Computational Statistics and Data Analysis 54 (2010) 3033–3043

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A goodness-of-fit test for Archimedean copula models in the presence of right censoring

Takeshi Emura, Chien-Wei Lin, Weijing Wang *

Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan, ROC

ARTICLE INFO

Article history:
Received 28 November 2008
Received in revised form 15 March 2010
Accepted 16 March 2010
Available online 2 April 2010

Keywords:
Archimedean copulas
Bivariate failure times
Clayton model
Concordance estimator
Goodness of fit test
U-statistics

ABSTRACT

A goodness-of-fit testing procedure for Archimedean copula (AC) models is developed based on right-censored data. The proposed approach extends an existing method, which is suitable for the Clayton model, to general AC models. Asymptotic properties of the proposed test statistics under the true model assumption are derived. Simulation analysis shows that the proposed test has reasonable performance. Finally, two real data examples are analyzed for illustrative purposes.

1. Introduction

Copula models are popular choices for describing multivariate data. They have the major advantage that the dependence structure can be studied separately from marginal distributions. Archimedean copula (AC) models form a useful family of copula models in which the dependence structure can be further characterized by a univariate function (Nelsen, 2006, Section 4). See Frees and Valdez (1998) for insurance applications and Wang and Wells (2000) for biomedical applications. In this article, we consider the bivariate case and propose a testing procedure for checking the goodness-of-fit for the imposed AC model assumption.

If the failure time variables (X, Y) follow a copula model, their joint survival distribution $S(x, y) = \Pr(X > x, Y > y)$ can be written as

$$S(x, y) = C\{S_X(x), S_Y(y)\},$$

where $S_X(x) = S(x, 0)$ and $S_Y(y) = S(0, y)$ are the marginal survival functions, and $C(u, v) : [0, 1]^2 \to [0, 1]$ is the copula function. For AC models, the bivariate copula function can be further simplified as

$$C_\alpha(u, v) = \phi_\alpha^{-1}[\phi_\alpha(u) + \phi_\alpha(v)], \tag{1}$$

where $\phi_\alpha(\cdot) : [0, 1] \to [0, \infty]$ is a univariate function which has two continuous derivatives satisfying $\phi_\alpha(1) = 0$, $\phi_\alpha'(t) = \frac{\partial\phi_\alpha(t)}{\partial t} < 0$ and $\phi_\alpha''(t) = \frac{\partial^2\phi_\alpha(t)}{\partial t^2} > 0$. The parameter α measures the degree of association related to Kendall's τ, which is a rank-invariant correlation measure. Specifically, let (X_i, Y_i) and (X_j, Y_j) be two random replications of (X, Y). Kendal's tau is defined as

$$\tau = \Pr(\Delta_{ij} = 1) - \Pr(\Delta_{ij} = 0),$$

* Corresponding author. Tel.: +886 3 5728746; fax: +886 2 5728745.
E-mail address: wjwang@stat.nctu.edu.tw (W. Wang).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.03.013

where $\Delta_{ij} = I\{(X_i - X_j)(Y_i - Y_j) > 0\}$. It holds that

$$\tau = 4\int_0^\infty \int_0^\infty S(x,y)\mathrm{d}S(x,y) - 1 = 4\int_0^1 \int_0^1 C(u,v)\mathrm{d}C(u,v) - 1 = 4\int_0^1 \frac{\phi_\alpha(t)}{\phi'_\alpha(t)}\mathrm{d}t + 1,$$

where the last identity specifies the relationship between τ and α for an AC model. The following cross ratio function (Oakes, 1989) is useful for describing the local association:

$$\theta^*(x,y) = \frac{S(x,y)\cdot \partial^2 S(x,y)/\partial x\partial y}{\partial S(x,y)/\partial x \cdot \partial S(x,y)/\partial y}.$$

For AC models, Oakes (1989) showed that $\theta^*(x,y)$ can be written as

$$\theta^*(x,y) = \theta_\alpha\{S(x,y)\}, \tag{2}$$

where $\theta_\alpha(v) = -v\phi''(v)/\phi'(v)$ is a univariate function.

Semiparametric inference of the association parameter α for a chosen AC model has been a popular research topic in the literature. The form of $\phi_\alpha(\cdot)$ is assumed but the marginal distributions are not specified. Therefore how to choose an appropriate $\phi_\alpha(\cdot)$ based on the data at hand is an important issue for practical applications. Oakes (1989) and Genest and Rivest (1993) derived analytic properties of AC models that are useful for model selection. Specifically, Oakes (1989) showed that

$$E[\Delta_{ij} \mid \tilde{X}_{ij} = x, \tilde{Y}_{ij} = y] = \frac{\theta_\alpha\{S(x,y)\}}{\theta_\alpha\{S(x,y)\}+1}, \tag{3}$$

where $\tilde{X}_{ij} = X_i \wedge X_j$ and $\tilde{Y}_{ij} = Y_i \wedge Y_j$, where $a\wedge b = \min(a,b)$. Genest and Rivest (1993) derived that the variable $V = S(X,Y)$ has the distribution function

$$K_\alpha(v) = \Pr(V \le v) = v - \lambda_\alpha(v), \tag{4}$$

where $\lambda_\alpha(v) = \phi_\alpha(v)/\phi'_\alpha(v)$ for $0 < v \le 1$.

Eq. (4) has been adopted for model selection by Genest and Rivest (1993) for complete data and by Wang and Wells (2000) for bivariate censored data. The idea is to measure the goodness-of-fit based on the distance between nonparametric and model-based estimators of $K(v) = \Pr(V \le v)$ which characterizes the model. However, the distribution of the distance statistics (i.e., the supremum norm or L_2 form of a function) is quite complicated. Alternatively, some authors suggested approximating the distribution of a test statistics under the null hypothesis based on re-sampling techniques such as parametric bootstrap procedures (Dobric and Schmid, 2007; Nikoloulopoulos and Karlis, 2008; Genest et al., 2009). However, the computational cost of re-sampling methods may restrict their applicability.

A different strategy, originally proposed by Gill and Schumacher (1987), is to compare the discrepancy between two point estimators which are derived under the same class of estimating functions but with different weight functions. If the imposed model assumption is correctly specified, the two estimators will be close to each other. On the other hand, when the model assumption is wrong, the values of the two estimators will fall apart if the weights are properly chosen. Formal goodness-of-fit tests can be constructed since asymptotic properties are easier to handle. Shih (1998) adopted this idea to check the assumption of the Clayton model with $\phi_\alpha(t) = (t^{-\alpha}-1)/\alpha$. In this article we extend this approach to general AC models based on Eq. (3). The proposed goodness-of-fit statistics are introduced in Section 2. Section 3 contains simulation results and Section 4 contains real data applications. Concluding remarks are given in Section 5.

2. The proposed test

We consider testing

$$H_0 : C(u,v) = \phi_\alpha^{-1}[\phi_\alpha(u) + \phi_\alpha(v)] \quad \text{for some } \alpha \in R.$$

The alternative hypothesis can be any copula other than the one specified as H_0. We will examine the power performance when the alternative hypothesis follows a different AC model or a non-AC model. We will first ignore censoring to illustrate the main idea. In Section 2.5, the proposed method will be modified to account for the censoring effect.

2.1. The main idea

In the absence of censoring, the observed data can be written as $\{(X_i, Y_i);\ (i = 1, \ldots, n)\}$, and the concordance indicator Δ_{ij} is observable. A class of estimating equations based on Eq. (3) can be constructed as

$$U_k(\alpha) = \sum_{i<j} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha)\left[\Delta_{ij} - \frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}+1}\right], \tag{5}$$

where $W_k(\cdot,\cdot,\cdot)$ is a weight function and $\hat{S}(x,y)$ is an estimator of $S(x,y)$. For complete data, we can use the empirical survival function $\hat{S}(x,y) = \frac{1}{n}\sum_i I(X_i \geq x, Y_i \geq y)$. To construct a goodness-of-fit test, we need to choose two weight functions, and then obtain $\hat{\alpha}_k$ which solves $U_k(\alpha) = 0$ for $k = 1, 2$. Under H_0 and some regularity conditions, $n^{1/2}\{\hat{\alpha}_1 - \hat{\alpha}_2\}$ converges to a mean-zero normal distribution. The power of the test depends on the choice of two weights. In principle, one weight function should utilize more model information while the other should be less model-dependent. It is expected that $n^{1/2}\{\hat{\alpha}_1 - \hat{\alpha}_2\}$ will show a large discrepancy when H_0 does not hold. We propose to choose one weight function using some likelihood information, which will be discussed in Section 2.2. For the other weight function, we suggest the unweighted version since it does not contain any model information. Sometimes, for better normal approximation, one may use $n^{1/2}\{\hat{\gamma}_1 - \hat{\gamma}_2\}$, where γ is a monotone function of α.

2.2. Conditional likelihood and the weight function

We generalize the likelihood approach proposed by Clayton (1978) and show that the resulting estimating function can also be written as the form of $U_k(\alpha)$ in (5). Define the set of grid points,

$$\psi = \left\{ (x,y) \mid \sum_{i=1}^{n} I(X_i = x, Y_i \geq y) = 1, \sum_{i=1}^{n} I(X_i \geq x, Y_i = y) = 1 \right\}.$$

Define $D(x,y) = \sum_{i=1}^{n} I(X_i = x, Y_i = y)$, which measures the number of observed failures at $(x,y) \in \psi$, and $R(x,y) = \sum_{i=1}^{n} I(X_i \geq x, Y_i \geq y)$, which measures the number at risk at $(x,y) \in \psi$. Assuming that the data contains no ties and conditional on $R(x,y)$ for $(x,y) \in \psi$, $D(x,y)$ is a Bernoulli random variable with the success probability $\Pr\{D(x,y) = 1 \mid R(x,y) = r\}$. Clayton (1978) derived this probability under the Clayton model with $\phi_\alpha(v) = (v^{-\alpha} - 1)/\alpha$. For an AC model, we show that

$$\Pr\{D(x,y) = 1 \mid R(x,y) = r\} = \frac{\theta_\alpha\{S(x,y)\}}{r - 1 + \theta_\alpha\{S(x,y)\}}.$$

Combining all the points in ψ under the working assumption of independence among different grids, the likelihood function can be written as

$$L(\alpha) = \prod_{(x,y)\in\psi} \left[\frac{\theta_\alpha\{S(x,y)\}}{R(x,y) - 1 + \theta_\alpha\{S(x,y)\}} \right]^{D(x,y)} \left[\frac{R(x,y) - 1}{R(x,y) - 1 + \theta_\alpha\{S(x,y)\}} \right]^{1-D(x,y)}. \tag{6}$$

Motivated by the paper of Oakes (1986), we apply some algebraic operations (given in Appendix A) to show that the resulting log-likelihood can be written as

$$l(\alpha) = \sum_i \log\left[\frac{\theta_\alpha\{S(X_i, Y_i)\}}{R_{ii} - 1 + \theta_\alpha\{S(X_i, Y_i)\}} \right] + \sum_{i<j} (1 - \Delta_{ij}) \log\left[\frac{R_{ij} - 1}{R_{ij} - 1 + \theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}} \right], \tag{7}$$

where $R_{ij} = R(\tilde{X}_{ij}, \tilde{Y}_{ij}) = n\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})$. The resulting score function becomes

$$\begin{aligned} U_1(\alpha) &= \sum_{i<j} \frac{\dot{\theta}_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}[\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1]}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}[R_{ij} - 1 + \theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}]} \left[\Delta_{ij} - \frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right] \\ &= \sum_{i<j} W_1(\tilde{X}_{ij}, \tilde{Y}_{ij}) \left[\Delta_{ij} - \frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right], \end{aligned} \tag{8a}$$

where $\dot{\theta}_\alpha(v) = \partial\theta_\alpha(v)/\partial\alpha$. The second estimating function is given by

$$U_2(\alpha) = \sum_{i<j} \left[\Delta_{ij} - \frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right]. \tag{8b}$$

By solving $U_k(\alpha) = 0$, we have $\hat{\alpha}_k$ for $k = 1, 2$.

Though we have assumed the absence of ties in the derivation, the resulting Eqs. (8a) and (8b) can still be calculated for tied data without any tie breaking method. Note that the validity of (8a) and (8b) is due to the moment equation (3) that is derived under the continuity assumption of $S(x,y)$. If the underlying distribution is continuous but ties occur due to random round-off of the observation, the resulting bias in (8a) and (8b) should be modest. On the other hand, if ties occur due to the discrete mass in the distribution, the proposed method may not be appropriate.

For the Clayton model with $\phi_\alpha(t) = (t^{-\alpha} - 1)/\alpha$, $\theta_\alpha\{S(x, y)\} = \alpha + 1$, which is a constant. In this special case, Eqs. (8a) and (8b) reduce to the results in Shih (1998). For the Gumbel model with $\phi_\alpha(v) = \{-\log(v)\}^{\alpha+1}$, we have $\theta_\alpha\{S(x, y)\} = 1 - \alpha/\{\log S(x, y)\}$ and

$$W_1(\tilde{X}_{ij}, \tilde{Y}_{ij}) = \frac{2 \log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) - \alpha}{\{\log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) - \alpha\}\{\alpha - R_{ij} \log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}.$$

Therefore,

$$U_1(\alpha) = \sum_{i<j} \frac{2 \log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) - \alpha}{\{\log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) - \alpha\}\{\alpha - R_{ij} \log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}} \left[\Delta_{ij} - \frac{\log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) - \alpha}{2 \log \hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij}) - \alpha} \right]. \tag{9}$$

We will use the Gumbel model for illustration in subsequent discussions.

2.3. Asymptotic distributional theory

In this section, we derive the asymptotic distribution of $n^{1/2}(\hat{\alpha}_1 - \hat{\alpha}_2)$ and also $n^{1/2}(\hat{\gamma}_1 - \hat{\gamma}_2)$ if the latter provides a better approximation to the normal distribution. The theoretical challenge comes from the fact that $U_1(\alpha)$ and $U_2(\alpha)$ involve the estimator $\hat{S}(x, y)$ (see Eq. (9) for the Gumbel case). Lemma 1 states that $U_1(\alpha)$ and $U_2(\alpha)$ can be approximated by the following two U-statistics:

$$\tilde{U}_1(\alpha) = \sum_{i<j} \frac{\dot{\theta}_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}[\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1]}{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}S(\tilde{X}_{ij}, \tilde{Y}_{ij})} \left[\Delta_{ij} - \frac{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right], \tag{10a}$$

and

$$\tilde{U}_2(\alpha) = \sum_{i<j} \left[\Delta_{ij} - \frac{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right]. \tag{10b}$$

Lemma 1. *Under the correct model and regularity conditions in Appendix B.1,*

$$n \binom{n}{2}^{-1} U_1(\alpha) = \binom{n}{2}^{-1} \tilde{U}_1(\alpha) + o_P(1), \qquad \binom{n}{2}^{-1} U_2(\alpha) = \binom{n}{2}^{-1} \tilde{U}_2(\alpha) + o_P(1),$$

where $o_P(1)$ is uniform in α.

If the parameter α has a positive value as in the Gumbel model, the natural logarithm transformation can improve the normal approximation. The following two results are derived by taking $\gamma = \log \alpha$ and $\hat{\gamma}_k = \log \hat{\alpha}_k$ for $k = 1, 2$.

Lemma 2. *Under the correct model and the regularity conditions in Appendix B.1,*

$$n^{1/2}(\hat{\gamma}_1 - \hat{\gamma}_2) = n^{1/2} \binom{n}{2}^{-1} \sum_{i<j} h\{(X_i, Y_i), (X_j, Y_j)\} + o_P(1),$$

where the function h is symmetric in its arguments and is defined as

$$h\{(X_i, Y_i), (X_j, Y_j)\} \equiv \frac{1}{\alpha} \left(\frac{\dot{\theta}_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}[\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1]}{A_L \theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}S(\tilde{X}_{ij}, \tilde{Y}_{ij})} - \frac{1}{A} \right) \left[\Delta_{ij} - \frac{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1} \right],$$

$$A \equiv E\left(\frac{\dot{\theta}_\alpha\{S(\tilde{X}_{12}, \tilde{Y}_{12})\}}{[\theta_\alpha\{S(\tilde{X}_{12}, \tilde{Y}_{12})\} + 1]^2} \right) \quad \text{and} \quad A_L \equiv E\left(\frac{[\dot{\theta}_\alpha\{S(\tilde{X}_{12}, \tilde{Y}_{12})\}]^2}{\theta_\alpha\{S(\tilde{X}_{12}, \tilde{Y}_{12})\}[\theta_\alpha\{S(\tilde{X}_{12}, \tilde{Y}_{12})\} + 1]} \right).$$

Based on Lemma 2 and applying the central limit theorem for U-statistics (p. 162, Theorem 12.3 of Van Der Vaart, 1998), we can obtain the following theorem.

Theorem 1. *Under the correct model and the regularity conditions in Appendix B.1, $n^{1/2}(\hat{\gamma}_1 - \hat{\gamma}_2)$ converges in distribution to a normal distribution with mean zero and variance $\sigma^2 = 4E[h\{(X_1, Y_1), (X_2, Y_2)\}h\{(X_1, Y_1), (X_3, Y_3)\}]$.*

2.4. The proposed testing procedure

For testing H_0, we propose to reject H_0 at the significance level κ if $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}$ is greater than the upper $\kappa/2$ percentile of the standard normal distribution, where $\hat{\sigma}$ is an estimator of $\sigma/\sqrt{n}$.

The asymptotic variance σ^2 can be estimated based on its analytical expression by averaging over all possible observation triples. However this analytic formula becomes complicated for right-censored data. Alternatively, we can use the jackknife estimator defined as

$$\hat{\sigma}^2_{\text{Jack}} = \frac{n-1}{n}\sum_{i=1}^{n}\{\hat{\gamma}_1^{(-i)} - \hat{\gamma}_2^{(-i)} - (\hat{\gamma}_1^{(\cdot)} - \hat{\gamma}_2^{(\cdot)})\}^2,$$

where $\hat{\gamma}_k^{(-i)}$ is the statistic $\hat{\gamma}_k$ ignoring the i-th observation and $\hat{\gamma}_1^{(\cdot)} - \hat{\gamma}_2^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\gamma}_1^{(-i)} - \hat{\gamma}_2^{(-i)})$.

2.5. Modification for right censoring

When (X_i, Y_i) are subject to right censoring by (A_i, B_i) such that one can only observe $(\breve{X}_i, \breve{Y}_i, \delta_i^X, \delta_i^Y)$ $(i = 1, \dots, n)$, where $\breve{X}_i = X_i \wedge A_i$, $\breve{Y}_i = Y_i \wedge B_i$, $\delta_i^X = I(X_i \le A_i)$ and $\delta_i^Y = I(Y_i \le B_i)$. We assume that (A_i, B_i) is independent of (X_i, Y_i). The order of X_i and X_j is known if and only if $\tilde{X}_{ij} \le \tilde{A}_{ij}$, where $\tilde{X}_{ij} = X_i \wedge X_j$ and $\tilde{A}_{ij} = A_i \wedge A_j$. Similarly, the order of Y_i and Y_j can be known if and only if $\tilde{Y}_{ij} \le \tilde{B}_{ij}$. Define $Z_{ij} = I(\tilde{X}_{ij} \le \tilde{A}_{ij}, \tilde{Y}_{ij} \le \tilde{B}_{ij})$, which indicates whether the ordering relationship is certain or not. We can modify $U_k(\alpha)$ in (5) by selecting only "orderable" pairs with $Z_{ij} = 1$ such that

$$U_k(\alpha) = \sum_{i<j} Z_{ij} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, \hat{S})\left[\Delta_{ij} - \frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1}\right] \quad (k = 1, 2), \tag{11a}$$

where $\Delta_{ij} = I\{(\breve{X}_i - \breve{X}_j)(\breve{Y}_i - \breve{Y}_j) > 0\}$ and $\hat{S}(x, y)$ is an estimator of $S(x, y)$ suitable for right-censored data. Note that, to simplify the presentation, we may use the same notations but change their definitions for the censoring case. The first suggested weight function has the form

$$W_1(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, S) = \frac{\dot{\theta}_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}[\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1]}{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}[R_{ij} - 1 + \theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}]},$$

where $R_{ij} = \sum_{l=1}^{n} I(\breve{X}_l \ge \tilde{X}_{ij}, \breve{Y}_l \ge \tilde{Y}_{ij})$ and the second one is $W_2(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, S) = 1$.

A possible candidate of $\hat{S}(x, y)$ is the nonparametric estimator proposed by Dabrowska (1988). Since the estimator is somewhat complicated, one may use the following model-based estimator:

$$\tilde{S}_\alpha(x, y) = \phi_\alpha^{-1}\{\phi_\alpha[\hat{S}_X(x)] + \phi_\alpha[\hat{S}_Y(y)]\},$$

where $\hat{S}_X(x)$ and $\hat{S}_Y(y)$ are marginal Kaplan–Meier estimators of $S_X(x)$ and $S_Y(y)$, respectively. To obtain $\hat{\alpha}_k$, one can solve the equation

$$\sum_{i<j} Z_{ij} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}, \alpha, \tilde{S}^{(l-1)})\left[\Delta_{ij} - \frac{\theta_\alpha\{\tilde{S}^{(l-1)}(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{\tilde{S}^{(l-1)}(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1}\right] = 0 \tag{11b}$$

successively for $l = 1, 2, \dots$, until some convergence criterion is met, where $\tilde{S}^{(l)}$ is defined as $\tilde{S}_\alpha(x, y)$, with α being the estimated value in the l-step. The initial value α_0 may be obtained by inverting a naïve estimate of Kendall's tau based on $\{(\breve{X}_i, \breve{Y}_i) : (i = 1, \dots, n)\}$. Our simulation analysis shows that the two methods for handling the survival function yield close results.

The asymptotic normality results can be established using similar arguments as in the uncensored case. The key step is to approximate $U_k(\alpha)$ in (11a) by a U-statistic. For instance,

$$\tilde{U}_2(\alpha) = \sum_{i<j} Z_{ij}\left[\Delta_{ij} - \frac{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\}}{\theta_\alpha\{S(\tilde{X}_{ij}, \tilde{Y}_{ij})\} + 1}\right],$$

is a U-statistic that can approximate $U_2(\alpha)$. Under regularity conditions such as those listed in Proposition 4.1 of Dabrowska (1988), $\hat{S}(x, y)$ (or $\tilde{S}_\alpha(x, y)$) converges in probability to $S(x, y)$ uniformly in $(x, y) \in [0, \tau_1) \times [0, \tau_2)$ for some (τ_1, τ_2). Then, it follows from the same arguments as the proof of Lemma 1 that

$$n\binom{n}{2}^{-1} U_2(\alpha) = \binom{n}{2}^{-1}\tilde{U}_2(\alpha) + o_P(1).$$

Applying a Taylor series expansion and the central limit theorem for U-statistics, as in the proof of Lemma 1, $n^{1/2}(\log\hat{\alpha}_1 - \log\hat{\alpha}_2)$ asymptotically follows a mean-zero normal distribution.

Table 1A
Empirical probabilities of not rejecting the Gumbel model at the 5% significance level.

	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$
$n = 100$					
CEN% = (0, 0)	0.99	0.97	0.96	0.96	0.95
CEN% = (20, 20)	0.98	0.93	0.95	0.97	0.96
CEN% = (50, 50)	0.97	0.96	0.95	0.96	0.99
$n = 200$					
CEN% = (0, 0)	0.97	0.95	0.92	0.93	0.93
CEN% = (20, 20)	0.95	0.94	0.94	0.93	0.96
CEN% = (50, 50)	0.95	0.97	0.98	0.97	0.96

Note: CEN% denotes the two marginal censoring rates, $100 \times \Pr(A < X)$ and $100 \times \Pr(B < Y)$. The probabilities are calculated based on 100 replications.

Table 1B
Means and standard deviations (in parentheses) of the test statistic $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}_{\text{Jack}}$ under H_0.

	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$
$n = 100$					
CEN% = (0, 0)	0.06 (0.89)	0.18 (0.90)	0.28 (1.00)	0.12 (0.89)	0.32 (0.91)
CEN% = (20, 20)	−0.01 (0.94)	0.30 (1.00)	0.13 (1.04)	0.19 (1.03)	0.14 (0.89)
CEN% = (50, 50)	0.15 (0.99)	−0.13 (0.94)	0.14 (0.98)	0.17 (0.81)	0.06 (0.81)
$n = 200$					
CEN% = (0, 0)	0.07 (0.94)	0.13 (1.01)	0.22 (0.94)	0.05 (0.94)	0.06 (0.94)
CEN% = (20, 20)	0.12 (0.99)	0.19 (1.02)	0.10 (1.02)	0.05 (0.89)	0.07 (0.97)
CEN% = (50, 50)	−0.01 (0.83)	0.03 (0.92)	0.12 (0.84)	0.13 (0.97)	0.04 (0.90)

Note: CEN% denotes the two marginal censoring rates, $100 \times \Pr(A < X)$ and $100 \times \Pr(B < Y)$. The means and standard deviations are calculated based on 100 replications.

3. Simulation results

The simulation study contains three components. We first examine the performance of the proposed test when H_0 is correctly specified. Then, we study the robustness of the proposed test under dependent censoring, a condition which violates the assumption of independent censorship. Finally, we study the power of the test when H_0 does not hold. The level of significance is set to be 5%.

Here we use the Gumbel model for illustration. The null hypothesis is given by

$$H_0 : C(u, v) = \exp\left\{-\left[(-\log u)^{\alpha+1} + (-\log v)^{\alpha+1}\right]^{\frac{1}{\alpha+1}}\right\} \quad \text{for some } \alpha > 0.$$

The two marginal distributions both follow an exponential distribution with the hazard rate equal to 1. Specifically, we let

$$X = -U^{1/(\alpha+1)} \log(V), \qquad Y = -(1-U)^{1/(\alpha+1)} \log(V),$$

where $U \sim \text{Uniform}(0, 1)$ and V has the distribution function $K_\alpha(v) = v - v\log(v)/(\alpha + 1)$. Five values of α whose corresponding τ is equal to 0.3, 0.4, 0.5, 0.6 and 0.7 are chosen. The bivariate censoring variables (A, B) are generated independently from exponential distributions such that the censoring proportion is 0, 0.2 and 0.5 respectively in each coordinate. Based on the data, we obtain $\hat{\alpha}_k$ which solves Eq. (11b), and then $\hat{\gamma}_k = \log \hat{\alpha}_k$ for $k = 1, 2$. The jackknife estimator $\hat{\sigma}^2_{\text{Jack}}$, illustrated in Section 2.4, is used for variance estimation. The null hypothesis is rejected if $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}_{\text{Jack}}$ exceeds 1.96.

Table 1A reports the empirical probabilities of not rejecting H_0 based on 100 replications. When $n = 100$, the type-I error rates are slightly lower than the nominal 5% level in some cases but the results improve as the sample size increases to $n = 200$. Table 1B reports the means and standard deviations of $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}_{\text{Jack}}$ which can be used to evaluate the validity of normal approximation. We see that, when n increases, the approximation also improves. In general, the means are close to zero and the standard deviations are slightly less than one. This may be due to the fact that the jackknife algorithm tends to overestimate the variance, which results in slightly lower type-I error rate.

We also evaluate the robustness of the proposed test under dependent censoring in which (A, B) and (X, Y) are no longer independent. Let $A = C_1 I(X < -\log(0.5)) + C_2 I(X \geq -\log(0.5))$, where C_1 and C_2 are independent exponential distributions with hazard rates $\theta/2$ and $1/2$, respectively. It is easy to see that X and A are positively correlated when $1 < \theta$ and negatively correlated when $0 < \theta < 1$. We define B in the same way.

Table 2A summarizes the empirical probabilities of not rejecting H_0 based on 100 replications under the dependent censoring setting. Surprisingly, the empirical type-I error rates are still close to the nominal 5% level. In fact, Table 2B shows that $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}_{\text{Jack}}$ can still be approximated by the standard normal distribution. However, we find that both $\hat{\gamma}_1$ and $\hat{\gamma}_2$ become biased under dependent censoring but the mean of $\hat{\gamma}_1 - \hat{\gamma}_2$ is still close to zero. This implies that the dependent

Table 2A
Empirical probabilities of not rejecting the Gumbel model at the 5% significance level under dependent censoring.

		$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$
$n = 100$						
$\theta = 1/4$	CEN% = (28, 28)	0.98	0.96	0.96	0.95	0.96
$\theta = 1/2$	CEN% = (30, 30)	0.98	0.97	0.95	0.96	0.95
$\theta = 2$	CEN% = (39, 39)	0.97	0.98	0.97	0.99	0.98
$\theta = 4$	CEN% = (47, 47)	0.99	0.96	0.94	0.99	0.97
$n = 200$						
$\theta = 1/4$	CEN% = (28, 28)	0.93	0.95	0.93	0.93	0.94
$\theta = 1/2$	CEN% = (30, 30)	0.95	0.95	0.96	0.95	0.97
$\theta = 2$	CEN% = (39, 39)	0.91	0.94	0.93	0.97	0.97
$\theta = 4$	CEN% = (47, 47)	0.97	0.97	0.97	0.97	0.96

Note: θ is related to the degree of dependence between failure times and censoring times. CEN% denotes the two marginal censoring rates, $100 \times \Pr(A < X)$ and $100 \times \Pr(B < Y)$. The probabilities are calculated based on 100 replications.

Table 2B
Means and standard deviations (in parentheses) of the test statistics $|\hat{\gamma}_1 - \hat{\gamma}_2|/\hat{\sigma}_{\text{Jack}}$ under H_0 and dependent censoring.

		$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$
$n = 100$						
$\theta = 1/4$	CEN% = (28, 28)	0.11 (0.84)	0.01 (0.96)	0.16 (0.91)	0.17 (0.95)	0.17 (0.90)
$\theta = 1/2$	CEN% = (30, 30)	0.21 (0.80)	−0.01 (0.98)	0.23 (0.90)	0.12 (0.85)	0.19 (0.99)
$\theta = 2$	CEN% = (39, 39)	0.08 (0.99)	0.13 (0.77)	0.06 (0.84)	−0.10 (0.82)	−0.01 (0.82)
$\theta = 4$	CEN% = (47, 47)	0.08 (0.85)	0.01 (1.03)	−0.31 (0.98)	−0.14 (0.79)	−0.06 (0.86)
$n = 200$						
$\theta = 1/4$	CEN% = (28, 28)	0.06 (1.15)	0.12 (0.99)	0.26 (1.05)	0.28 (1.01)	0.27 (0.95)
$\theta = 1/2$	CEN% = (30, 30)	0.06 (0.98)	0.15 (0.94)	0.20 (0.97)	−0.05 (0.91)	0.17 (0.81)
$\theta = 2$	CEN% = (39, 39)	0.13 (1.01)	0.12 (1.00)	0.14 (1.06)	−0.04 (0.92)	−0.10 (0.95)
$\theta = 4$	CEN% = (47, 47)	0.03 (0.90)	0.01 (0.94)	−0.01 (0.93)	−0.28 (0.85)	−0.24 (0.87)

Note: θ is related to the degree of dependence between failure times and censoring times. τ denotes the Kendall's tau between X and Y. CEN% denotes the two marginal censoring rates, $100 \times \Pr(A < X)$ and $100 \times \Pr(B < Y)$. The means and standard deviations are calculated based on 100 replications.

censoring affects the biasness of the two estimators in the same direction and with similar magnitude. We also investigated different setups of dependent censoring, and the results were similar. More thorough analysis for the robustness under dependent censoring deserves future study.

To assess the power of the test, data is used from the Clayton and Frank models, respectively. The power is the probability of correctly rejecting Gumbel's assumption. The empirical power functions are plotted in Figs. 1A for $n = 100$ and Fig. 1B for $n = 200$, respectively. In each figure, we see that the power deceases as τ decreases. This is reasonable since these three models coincide with each other as Kendall's τ approaches zero, and it becomes more difficult to distinguish between similar models. As the level of censoring increases, the power performance becomes worse. As expected, the powers become larger when the sample size increases.

We further examine the power when the alternative model is similar to that specified in the null hypothesis. Here we consider the Galambos copula model, which is not an AC model but is similar to the Gumbel AC model in some key aspects. For example, both copulas have upper tail dependences. In terms of the Jeffreys divergence measure, the distance between the two models is only 0.001 even when $\tau \geq 0.5$ (Nikoloulopoulos and Karlis, 2008). We report the results under $\tau = 0.5$. When the censoring proportion is 50%, the power is only 4% for $n = 100$ and 11% for $n = 200$. Nevertheless, the power increases to 23% for $n = 100$ and 40% for $n = 200$ when the censoring proportion reduces to 20%. In the absence of censoring, the power further increases to 48% for $n = 100$ and 79% for $n = 200$.

It is worth noting that the method in Shih (1998) was developed to verify the Clayton model assumption. In her simulations, the Clayton assumption was set as the null hypothesis and the power under the Gumbel alternative was evaluated. Comparing Shih's analysis with the left panels of Figs. 1A and 1B, the two results are close, which implies that reversing the roles of the null and alternative hypotheses does not matter.

4. Real data applications

The first example is from the Australian Twin Study (Duffy et al., 1990), in which (X, Y) represents the ages at appendicectomy measured for each twin pair. As in Prentice and Hsu (1997), a subset of the original data containing 748 dizygotic pairs is chosen. In the sample, 82 observations are uncensored, 117 are censored for X, 105 are censored for Y and 444 are censored for both X and Y. The proposed method is applied to test the goodness-of-fit for four AC model candidates individually, and the results are summarized in Table 3A. The Gumbel model provides the best fit to the data while the Clayton and Log-copula models are rejected at the 5% significance level.

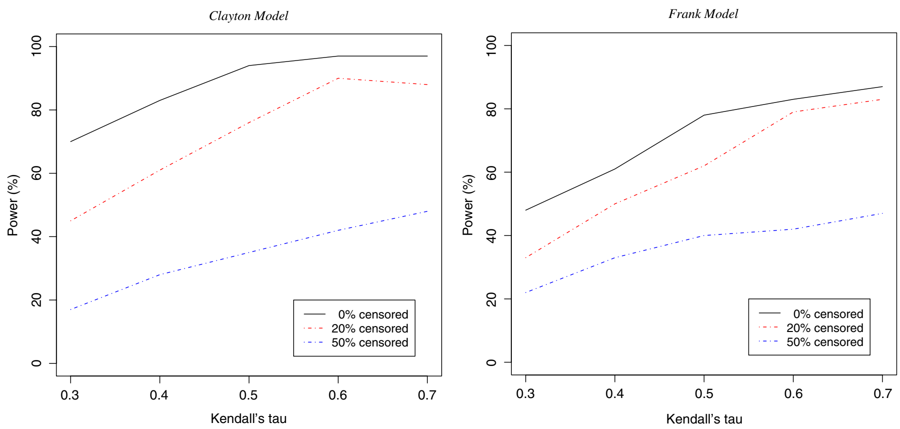

Fig. 1A. Empirical powers with $n = 100$ under H_0: Gumbel versus H_a: Not Gumbel. Powers are the rates of rejecting H_0 with 5% significance during 100 replications.

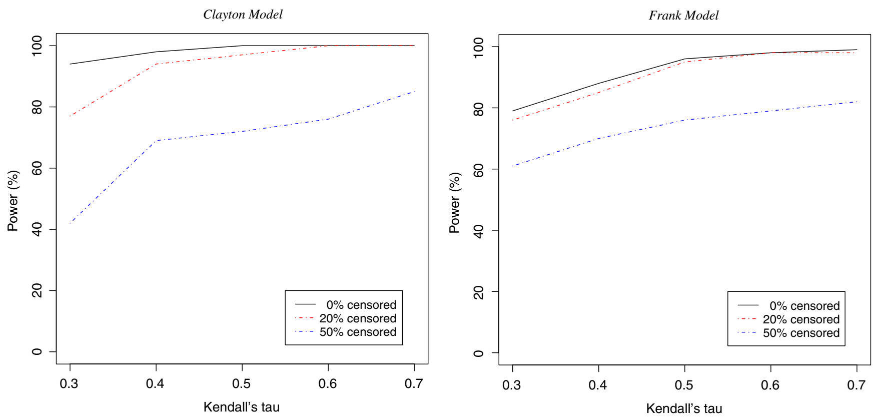

Fig. 1B. Empirical powers with $n = 200$ under H_0: Gumbel versus H_a: Not Gumbel. Powers are the rates of rejecting H_0 with 5% significance during 100 replications.

Table 3A
The goodness-of-fit test results for four AC models based on the Australian Twin Study (Duffy et al., 1990).

	$\hat{\alpha}_1$ ($\hat{\tau}_1$)	$\hat{\alpha}_2$ ($\hat{\tau}_2$)	$(\hat{\gamma}_1 - \hat{\gamma}_2)/\hat{\sigma}_{\text{Jack}}$	p-value
Clayton	1.446 (0.182)	1.717 (0.264)	−1.867	0.000
Frank	1.308 (0.143)	1.496 (0.163)	−1.090	0.117
Gumbel	0.115 (0.103)	0.114 (0.102)	0.084	0.497
Log-copula	1.447 (0.295)	1.147 (0.248)	1.351	0.034

In the second example, X and Y represent the time to the first and second recurrence of infection in kidney patients on dialysis (McGilchrist and Aisbett, 1991). Among 38 patients, 6 observations are censored for X, 12 observations are censored for Y and 3 observations are censored for both X and Y. This data has been analyzed by Wang and Wells (2000), in which Eq. (4) was used to construct a distance measure between nonparametric and model-based curves. For the kidney data, their approach selected the Gumbel model which gives the smallest distance among several alternatives. Note that their method is not a formal statistical test. Applying the proposed test to four model candidates, the p-values are 0.452 (Gumbel), 0.189 (Clayton), 0.365 (Frank) and 0.423 (Log-copula). Due to the small sample size, no model is rejected. However, the Gumbel model still gives the largest p-value, which supports the conclusion of Wang and Wells (2000). Under the Gumbel model, we obtain $\hat{\alpha}_1 = 0.282$ ($\hat{\tau}_1 = 0.220$), $\hat{\alpha}_2 = 0.262$ ($\hat{\tau}_2 = 0.208$).

5. Conclusion

Copula models are often defined without specifying the marginal distributions. Likelihood-based strategies for model selection, such as the Akaike Information Criterion (AIC), cannot be directly applied to such a semiparametric structure. The proposed test extends the approach used by Shih (1998) from the Clayton model to a class of AC models. The property of the conditional moment in (3) is utilized to construct a class of estimating functions for the association parameter. One suggested weight function is derived based on the conditional likelihood function, while the other is the unweighted version. The contrast between the two weights provides a way to detect any violation of the model assumption. Compared with existing methods for model selection for copulas, the proposal yields a formal goodness-of-fit test.

In the presence of right censoring, we propose to delete non-orderable pairs from the estimating equation since the remaining observations still maintain the unbiased property. This strategy has been used by Oakes (1982, 1986) and Shih (1998). Despite its simplicity, data deletion may result in efficiency loss especially when censoring is heavy. The techniques of handling missing data, including imputation and weighting, may be applied to handle the situation that Δ_{ij} is unknown due to censoring (Hsieh, 2010). This direction deserves further investigation. However, if alternative approaches involve extra estimation for nuisance parameters, efficiency gain is not guaranteed.

Acknowledgements

The authors are grateful to the editor and the two referees for their helpful comments on the first version of this paper. We also thank Ms. Sophia Shing for proofreading the article. The research was financially supported by the research Grant 098-2811-M-009-025 funded by National Science Council of Taiwan, ROC.

Appendix A. Derivation of the log-likelihood function

Taking the log of Eq. (6), we obtain the log-likelihood function

$$\begin{aligned} l(\alpha) &= \sum_{(x,y)\in\varphi} D(x,y)\log\left[\frac{\theta_\alpha\{S(x,y)\}}{R(x,y)-1+\theta_\alpha\{S(x,y)\}}\right] + \{1-D(x,y)\}\log\left[\frac{R(x,y)-1}{R(x,y)-1+\theta_\alpha\{S(x,y)\}}\right] \\ &= \sum_i \sum_{j:X_j\ge X_i, Y_j\le Y_i} D(X_i,Y_j)\log\left[\frac{\theta_\alpha\{S(X_i,Y_j)\}}{R(X_i,Y_j)-1+\theta_\alpha\{S(X_i,Y_j)\}}\right] \\ &\quad + \{1-D(X_i,Y_j)\}\log\left[\frac{R(X_i,Y_j)-1}{R(X_i,Y_j)-1+\theta_\alpha\{S(X_i,Y_j)\}}\right] \\ &= \sum_i \log\left[\frac{\theta_\alpha\{S(X_i,Y_i)\}}{R(X_i,Y_i)-1+\theta_\alpha\{S(X_i,Y_i)\}}\right] + \sum_i \sum_{j:X_j>X_i, Y_j<Y_i} \log\left[\frac{R(X_i,Y_j)-1}{R(X_i,Y_j)-1+\theta_\alpha\{S(X_i,Y_j)\}}\right]. \end{aligned}$$

As long as $X_j > X_i$ and $Y_j < Y_i$ hold, $X_i = \tilde{X}_{ij}$ and $Y_j = \tilde{Y}_{ij}$, and hence $S(\tilde{X}_{ij},\tilde{Y}_{ij}) = S(X_i,Y_j)$ and $R(X_i,Y_j) = R_{ij}$. Thus, the second term in the previous equation becomes

$$\begin{aligned} L &\equiv \sum_i \sum_{j:X_j>X_i, Y_j<Y_i} \log\left[\frac{R_{ij}-1}{R_{ij}-1+\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}\right] \\ &= \sum_i \sum_{j:X_j>X_i} I(Y_j<Y_i)\log\left[\frac{R_{ij}-1}{R_{ij}-1+\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}\right]. \end{aligned}$$

It follows that

$$L = \sum_i \sum_{j:X_j>X_i} (1-\Delta_{ij})\log\left[\frac{R_{ij}-1}{R_{ij}-1+\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}\right] = \sum_{i<j} (1-\Delta_{ij})\log\left[\frac{R_{ij}-1}{R_{ij}-1+\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}\right].$$

Appendix B

B.1. Regularity conditions

(i) A parameter space Θ for α is open and the true value lies inside Θ.
(ii) $\theta_\alpha(v)$ is twice differentiable with respect to α and differentiable with respect to v.
(iii) $1/\theta_\alpha(v)$ and $\theta'_\alpha(v) = \frac{\partial}{\partial v}\theta_\alpha(v)$ are bounded with respect to parameters (α, v).
(iv) A and A_L exist and are strictly positive.

B.2. Proof of Lemma 1

First, we show that $\binom{n}{2}^{-1}\{U_2(\alpha)-\tilde{U}_2(\alpha)\}$ converges in probability to zero uniformly in α. It is straightforward to see that

$$|U_2(\alpha)-\tilde{U}_2(\alpha)| = \sum_{i<j}\left|\frac{\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}}{1+\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}}-\frac{\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}{1+\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}\right|$$
$$\leq \sum_{i<j}\left|\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}-\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}\right|.$$

Let $M=\sup_{\alpha,v}|\theta'_\alpha(v)|<\infty$ based on the regularity condition (iii) of Appendix B.1. Applying the Taylor expansion to the last term, it follows that

$$\binom{n}{2}^{-1}|U_2(\alpha)-\tilde{U}_2(\alpha)|\leq M\sup_{x,y}\left|\hat{S}(x,y)-S(x,y)\right|.$$

Based on the Glivenko–Cantelli theorem, the last term converges almost surely to zero uniformly in α.

Next, we show that $\binom{n}{2}^{-1}\{nU_1(\alpha)-\tilde{U}_1(\alpha)\}$ converges in probability to zero uniformly in α. Similarly,

$$|nU_1(\alpha)-\tilde{U}_1(\alpha)| \leq \sum_{i<j}\left|\frac{\dot{\theta}_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}[1+\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}]}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}[\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-1/n+\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}/n]}-\frac{\dot{\theta}_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}[1+\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}]}{\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}S(\tilde{X}_{ij},\tilde{Y}_{ij})}\right|$$
$$+\sum_{i<j}\left|\frac{\dot{\theta}_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}[1+\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}]}{\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}[\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})-1/n+\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}/n]}\right||\theta_\alpha\{\hat{S}(\tilde{X}_{ij},\tilde{Y}_{ij})\}-\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}|.$$

Now we show that the last two terms, multiplied by $\binom{n}{2}^{-1}$, converge in probability to zero uniformly in α. According to the Chebyshev inequality, it is sufficient to show that the quantities

$$I_1=E\left|\frac{\dot{\theta}_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}[1+\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}]}{\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12j})\}[\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})-1/n+\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}/n]}-\frac{\dot{\theta}_\alpha\{S(\tilde{X}_{12},\tilde{Y}_{12})\}[1+\theta_\alpha\{S(\tilde{X}_{12},\tilde{Y}_{12})\}]}{\theta_\alpha\{S(\tilde{X}_{12},\tilde{Y}_{12})\}S(\tilde{X}_{12},\tilde{Y}_{12})}\right|$$

and

$$I_2=E\left|\frac{\dot{\theta}_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}[1+\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}]}{\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12j})\}[\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})-1/n+\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}/n]}\right||\theta_\alpha\{\hat{S}(\tilde{X}_{12},\tilde{Y}_{12})\}-\theta_\alpha\{S(\tilde{X}_{12},\tilde{Y}_{12})\}|$$

converge to zero. Showing this requires substantial technical efforts due to the instability of the denominator terms. The key property in the proof is

$$E\left[\frac{1}{\theta_\alpha\{S(\tilde{X}_{12},\tilde{Y}_{12})\}S(\tilde{X}_{12},\tilde{Y}_{12})}\right]<\infty,$$

which can be shown from Corollary 1 of Shih (1998), under the regularity conditions listed in Appendix B.1. After tedious evaluations similar to the previous arguments and applying the Glivenko–Cantelli theorem to $\hat{S}$, we can show that I_1 and I_2 converge to zero.

B.3. Proof of Lemma 2

Lemma 1 and the Taylor expansion for $U_2(e^\gamma)$ allow us to use the asymptotic expansions

$$n^{1/2}(\gamma_2-\gamma)=\left\{-\binom{n}{2}^{-1}\alpha\dot{\tilde{U}}_2(\alpha^*)\right\}^{-1}\left\{n^{1/2}\binom{n}{2}^{-1}\tilde{U}_2(\alpha)+o_P(1)\right\},$$

where $\dot{\tilde{U}}_2(\alpha)=\partial\tilde{U}_2(\alpha)/\partial\alpha$ and α^* is between α and $\hat{\alpha}_2$. Based on the strong law of large number for U-statistics, the term $-\binom{n}{2}^{-1}\dot{\tilde{U}}_2(\alpha^*)$ converges almost surely to a constant A. Applying the same argument to $U_1(e^\gamma)$, we obtain the asymptotic expansion

$$n^{1/2}(\hat{\gamma}_1-\gamma)=n^{1/2}\binom{n}{2}^{-1}\frac{\tilde{U}_1(\alpha)}{\alpha A_L}+o_P(1),\qquad n^{1/2}(\hat{\gamma}_2-\gamma)=n^{1/2}\binom{n}{2}^{-1}\frac{\tilde{U}_2(\alpha)}{\alpha A}+o_P(1).$$

Combining these formulas, we obtain the U-statistics expression

$$n^{1/2}(\hat{\gamma}_1 - \hat{\gamma}_2) = n^{1/2}\binom{n}{2}^{-1}\sum_{i<j}\frac{1}{\alpha}\left(\frac{\dot{\theta}_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}[\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}+1]}{A_L\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}S(\tilde{X}_{ij},\tilde{Y}_{ij})} - \frac{1}{A}\right)\left[\Delta_{ij} - \frac{\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}}{\theta_\alpha\{S(\tilde{X}_{ij},\tilde{Y}_{ij})\}+1}\right] + o_P(1).$$

Appendix C. Two examples of AC models

Example 1 (*Clayton Model*)**.** The generating function can be written as $\phi_\alpha(v) = (v^{-\alpha} - 1)/\alpha$ for $\alpha \in (0, \infty)$. The joint survival function can be written as

$$\Pr(X > x, Y > y) = \left[S_X(x)^{-\alpha} + S_Y(y)^{-\alpha} - 1\right]^{-1/\alpha}.$$

It follows that $\tau = \alpha/(\alpha + 2)$ and

$$K_\alpha(v) = v - \phi_\alpha(v)/\phi'_\alpha(v) = v + v(1 - v^\alpha)/\alpha.$$

A special property of the Clayton model reflects in its local odds ratio, which can be expressed as $\theta^*(x, y) = \alpha + 1$. The odds ratio does not depend on (x, y).

Example 2 (*Gumbel Model*)**.** The generating function can be written as $\phi_\alpha(v) = \{-\log(v)\}^{\alpha+1}, \alpha \in [0, \infty)$. The joint survival function can be written as

$$\Pr(X > x, Y > y) = \exp\left\{-\left[\{-\log S_X(x)\}^{\alpha+1} + \{-\log S_Y(y)\}^{\alpha+1}\right]^{\frac{1}{\alpha+1}}\right\}.$$

It follows that $\tau = \alpha/(\alpha + 1)$ and

$$K_\alpha(v) = v - \phi_\alpha(v)/\phi'_\alpha(v) = v - v\log(v)/(\alpha + 1).$$

The local odds ratio can be expressed as $\theta^*(x, y) = 1 - \frac{\alpha}{\log S(x,y)}$. Compared with the Clayton model, however, $\theta^*(x, y)$ depends on (x, y).

References

Clayton, D.G., 1978. A model for association in bivariate life tables and its application to epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141–151.

Dabrowska, D., 1988. Kaplan–Meier estimate on the plane. The Annals of Statistics 16, 1475–1489.

Dobric, J., Schmid, F., 2007. A goodness of fit test for copulas based on Rosenblatt's transformation. Computational Statistics and Data Analysis 51, 4633–4642.

Duffy, D.L., Martin, N.G., Matthews, J.D., 1990. Appendectomy in Australian twins. American Journal of Human Genetics 47, 590–592.

Frees, E.W., Valdez, E., 1998. Understanding the relationships using copulas. North American Actuarial Journal 2, 1–25.

Genest, C., Remillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: a review and a power study. Insurance: Mathematics and Economics 44, 199–213.

Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88, 1034–1043.

Gill, R., Schumacher, M., 1987. A simple test of the proportional hazards assumption. Biometrika 74, 289–300.

Hsieh, , 2010. Estimation of Kendall's tau from censored data. Computational Statistics and Data Analysis 54, 1613–1621.

McGilchrist, C.A., Aisbett, C.W., 1991. Regression with frailty in survival analysis. Biometrics 47, 461–466.

Nelsen, R.B., 2006. An Introduction to Copulas, 2nd ed. In: Springer Series in Statistics, Springer-Verlag, New York.

Nikoloulopoulos, A.K., Karlis, D., 2008. Copula model evaluation based on parametric bootstrap. Computational Statistics and Data Analysis 52, 3342–3353.

Oakes, D., 1982. A model for association in bivariate survival data. Journal of the Royal Statistical Society, Series B 44, 414–422.

Oakes, D., 1986. Semiparametric inference in a model for association in bivariate survival data. Biometrika 73, 353–361.

Oakes, D., 1989. Bivariate survival models induced by frailties. Journal of the American Statistical Association 84, 487–493.

Prentice, R.L., Hsu, Li., 1997. Regression on hazard ratios and cross ratios in multivariate failure time analysis. Biometrika 84, 349–363.

Shih, J.H., 1998. A goodness-of-fit test for association in a bivariate survival model. Biometrika 85, 189–200.

Van Der Vaart, A.W., 1998. Asymptotic Statistics. In: Cambridge Series in Statistics and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Wang, W., Wells, M., 2000. Model selection and semiparametric inference for bivariate failure-time data. Journal of the American Statistical Association 95, 62–72.