. Winkelmann, "Copula bivariate probit models: with an application to medical expenditures." *Health Economics*, vol. 21, no. 12, pp. 1444-1455, 2012.