

**ESTIMATION AND MODEL SELECTION OF
SEMIPARAMETRIC MULTIVARIATE SURVIVAL FUNCTIONS
UNDER GENERAL CENSORSHIP**

By

Xiaohong Chen, Yanqin Fan, Demian Pouzo, and Zhiliang Ying

November 2008

COWLES FOUNDATION DISCUSSION PAPER NO. 1683



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281**

<http://cowles.econ.yale.edu/>

Estimation and model selection of semiparametric multivariate survival functions under general censorship¹

Xiaohong Chen², Yanqin Fan, Demian Pouzo, Zhiliang Ying

First version: March 2005; This version: March 2007

Abstract

Many models of semiparametric multivariate survival functions are characterized by nonparametric marginal survival functions and parametric copula functions, where different copulas imply different dependence structures. This paper considers estimation and model selection for these semiparametric multivariate survival functions, allowing for misspecified parametric copulas and data subject to general censoring. We first establish convergence of the two-step estimator of the copula parameter to the pseudo-true value defined as the value of the parameter that minimizes the KLIC between the parametric copula induced multivariate density and the unknown true density. We then derive its root- n asymptotically normal distribution and provide a simple consistent asymptotic variance estimator by accounting for the impact of the nonparametric estimation of the marginal survival functions. These results are used to establish the asymptotic distribution of the penalized pseudo-likelihood ratio statistic for comparing multiple semiparametric multivariate survival functions subject to copula misspecification and general censorship. An empirical application of the model selection test to the Loss-ALAE insurance data set is provided.

JEL Classification: C14; C22; G22

Keywords: Multivariate survival models; Misspecified copulas; Penalized pseudo-likelihood ratio; Fixed or random censoring; Kaplan-Meier estimator

¹We thank Professors Frees and Valdez for kindly providing the loss-ALAE data, which were collected by the US Insurance Services Office (ISO). Chen and Fan acknowledge financial support from the National Science Foundation. Ying acknowledges financial support from the National Science Foundation and the National Institute of Health. Part of the work was initiated during Chen and Ying's visit to the Institute for Mathematical Sciences at the National University of Singapore whose hospitality and support are acknowledged.

²Corresponding author. Tel.: +212 998 8970; fax: +212 995 4186.

Chen and Pouzo: Department of Economics, New York University, 19 West 4th Street, 6FL, New York, NY 10012, USA; xiaohong.chen@nyu.edu, dgp219@nyu.edu.

Fan: Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA; yanqin.fan@vanderbilt.edu.

Ying: Department of Statistics, Columbia University, 1255 Amsterdam Avenue, 10th Floor, New York, NY 10027, USA; zying@stat.columbia.edu.

1 Introduction

Economic, financial, and medical multivariate survival data are typically non-normally distributed and exhibit nonlinear dependence among their component variables. A class of semiparametric multivariate survival models that have proven to be useful in modeling such data is the class of semiparametric copula-based multivariate survival functions in which the marginal survival functions are nonparametric, but the copula functions characterizing the dependence structure between the component variables are parametrized. More specifically, let $X = (X_1, \dots, X_d)'$ be the survival variables of interest with a d -variate joint survival function: $F^o(x_1, \dots, x_d) = P(X_1 > x_1, \dots, X_d > x_d)$ and marginal survival functions $F_j^o(\cdot)$ ($j = 1, \dots, d$). Assume that F_j^o ($j = 1, \dots, d$) are continuous. A straightforward application of Sklar's (1959) theorem shows that there exists a unique d -variate copula function C^o such that $F^o(x_1, \dots, x_d) \equiv C^o(F_1^o(x_1), \dots, F_d^o(x_d))$, where the copula $C^o(\cdot) : [0, 1]^d \rightarrow [0, 1]$ is itself a multivariate probability distribution function; it captures the dependence structure among the component variables X_1, \dots, X_d . This decomposition of the joint survival function leads naturally to the class of semiparametric multivariate survival functions in which the marginal survival functions are unspecified, but the copula function is parameterized: $C^o(u_1, \dots, u_d) = C^o(u_1, \dots, u_d; \alpha^o)$ for some parametric copula function $C^o(u_1, \dots, u_d; \alpha)$ and some value $\alpha^o \in \mathcal{A}$. As a multivariate survival function in this class depends on nonparametric functions of only one dimension, it achieves dimension reduction while maintaining a more flexible form than purely parametric survival functions. This class of semiparametric multivariate survival functions has been used widely in survival analysis, where modeling and estimating the dependence structure between survival variables is of importance. See Joe (1997), Nelsen (1999), Clayton (1978), Oakes (1989, 1994), Frees and Valdez (1998) and Li (2000) for examples of such applications.

A semiparametric copula-based multivariate survival model has two sets of unknown parameters: the unknown marginal survival functions F_j^o , $j = 1, \dots, d$, and the copula parameter α^o of the parametric copula function $C^o(u_1, \dots, u_d; \alpha^o)$. For complete data (i.e., data without censoring or truncation), Oakes (1994) and Genest et al. (1995) propose a two-step estimation procedure: in first step the marginal distribution functions $1 - F_j^o$, $j = 1, \dots, d$ are estimated by the rescaled empirical distribution functions, in the second step the copula parameter α^o is estimated by maximizing the estimated log-likelihood function. For randomly right censored data, Shih and Louis (1995) independently propose the same two-step

procedure, except that the Kaplan-Meier estimators of marginal survival functions are used in the first step. For a random sample of size n , Genest et al. (1995) establish the root- n consistency and asymptotic normality of their two-step estimator of α° . For randomly right censored data, Shih and Louis (1995) derive similar large sample properties of their two-step estimator of α° under the assumption of bounded partial derivatives of score functions. Unfortunately, this assumption is violated by many commonly used copulas including the Gaussian copula, the Student's t copula, Clayton copula and Gumbel copula. In addition, Shih and Louis (1995) assume that the censoring scheme is i.i.d. random and the parametric copula function is correctly specified.

A closely related important issue in applying this class of semiparametric survival functions to a given data set is how to choose an appropriate parametric copula, as different parametric copulas lead to survival functions that may have very different dependence properties. A number of existing papers have attempted to address this issue. For complete data, we refer to Chen and Fan (2005, 2006a) for a detailed discussion of existing approaches and references. For bivariate censored data, existing work include Frees and Valdez (1998), Klugman and Parsa (1999), Wang and Wells (2000), Chen and Fan (2007), and Denuit et al. (2004). Frees and Valdez (1998) and Klugman and Parsa (1999) consider fully parametric models of bivariate distribution (or survival) functions, and they address model selection of parametric copulas and parametric marginals for insurance company data on losses and allocated loss adjustment expenses (ALAEs). The particular data set they use were collected by the US Insurance Services Office in which loss is censored by a fixed censoring mechanism and ALAE is not censored. Using various model selection techniques including AIC/BIC, Frees and Valdez (1998) select the Pareto marginal distributions and the Gumbel copula, while Klugman and Parsa (1999) select inverse paralogistic for loss marginal distribution, inverse Burr for ALAE marginal distribution and Frank copula. Wang and Wells (2000), Denuit et al. (2004) and Chen and Fan (2007) consider model selection of semiparametric bivariate distribution (or survival) functions in which they do not specify marginals, but restrict the parametric copulas to be in the Archimedean family. In particular, Wang and Wells (2000) propose a model selection procedure for comparing copulas in the one-parameter Archimedean family, allowing for various censoring mechanisms, as long as a consistent nonparametric estimator for the bivariate joint distribution (or survival) function is available. Their selection procedure is based on comparing point estimates of the

integrated squared difference between the true Archimedean copula and a parametric copula; the one with the smallest value of the integrated squared difference is chosen over the rest of the one-parameter Archimedean copulas. Denuit et al. (2004) apply Wang and Wells's (2000) procedure to copula model selection for the same Loss-ALAE data set studied in Frees and Valdez (1998). They use a nonparametric estimator of the bivariate distribution that takes into account the fixed censoring mechanism underlying the Loss-ALAE data. They examine four one-parameter Archimedean copulas (Gumbel, Clayton, Frank and Joe) and select Gumbel copula since it yields the smallest estimated integrated squared difference. Chen and Fan (2007) propose a model selection test for comparing multiple semiparametric bivariate survival functions by taking into account the randomness in the estimated integrated squared difference. However, their test is still only applicable to model selection of parametric copulas within Archimedean family only. It is known that one or two-parameter Archimedean copula family could be too restrictive to capture various dependence structures among multivariate variables. In addition, the semiparametric model selection procedures in Wang and Wells (2000), Denuit et al. (2004) and Chen and Fan (2007) require consistent nonparametric estimation of the joint distribution function and the limiting distributions are complicated. As a result, even for parametric Archimedean copula family, these tests are difficult to implement for multivariate (higher than bivariate) data with general censorship.

In this paper we bridge the gap in existing work for estimating and selecting a semiparametric multivariate copula-based survival model by (i) allowing for data to be censored under various censoring mechanisms, (ii) using nonparametric estimation of marginal survival functions only, (iii) permitting any parametric copula specification, which may be misspecified, non-Archimedean, and its score function may have unbounded partial derivatives. For random samples without censoring, Chen and Fan (2005) already consider the Pseudo-likelihood estimation of copula parameters and Pseudo-likelihood ratio (PLR) model selection test for semiparametric multivariate copula-based distribution models, accounting for (ii) and (iii). In this paper, we extend their results to allow for general right censorship. In particular, we first establish convergence of the two-step estimator of the copula parameter to the pseudo-true value defined as the value of the parameter that minimizes the Kullback-Leibler Information Criterion (KLIC) between the parametric copula induced multivariate density and the unknown true density. We then derive its root- n asymptotically normal distribution and provide a simple consistent asymptotic variance estimator by accounting for (i), (ii) and

(iii). These results are used to establish the asymptotic distribution of the penalized PLR statistic for comparing multiple semiparametric multivariate survival functions subject to copula misspecification and general censorship. We also propose a standardized version of the test, whose limiting null distribution is easy to simulate. To illustrate the usefulness of our testing procedure, we apply it to copula model selection for the loss-ALAE data, taking into account the underlying censoring mechanism in the data and allowing parametric copulas to exhibit more flexible dependence structures than those in the Archimedean family. We find that the standardized test is generally more powerful than the non-standardized test.

The rest of this paper is organized as follows. Section 2 introduces the model selection criterion function and the two-step estimation of the copula dependence parameter. In Section 3, we study the large sample properties of the pseudo-likelihood estimator of the copula parameter allowing for independent but general right censorship and misspecified parametric copulas. In Section 4, we present the limiting null distributions of the (penalized) PLR test statistics for model selection among multiple semiparametric copula models for multivariate censored data. Section 5 provides an empirical application to the Loss-ALAE data set and Section 6 briefly concludes. All technical proofs are gathered into the Appendix.

2 Model selection criterion and parameter estimation

To simplify notation, we shall present our results for bivariate survival models only. Obviously, all these results have straightforward extensions to multivariate copula models for survival data with any finite dimension.

In the following we shall use (D_1, D_2) to denote the censoring variables. Thus under the right censorship, one observes $(\widetilde{X}_1, \widetilde{X}_2) = (X_1 \wedge D_1, X_2 \wedge D_2)$ and a pair of indicators, $(\delta_1, \delta_2) = (I\{X_1 \leq D_1\}, I\{X_2 \leq D_2\})$, where $a \wedge b = \min(a, b)$ for real numbers a and b and $I\{\cdot\}$ is the indicator function. We assume that the censoring variables (D_1, D_2) are independent of the survival variables (X_1, X_2) . Let $F_j^o(x_j) = P(X_j > x_j)$ denote the true but unknown marginal survival function of X_j for $j = 1, 2$. Suppose n independent (but possibly non-identically distributed) observations $\{(\widetilde{X}_{1t}, \widetilde{X}_{2t}, \delta_{1t}, \delta_{2t})\}_{t=1}^n$ are available, where $(\widetilde{X}_{1t}, \widetilde{X}_{2t}) = (X_{1t} \wedge D_{1t}, X_{2t} \wedge D_{2t})$ and $(\delta_{1t}, \delta_{2t}) = (I\{X_{1t} \leq D_{1t}\}, I\{X_{2t} \leq D_{2t}\})$. Denote $U_t = (U_{1t}, U_{2t}) = (F_1^o(\widetilde{X}_{1t}), F_2^o(\widetilde{X}_{2t}))$.

2.1 Model selection criterion

Let $\{C_i(u_1, u_2; \alpha_i) : \alpha_i \in \mathcal{A}_i \subset \mathcal{R}^{p_i}\}$ be a class of parametric copulas with $i = 1, 2, \dots, M$. By Sklar's (1959) theorem, each parametric copula family i corresponds to a parametric likelihood $L_{i,n}(\alpha_i) \equiv \frac{1}{n} \sum_{t=1}^n \ell_i(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_i)$, where

$$\begin{aligned} \ell_i(u_{1t}, u_{2t}, \delta_{1t}, \delta_{2t}; \alpha_i) &= \delta_{1t}\delta_{2t} \log c_i(u_{1t}, u_{2t}; \alpha_i) + \delta_{1t}(1 - \delta_{2t}) \log \frac{\partial C_i(u_{1t}, u_{2t}; \alpha_i)}{\partial u_1} \\ &+ \delta_{2t}(1 - \delta_{1t}) \log \frac{\partial C_i(u_{1t}, u_{2t}; \alpha_i)}{\partial u_2} + (1 - \delta_{1t})(1 - \delta_{2t}) \log C_i(u_{1t}, u_{2t}; \alpha_i), \end{aligned}$$

where $c_i(u_1, u_2; \alpha_i) = \frac{\partial^2 C_i(u_1, u_2; \alpha_i)}{\partial u_1 \partial u_2}$ is the density function of copula $C_i(u_1, u_2; \alpha_i)$.

In this paper, we are interested in testing whether a benchmark model (say copula model 1) performs significantly better than the rest of the copula models according to the KLIC. Let E^0 denote the expectation with respect to the true probability measure. Define

$$\alpha_{in}^* = \arg \max_{\alpha_i \in \mathcal{A}_i} n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_i)]$$

as the pseudo-true value that minimizes the KLIC between the i -th parametric copula family induced multivariate density and the unknown true density. To conclude that copula model 1 performs significantly better than the rest of the copula models calls for a formal statistical test, where the null hypothesis is:

$$H_0 : \max_{i=2, \dots, M} n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_{1n}^*)] \leq 0,$$

meaning that none of the copula models $2, \dots, M$ is closer to the true model (according to KLIC) than model 1, and the alternative hypothesis is:

$$H_1 : \max_{i=2, \dots, M} n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_{1n}^*)] > 0,$$

meaning that there exists a copula model from $2, \dots, M$ that is closer to the true model (according to KLIC) than model 1.

2.2 Two-step estimation

To construct a test statistic for the null hypothesis H_0 against the alternative H_1 , we need estimates of $(U_{1t}, U_{2t}) = (F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}))$ and α_{in}^* for $i = 1, \dots, M$.

For $j = 1, 2$, let $\tilde{F}_j(\cdot)$ be the Kaplan-Meier estimator of $F_j^o(\cdot) = P(X_j > \cdot)$:

$$\tilde{F}_1(x) = \prod_{\tilde{X}_{1(t)} \leq x} \left(1 - \frac{1}{n - t + 1}\right)^{\delta_{1(t)}}, \quad \tilde{F}_2(x) = \prod_{\tilde{X}_{2(t)} \leq x} \left(1 - \frac{1}{n - t + 1}\right)^{\delta_{2(t)}},$$

where $\tilde{X}_{j(1)} \leq \tilde{X}_{j(2)} \leq \dots \leq \tilde{X}_{j(n)}$ are order statistics of $\{\tilde{X}_{jt}\}_{t=1}^n$ for $j = 1, 2$, and $\{\delta_{j(t)}\}_{t=1}^n$ ($j = 1, 2$) are similarly defined. Then under independent censoring, $\tilde{F}_j(\cdot)$ is consistent for $F_j^o(\cdot)$, $j = 1, 2$; see e.g., Lai and Ying (1991).

Given the definition of α_{in}^* , a natural estimator for it is the pseudo-likelihood estimator $\hat{\alpha}_{in}$:

$$\hat{\alpha}_{in} = \arg \max_{\alpha_i \in \mathcal{A}_i} n^{-1} \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}), \delta_{1t}, \delta_{2t}; \alpha_i), \quad i = 1, \dots, M.$$

Since this estimation procedure involves the first-step nonparametric estimation of the marginal survival functions $F_j^o(\cdot)$, $j = 1, 2$, the estimator $\hat{\alpha}_{in}$ is also called the ‘‘two-step’’ estimator.

Note that no assumption is made on the censoring variables (D_{1t}, D_{2t}) other than their independence with the survival variables (X_{1t}, X_{2t}) . As a result, various censoring mechanisms are allowed, including the simple random censoring, fixed censoring, and of course no censoring. If the censoring variables are fixed at $D_{jt} = +\infty$ for $j = 1, 2$, $\hat{\alpha}_{in}$ becomes the estimator proposed in Genest et al. (1995). If the censoring variables (D_{1t}, D_{2t}) are i.i.d. with a continuous joint survival function, $\hat{\alpha}_{in}$ becomes the estimator proposed in Shih and Louis (1995). Assuming that the parametric copula density $c_i(u_1, u_2; \alpha_i)$ is correctly specified and that $\log c_i(u_1, u_2; \alpha_i)$ has bounded partial derivatives with respect to u_1, u_2 , Shih and Louis (1995) establish the root- n asymptotic normality of $\hat{\alpha}_{in}$ and provide a consistent estimator of its asymptotic variance for i.i.d. randomly censored data.

The censoring mechanism for the loss-ALAE data is non-random; ALAE is not censored and Loss is censored by a constant which differs from each individual to another. Results in Shih and Louis (1995) may not be directly applicable to this data set even under correct specification of the copula function. Moreover, for model selection, we need to establish the asymptotic properties of the two-step estimator under copula misspecification. This will be done in the next section for a general censoring mechanism.

2.3 Penalized pseudo-likelihood ratio criteria

To test the null hypothesis H_0 against the alternative H_1 , we use the PLR statistic:

$$LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) = \tilde{L}_{i,n}(\hat{\alpha}_{in}) - \tilde{L}_{1,n}(\hat{\alpha}_{1n}), \quad i = 2, \dots, M,$$

where

$$\tilde{L}_{i,n}(\hat{\alpha}_{in}) \equiv \frac{1}{n} \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}), \delta_{1t}, \delta_{2t}; \hat{\alpha}_{in}), \quad i = 1, \dots, M.$$

In most applications, several parametric copula families are compared which may have different numbers of parameters. To take this into account, we follow the approach in Sin and White (1996) by adopting a general penalization of model complexity. Let $Pen(p_i, n)$ denote a penalization term such that $Pen(p_i, n)$ increases with $p_i \equiv \dim(\mathcal{A}_i)$, decreases with n , and $Pen(p_i, n)/n \rightarrow 0$. Then the penalized PLR statistic is

$$\begin{aligned} PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) &= LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - \frac{Pen(p_i, n) - Pen(p_1, n)}{n} \\ &= LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) + o_p(1). \end{aligned}$$

We note that $Pen(p_i, n) = p_i$ corresponds to AIC, and $Pen(p_i, n) = 0.5p_i \log n$ corresponds to BIC criterion.

In many existing applications of copula models, AIC has been used to compare different families of parametric copula models. To be more specific, let

$$AIC_i = -\frac{2}{n} \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}), \delta_{1t}, \delta_{2t}; \hat{\alpha}_{in}) + \frac{2p_i}{n}, \quad i = 1, \dots, M.$$

Then the values of AIC_i for $i = 1, \dots, M$ are compared; copula model 1 will be selected if $AIC_1 = \min\{AIC_i : 1 \leq i \leq M\}$ or equivalently if

$$LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - \frac{p_i - p_1}{n} < 0, \quad i = 2, \dots, M. \quad (2.1)$$

Noting, however, that $PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})$ (such as AIC_i) is a random variable, the fact that $PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) < 0$ for $i = 2, \dots, M$ (or inequality (2.1) holds) for one sample $\{\tilde{X}_{1t}, \tilde{X}_{2t}, \delta_{1t}, \delta_{2t}\}_{t=1}^n$ may not imply that copula model 1 performs significantly better than the rest of the models; it may occur by chance. As we will show in the next section, $PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) = n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}, \delta_{1t}, \delta_{2t}; \alpha_{1n}^*)] + o_p(1)$ for $i = 2, \dots, M$. To conclude that copula model 1 performs significantly better than the rest of the models we need to perform a formal statistical test for H_0 against H_1 .

To test H_0 , we have to take into account the randomness of the (penalized) PLR statistic. More precisely, we need to derive the asymptotic distributions of $\hat{\alpha}_{in}$ and the test statistics under the null hypothesis. This will be accomplished in Sections 3 and 4 of this paper.

3 Asymptotic properties of the two-step estimator under copula misspecification

As mentioned in the previous section, asymptotic properties of the two-step estimator are established for randomly censored data in Shih and Louis (1995) under the assumptions that

the parametric copula density correctly specifies the true copula density and that its score function has bounded partial derivatives. In this section, we will extend their results to a more general censoring mechanism and allow for misspecified parametric copulas whose score functions may have unbounded partial derivatives.

Recall that $\mathcal{A} \subset \mathcal{R}^p$ is the parameter space. For $\alpha, \alpha^* \in \mathcal{A}$, we use $\|\alpha - \alpha^*\|$ to denote the usual Euclidean metric. To simplify notation, we now let

$$\begin{aligned} \ell(u_1, u_2; \alpha) &= \delta_1 \delta_2 \log c(u_1, u_2; \alpha) + \delta_1 (1 - \delta_2) \log \frac{\partial C(u_1, u_2; \alpha)}{\partial u_1} \\ &\quad + \delta_2 (1 - \delta_1) \log \frac{\partial C(u_1, u_2; \alpha)}{\partial u_2} + (1 - \delta_1)(1 - \delta_2) \log C(u_1, u_2; \alpha), \end{aligned}$$

where $c(u_1, u_2; \alpha)$ is the density of the parametric copula $C(u_1, u_2; \alpha)$. Then the pseudo-true copula parameter value is $\alpha_n^* = \arg \max_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n E^0[\ell(U_{1t}, U_{2t}; \alpha)]$, and its two-step estimator is $\hat{\alpha}_n = \arg \max_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n \ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha)$.

Finally we denote $\ell_\alpha(u_1, u_2; \alpha) = \frac{\partial \ell(u_1, u_2; \alpha)}{\partial \alpha}$, $\ell_j(u_1, u_2; \alpha) = \frac{\partial \ell(u_1, u_2; \alpha)}{\partial u_j}$ ($j = 1, 2$), $\ell_{\alpha\alpha}(u_1, u_2; \alpha) = \frac{\partial^2 \ell(u_1, u_2; \alpha)}{\partial \alpha^2}$ and $\ell_{\alpha j}(u_1, u_2; \alpha) = \frac{\partial^2 \ell(u_1, u_2; \alpha)}{\partial u_j \partial \alpha}$ for $j = 1, 2$.

3.1 Consistency

The following conditions are sufficient to ensure the convergence of the two-step estimator $\hat{\alpha}_n$ to the pseudo true value α_n^* .

C1. (i) The sequence of survival variables, $\{(X_{1t}, X_{2t})\}_{t=1}^n$, is an i.i.d. sample from an unknown survival function $F^o(x_1, x_2)$ with continuous marginal survival functions $F_j^o(\cdot)$, $j = 1, 2$;

(ii) The sequence of censoring variables $\{D_{1t}, D_{2t}\}_{t=1}^n$ is an independent sample with joint survival functions $\{G_t(x_1, x_2)\}_{t=1}^n = \{P(D_{1t} > x_1, D_{2t} > x_2)\}_{t=1}^n$ and marginal survival functions $\{G_{jt}(\cdot)\}_{t=1}^n$, $j = 1, 2$;

(iii) The censoring variables (D_{1t}, D_{2t}) are independent of survival variables (X_{1t}, X_{2t}) and there is no mass concentration at 0 in the sense that $\limsup_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n (1 - G_{jt}(\eta)) \rightarrow 0$ as $\eta \rightarrow 0$.

C2. Let \mathcal{A} be a compact subset of \mathcal{R}^p . For every $\epsilon > 0$,

$$\liminf_{\alpha \in \mathcal{A}: \|\alpha - \alpha_n^*\| \geq \epsilon} \frac{1}{n} \sum_{t=1}^n \left[E^0\{\ell(U_{1t}, U_{2t}; \alpha_n^*)\} - E^0\{\ell(U_{1t}, U_{2t}; \alpha)\} \right] > 0.$$

C3. The true (unknown) copula function $C^o(u_1, u_2)$ has continuous partial derivatives.

C4. (i) For any $(u_1, u_2) \in (0, 1)^2$, $\ell(u_1, u_2; \alpha)$ is a continuous function of $\alpha \in \mathcal{A}$.

(ii) Let $L_t = \sup_{\alpha \in \mathcal{A}} |\ell(U_{1t}, U_{2t}; \alpha)|$ and $L_{t\alpha} = \sup_{\alpha \in \mathcal{A}} |\ell_\alpha(U_{1t}, U_{2t}; \alpha)|$. Then,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n E^0 \{L_t I(L_t \geq K) + L_{t\alpha} I(L_{t\alpha} \geq K)\} = 0;$$

(iii) For any $\eta > 0$, $\epsilon > 0$, there is $K > 0$ such that $|\ell(u_1, u_2; \alpha)| \leq K |\ell(u'_1, u'_2; \alpha)|$ for all $\alpha \in \mathcal{A}$ and all $u_j \in [\eta, 1)$ such that $1 - u_j \geq \epsilon(1 - u'_j)$, $j = 1, 2$.

C5. If $\{X_{jt}\}_{t=1}^n$ are subject to non-trivial censoring (i.e., $D_{jt} \neq \infty$), then \tilde{F}_j is truncated at the tail in the sense that for some τ_j , $\tilde{F}_j(x_j) = \tilde{F}_j(\tau_j)$ for all $x_j \geq \tau_j$ and $\liminf n^{-1} \sum_{t=1}^n G_{jt}(\tau_j) F^o(\tau_j) > 0$.

Note that in contrast to the censoring mechanism in Shih and Louis (1995), Condition C1(ii) allows the censoring variables $\{(D_{1t}, D_{2t})\}_{t=1}^n$ to be non-identically distributed. In addition, no assumption is made on the joint survival function $G_t(x_1, x_2)$ of the censoring variables (D_{1t}, D_{2t}) . Hence Condition C1(ii) includes the fixed censoring mechanism in which each survival variable (X_{1t}, X_{2t}) is censored at a pre-specified, fixed time (D_{1t}, D_{2t}) which may differ from one observation to another, in which case, the survival function $G_t(x_1, x_2)$ is degenerate at (D_{1t}, D_{2t}) . It also allows the variables X_{1t} and X_{2t} to have different censoring mechanisms, one random and the other fixed or one censored and the other uncensored. For example, the censoring mechanism for the Loss-ALAE data is such that Loss is censored by a fixed censoring mechanism and ALAE is uncensored. As a result, the observed variables $\{(\tilde{X}_{1t}, \tilde{X}_{2t})\}_{t=1}^n$ may not be identically distributed and the identifiably unique maximizer α_n^* defined in Condition C2 may depend on n . Condition C5 is imposed to handle the possible tail instability of the Kaplan-Meier estimator, especially for non-identically distributed censoring times. The truncation can be achieved by simply using $D_{jt} \wedge \tau_j$ as the censoring variables. Thus, without loss of generality, we shall assume that $D_{jt} \wedge \tau_j$ are the censoring variables so that $\tilde{X}_{jt} \leq \tau_j$. The simple truncation at τ_j can be changed to the more elaborate tail modification. We refer to Lai and Ying (1991) for the issue of tail instability and modification. Finally, because we allow the left tail of the copula to blow up as well, we shall set $\ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) = 0$ whenever $\tilde{F}_j(\tilde{X}_{jt}) = 1$ for $j = 1$ or 2 .

Proposition 3.1 *Under conditions C1-C5, we have: (1) $\|\hat{\alpha}_n - \alpha_n^*\| = o_p(1)$;*

$$(2) \frac{1}{n} \sum_{t=1}^n \ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_n) = \frac{1}{n} \sum_{t=1}^n E^0 \{ \ell(F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}); \alpha_n^*) \} + o_p(1).$$

Proposition 3.1(1) states that the two-step estimator $\hat{\alpha}_n$ is a consistent estimator of the pseudo true value α_n^* . If the censoring mechanism is random, then $\alpha_n^* = \alpha^*$ which does not

depend on n . In addition, if the parametric copula correctly specifies the true copula, then $\alpha^* = \alpha^o$, where α^o is such that $C(u_1, u_2; \alpha^o) = C^o(u_1, u_2)$ for almost all $(u_1, u_2) \in (0, 1)^2$.

3.2 Asymptotic normality

Recall that $(U_{1t}, U_{2t}) = (F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}))$. For $j = 1, 2$, we denote

$$W_j(\tilde{X}_{jt}, \delta_{jt}; \alpha_n^*) \equiv E^0[\ell_{\alpha_j}(U_{1s}, U_{2s}; \alpha_n^*) I_j^o(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js}) \mid \tilde{X}_{jt}, \delta_{jt}],$$

$$I_j^o(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js}) \equiv -F_j^o(\tilde{X}_{js}) \left[\int_{-\infty}^{\tilde{X}_{js}} \frac{dN_{jt}(u)}{P_{n,j}(u)} - \int_{-\infty}^{\tilde{X}_{js}} \frac{I\{\tilde{X}_{jt} \geq u\} d\Lambda_j(u)}{P_{n,j}(u)} \right],$$

with $\Lambda_j(u) \equiv -\log(F_j^o(u))$ the cumulative hazard function of X_j , $N_{jt}(u) \equiv \delta_{jt} I\{\tilde{X}_{jt} \leq u\}$ and $dN_{jt}(u) = N_{jt}(u) - N_{jt}(u-)$, and $P_{n,j}(u) \equiv n^{-1} \sum_{k=1}^n P(\tilde{X}_{jk} \geq u) = n^{-1} \sum_{k=1}^n G_{jk}(u) F_j^o(u)$.

Let Var^0 denote the variance with respect to the true probability measure. The following conditions are sufficient to ensure the asymptotic normality of $\hat{\alpha}_n$.

A1. (i) C2 holds with $\alpha_n^* \in \text{int}(A^*)$ for all n , where A^* is a compact subset of \mathcal{A} ;

(ii) $B_n \equiv -n^{-1} \sum_{t=1}^n E^0\{\ell_{\alpha\alpha}(U_{1t}, U_{2t}; \alpha_n^*)\}$ has all its eigenvalues bounded below and above by some finite positive constants;

(iii) $\Sigma_n \equiv n^{-1} \sum_{t=1}^n Var^0\{\ell_{\alpha}(U_{1t}, U_{2t}; \alpha_n^*) + W_1(\tilde{X}_{1t}, \delta_{1t}; \alpha_n^*) + W_2(\tilde{X}_{2t}, \delta_{2t}; \alpha_n^*)\}$ has all its eigenvalues bounded below and above by some finite positive constants;

(iv) $\{\ell_{\alpha}(U_{1t}, U_{2t}; \alpha_n^*) + W_1(\tilde{X}_{1t}, \delta_{1t}; \alpha_n^*) + W_2(\tilde{X}_{2t}, \delta_{2t}; \alpha_n^*)\}_{t=1}^n$ satisfies Lindeberg condition.

A2. Functions $\ell_{\alpha\alpha}(u_1, u_2; \alpha)$ and $\ell_{\alpha_j}(u_1, u_2; \alpha)$, $j = 1, 2$, are well-defined and continuous in $(u_1, u_2, \alpha) \in (0, 1)^2 \times \mathcal{A}$.

A3. (i) $|\ell_{\alpha}(u_1, u_2; \alpha_n^*)| \leq q\{u_1(1-u_1)\}^{-a_1}\{u_2(1-u_2)\}^{-a_2}$ for some $q > 0$ and $a_j \geq 0$ such that $\limsup n^{-1} \sum_{t=1}^n E^0[\{U_{1t}(1-U_{1t})\}^{-2a_1}\{U_{2t}(1-U_{2t})\}^{-2a_2}] < \infty$;

(ii) $|\ell_{\alpha_j}(u_1, u_2; \alpha_n^*)| \leq \text{const.}\{u_j(1-u_j)\}^{-b_j}\{u_k(1-u_k)\}^{-a_k}$ for some b_k, a_k and $j \neq k$ such that $\limsup n^{-1} \sum_{t=1}^n E^0[\{U_{jt}(1-U_{jt})\}^{\xi_j - b_j}\{U_{kt}(1-U_{kt})\}^{-a_k}] < \infty$ for some $\xi_j \in (0, 1/2)$.

A4. (i) Let $L_{t\alpha_j} = \sup_{\alpha \in \mathcal{A}} |\ell_{\alpha_j}(U_{1t}, U_{2t}; \alpha)|$ and $L_{t\alpha\alpha} = \sup_{\alpha \in \mathcal{A}} |\ell_{\alpha\alpha}(U_{1t}, U_{2t}; \alpha)|$. Then,

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n E^0\{L_{t\alpha_j} I(L_{t\alpha_j} \geq K) + L_{t\alpha\alpha} I(L_{t\alpha\alpha} \geq K)\} = 0;$$

(ii) For any $\eta > 0$ and any $\epsilon > 0$, there is $K > 0$, such that

$$|\ell_{\alpha}(u_1, u_2; \alpha)| + |\ell_{\alpha\alpha}(u_1, u_2; \alpha)| \leq K\{|\ell_{\alpha}(u'_1, u'_2; \alpha)| + |\ell_{\alpha\alpha}(u'_1, u'_2; \alpha)|\}$$

for all $\alpha \in \mathcal{A}$ and all $u_j \in [\eta, 1)$ such that $1 - u_j \geq \epsilon(1 - u'_j)$, $j = 1, 2$.

Shih and Louis (1995) require bounded $\ell_\alpha(u_1, u_2; \alpha_n^*)$ and $\ell_{\alpha_j}(u_1, u_2; \alpha_n^*)$ for $j = 1, 2$, however, this requirement is not satisfied by many popular copula functions such as Gaussian copula, t-copula, Gumbel copula and Clayton copula. Conditions A3 and A4 relax the boundedness requirement, and allow the score function and its partial derivatives with respect to the first two arguments to blow up at the boundaries. Similar conditions have been verified for Gaussian, Frank and Clayton copulas in Chen and Fan (2006b).

Proposition 3.2 *Under conditions C1-C5 and A1-A4, we have: $B_n \Sigma_n^{-1/2} \sqrt{n}(\hat{\alpha}_n - \alpha_n^*) \rightarrow N(0, I_p)$ in distribution, where B_n and Σ_n are defined in A1.*

Proposition 3.2 extends Theorem 2 in Shih and Louis (1995) in two directions: (i) it allows for more general censoring mechanisms than the simple random censoring in Shih and Louis (1995), and (ii) it allows for the possibility that the parametric copula may not specify the true copula correctly. As a result, there are several differences between Proposition 3.2 and Theorem 2 in Shih and Louis (1995): First, since the censoring variables $\{(D_{1t}, D_{2t})\}_{t=1}^n$ may not be identically distributed, B_n and Σ_n may depend on n ; Second, since the parametric copula may misspecify the true copula, the information matrix equality may not hold. Consequently, the asymptotic variance of $\sqrt{n}(\hat{\alpha}_n - \alpha_n^*)$, $B_n^{-1} \Sigma_n B_n^{-1}$, can not be reduced to $[B_n^{-1} + n^{-1} B_n^{-1} \sum_{t=1}^n Var^0\{W_1(\tilde{X}_{1t}, \delta_{1t}; \alpha_n^*) + W_2(\tilde{X}_{2t}, \delta_{2t}; \alpha_n^*)\} B_n^{-1}]$ as in Shih and Louis (1995). For complete data, Proposition 3.2 reduces to that in Chen and Fan (2005a).

To estimate the asymptotic variance $B_n^{-1} \Sigma_n B_n^{-1}$ of $\sqrt{n}(\hat{\alpha}_n - \alpha_n^*)$, we let

$$\begin{aligned} \hat{B}_n &= -n^{-1} \sum_{t=1}^n \ell_{\alpha\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_n), \\ \hat{\Sigma}_n &= \frac{1}{n} \sum_{t=1}^n \left\{ \ell_\alpha(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_n) + \widehat{W}_1(\tilde{X}_{1t}, \delta_{1t}; \hat{\alpha}_n) + \widehat{W}_2(\tilde{X}_{2t}, \delta_{2t}; \hat{\alpha}_n) \right\} \times \\ &\quad \left\{ \ell_\alpha(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_n) + \widehat{W}_1(\tilde{X}_{1t}, \delta_{1t}; \hat{\alpha}_n) + \widehat{W}_2(\tilde{X}_{2t}, \delta_{2t}; \hat{\alpha}_n) \right\}', \end{aligned}$$

with

$$\begin{aligned} \widehat{W}_1(\tilde{X}_{1t}, \delta_{1t}; \hat{\alpha}_n) &= \frac{1}{n} \sum_{s \neq t, s=1}^n \ell_{\alpha 1}(\tilde{F}_1(\tilde{X}_{1s}), \tilde{F}_2(\tilde{X}_{2s}); \hat{\alpha}_n) \hat{I}_1^o(\tilde{X}_{1t}, \delta_{1t})(\tilde{X}_{1s}), \\ \widehat{W}_2(\tilde{X}_{2t}, \delta_{2t}; \hat{\alpha}_n) &= \frac{1}{n} \sum_{s \neq t, s=1}^n \ell_{\alpha 2}(\tilde{F}_1(\tilde{X}_{1s}), \tilde{F}_2(\tilde{X}_{2s}); \hat{\alpha}_n) \hat{I}_2^o(\tilde{X}_{2t}, \delta_{2t})(\tilde{X}_{2s}), \end{aligned}$$

in which for $j = 1, 2$,

$$\begin{aligned} & \hat{I}_j^\circ(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js}) \\ &= -\tilde{F}_j(\tilde{X}_{js}) \left[\frac{I\{\tilde{X}_{jt} \leq \tilde{X}_{js}\} \delta_{jt}}{n^{-1} \sum_{k=1}^n I\{\tilde{X}_{jk} \geq \tilde{X}_{jt}\}} - \frac{1}{n} \sum_{l=1}^n \frac{I\{\tilde{X}_{js} \geq \tilde{X}_{jl}\} I\{\tilde{X}_{jt} \geq \tilde{X}_{jl}\} \delta_{jl}}{\left[n^{-1} \sum_{k=1}^n I\{\tilde{X}_{jk} \geq \tilde{X}_{jl}\} \right]^2} \right]. \end{aligned} \quad (3.1)$$

We note that an alternative expression for $\hat{I}_j^\circ(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js})$ is:

$$\hat{I}_j^\circ(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js}) = -\tilde{F}_j(\tilde{X}_{js}) \left[\frac{I\{\tilde{X}_{jt} \leq \tilde{X}_{js}, \delta_{jt} = 1\}}{\hat{P}_{n,j}(\tilde{X}_{jt})} - \sum_{\tilde{X}_{jl} \leq \tilde{X}_{js}} \frac{I\{\tilde{X}_{jt} \geq \tilde{X}_{jl}\} \Delta \hat{\Lambda}_j(\tilde{X}_{jl})}{\hat{P}_{n,j}(\tilde{X}_{jl})} \right],$$

where $\hat{P}_{n,j}(u) \equiv n^{-1} \sum_{k=1}^n I\{\tilde{X}_{jk} \geq u\}$,

$$\Delta \hat{\Lambda}_j(u) = \frac{I\{\bar{Y}_j(u) > 0\}}{\bar{Y}_j(u)} d\bar{N}_j(u), \quad \bar{Y}_j(u) = \sum_{k=1}^n I\{\tilde{X}_{jk} \geq u\}, \quad \bar{N}_j(u) = \sum_{k=1}^n N_{jk}(u),$$

in which $\Delta \hat{\Lambda}_j(u)$ is so-called Nelson's estimator. This is because

$$\begin{aligned} \sum_{\tilde{X}_{jl} \leq \tilde{X}_{js}} \frac{I\{\tilde{X}_{jt} \geq \tilde{X}_{jl}\} \Delta \hat{\Lambda}_j(\tilde{X}_{jl})}{\hat{P}_{n,j}(\tilde{X}_{jl})} &= \sum_{\tilde{X}_{jl} \leq \tilde{X}_{js}} \frac{I\{\tilde{X}_{jt} \geq \tilde{X}_{jl}\} \delta_{jl}}{\hat{P}_{n,j}(\tilde{X}_{jl}) \sum_{k=1}^n I\{\tilde{X}_{jk} \geq \tilde{X}_{jl}\}} \\ &= \frac{1}{n} \sum_{l=1}^n \frac{I\{\tilde{X}_{js} \geq \tilde{X}_{jl}\} I\{\tilde{X}_{jt} \geq \tilde{X}_{jl}\} \delta_{jl}}{\left[n^{-1} \sum_{k=1}^n I\{\tilde{X}_{jk} \geq \tilde{X}_{jl}\} \right]^2}. \end{aligned}$$

By the consistency of the Kaplan-Meier estimators and $\hat{\alpha}_n$, and by applying the law of large numbers to independent observations, we can prove the following result, which provides a consistent variance estimator.

Proposition 3.3 *Under conditions C1-C5 and A1-A4, the asymptotic variance of $n^{1/2} \hat{\alpha}_n$ can be consistently estimated by $\hat{B}_n^- \hat{\Sigma}_n \hat{B}_n^-$, where \hat{B}_n^- is the generalized inverse of \hat{B}_n .*

4 Pseudo-likelihood ratio test for model comparison

By applying Proposition 3.1(2) we immediately obtain the probability limit of the PLR statistic.

Proposition 4.1 *Suppose for $i = 1, \dots, M$, the copula model i satisfies the conditions of Proposition 3.1. Then*

$$LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) = \frac{1}{n} \sum_{t=1}^n E^0 \{ \ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*) \} + o_p(1),$$

where $U_{jt} = F_j^\circ(\tilde{X}_{jt})$ for $j = 1, 2$.

In the following, we adopt the convention that all the notations involving the copula function $C(u_1, u_2; \alpha)$ introduced in Section 3 are now indexed by a subscript i for $i = 1, \dots, M$ to make explicit their dependence on the parametric copula model i . In addition, we define $U_t = (U_{1t}, U_{2t}) = (F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}))$,

$$\begin{aligned} e_t &= (e_{2t}, \dots, e_{Mt})', \\ e_{it} &\equiv \{\ell_i(U_t; \alpha_{in}^*) - \ell_1(U_t; \alpha_{1n}^*)\} + \sum_{j=1}^2 \{Q_{i,j}(\tilde{X}_{jt}, \delta_{jt}; \alpha_{in}^*) - Q_{1,j}(\tilde{X}_{jt}, \delta_{jt}; \alpha_{1n}^*)\}, \end{aligned}$$

where for $i = 1, \dots, M$ and for $j = 1, 2$,

$$Q_{i,j}(\tilde{X}_{jt}, \delta_{jt}; \alpha_{in}^*) \equiv E^0 \left[\ell_{i,j}(U_{1s}, U_{2s}; \alpha_{in}^*) I_j^o(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js}) \mid \tilde{X}_{jt}, \delta_{jt} \right].$$

It is easy to see that $\frac{1}{\sqrt{n}} \sum_{t=1}^n \{e_t - E^0(e_t)\}$ has the same asymptotic distribution as a multivariate normal random variable with mean zero and variance Ω_n , where

$$\begin{aligned} \Omega_n &= \frac{1}{n} \sum_{t=1}^n E^0 \left[(e_t - E^0\{e_t\})(e_t - E^0\{e_t\})' \right] = (\sigma_{ik})_{i,k=2}^M, \\ \sigma_{ik} &= \frac{1}{n} \sum_{t=1}^n E^0 \left[(e_{it} - E^0\{e_{it}\})(e_{kt} - E^0\{e_{kt}\}) \right]. \end{aligned}$$

It is easy to compute a consistent estimator $\hat{\Omega}_n$ for Ω_n :

$$\begin{aligned} \hat{\Omega}_n &= \frac{1}{n} \sum_{t=1}^n \left[\left(\hat{e}_t - \frac{1}{n} \sum_{s=1}^n \hat{e}_s \right) \left(\hat{e}_t - \frac{1}{n} \sum_{s=1}^n \hat{e}_s \right)' \right] = (\hat{\sigma}_{ik})_{i,k=2}^M, \\ \hat{\sigma}_{ik} &= \frac{1}{n} \sum_{t=1}^n \left(\hat{e}_{it} - \frac{1}{n} \sum_{s=1}^n \hat{e}_{is} \right) \left(\hat{e}_{kt} - \frac{1}{n} \sum_{s=1}^n \hat{e}_{ks} \right), \end{aligned} \tag{4.1}$$

where $\hat{e}_t = (\hat{e}_{2t}, \dots, \hat{e}_{Mt})'$ and for $i = 2, \dots, M$,

$$\begin{aligned} \hat{e}_{it} &\equiv \left\{ \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{in}) - \ell_1(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{1n}) \right\} \\ &\quad + \sum_{j=1}^2 \left\{ \hat{Q}_{i,j}(\tilde{X}_{jt}, \delta_{jt}; \hat{\alpha}_{in}) - \hat{Q}_{1,j}(\tilde{X}_{jt}, \delta_{jt}; \hat{\alpha}_{1n}) \right\}, \end{aligned}$$

in which

$$\hat{Q}_{i,j}(\tilde{X}_{jt}, \delta_{jt}; \hat{\alpha}_{in}) = \frac{1}{n} \sum_{s \neq t, s=1}^n \ell_{i,j}(\tilde{F}_1(\tilde{X}_{1s}), \tilde{F}_2(\tilde{X}_{2s}); \hat{\alpha}_{in}) \hat{I}_j^o(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js}),$$

for $i = 1, \dots, M$ and $j = 1, 2$ with $\hat{I}_j^o(\tilde{X}_{jt}, \delta_{jt})(\tilde{X}_{js})$ given in (3.1).

Before we present the test statistics, we recall the following definition from Chen and Fan (2005): For model $i \in \{2, \dots, M\}$,

Models 1 and i are *generalized non-nested* if the set $\{(v_1, v_2) : c_1(v_1, v_2; \alpha_{1n}^*) \neq c_i(v_1, v_2; \alpha_{in}^*)\}$ has positive Lebesgue measure;

Models 1 and i are *generalized nested* if $c_1(v_1, v_2; \alpha_{1n}^*) = c_i(v_1, v_2; \alpha_{in}^*)$ for almost all $(v_1, v_2) \in (0, 1)^2$.

Given the definition of the pseudo true value α_{in}^* , the closest $c_i(\cdot; \alpha_{in}^*)$ to the true copula c^0 (according to KLIC) in a parametric class of copulas $\{c_i(\cdot; \alpha_i) : \alpha_i \in \mathcal{A}_i\}$ depends on the true (but unknown) copula. Hence it is not obvious a priori whether two parametric classes of copulas are generalized non-nested or generalized nested.

Remark 4.1: Define

$$\sigma_{ii}^a \equiv \frac{1}{n} \sum_{t=1}^n \text{Var}^0[\ell_i(U_t; \alpha_{in}^*) - \ell_1(U_t; \alpha_{1n}^*)].$$

It is obvious that if models 1 and i are generalized nested, then $\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) = \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)$ almost surely, $e_{it} = 0$ almost surely, and $\sigma_{ii}^a = 0$, $\sigma_{ii} = 0$. Following the proof of proposition 3 in Chen and Fan (2005), we can show that if $\sigma_{ii}^a = 0$ then models 1 and i are generalized nested, and $\sigma_{ii} = 0$. Therefore it is easy to test whether the models 1 and i are generalized nested by testing $\sigma_{ii}^a = 0$, which may be done by using its consistent estimator:

$$\hat{\sigma}_{ii}^a = \frac{1}{n} \sum_{t=1}^n \left[\left\{ \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{in}) - \ell_1(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{1n}) \right\} - LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) \right]^2.$$

See Chen and Fan (2005) for details.

The following proposition provides the basis for our tests. Note that we allow for some but not all of the candidate models $i \in \{2, \dots, M\}$ to be generalized nested with the benchmark model 1.

Proposition 4.2 *For $i = 1, 2, \dots, M$, assume that the copula model i satisfies conditions of Proposition 3.2 and that $\{e_{it} : t = 1, \dots, n\}$ satisfies Lindeberg condition. If $\Omega_n = (\sigma_{ik})_{i,k=2}^M$ is finite and its largest eigenvalue is positive uniformly in n , then: (1)*

$$\begin{aligned} & n^{1/2} \left[LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right]_{i=2, \dots, M} \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \{e_t - E^0(e_t)\} + o_p(1), \\ &\rightarrow (Z_2, \dots, Z_M)' \quad \text{in distribution, with } (Z_2, \dots, Z_M)' \sim N(0, \Omega_n). \end{aligned}$$

(2) $\hat{\Omega}_n = \Omega_n + o_p(1)$.

Proposition 4.2 and the continuous mapping theorem imply

$$\begin{aligned} & \max_{i=2,\dots,M} n^{1/2} \left\{ LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right\} \\ & \rightarrow \max_{i=2,\dots,M} Z_i. \end{aligned}$$

Define

$$T_n \equiv \max_{i=2,\dots,M} [n^{1/2} LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})].$$

Proposition 4.2 implies that under the Least Favorable Configuration (LFC), i.e.,

$$n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] = 0 \quad \text{for } i = 2, \dots, M,$$

$T_n \rightarrow \max_{i=2,\dots,M} Z_i$ in distribution. This allows us to construct a test for H_0 . Suppose the largest eigenvalue of Ω_n is positive uniformly in n , then we will reject H_0 if $T_n > Z_\alpha$, where Z_α is the upper α -percentile of the distribution of $\max_{i=2,\dots,M} Z_i$.

The asymptotic power properties of this test against fixed alternatives and Pitman local alternatives follow immediately from Proposition 4.2 and are summarized in the following proposition.

Proposition 4.3 *Suppose all conditions of Proposition 4.2 are satisfied. Then the test based on T_n is consistent against fixed alternatives of the form H_1 and has non-trivial power against local alternatives satisfying*

$$\max_{i=2,\dots,M} \lim_{n \rightarrow \infty} \left\{ n^{-1/2} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right\} > 0.$$

Note that if the censoring mechanism is random, then the local alternatives in Proposition 4.3 can be written in the more familiar form:

$$\max_{i=2,\dots,M} E^0[\ell_i(U_{1t}, U_{2t}; \alpha_i^*) - \ell_1(U_{1t}, U_{2t}; \alpha_1^*)] = K \frac{1}{\sqrt{n}},$$

for a positive constant K .

In general, the distribution of $\max_{i=2,\dots,M} Z_i$ is unknown, since the asymptotic variance Ω_n of (Z_2, \dots, Z_M) depends on $\alpha_{1n}^*, \dots, \alpha_{Mn}^*$. Following White (2000), one can use either ‘‘Monte-Carlo RC’’ p-value or ‘‘bootstrap RC’’ p-value to implement this test. As noted in Chen and Fan (2005), Hansen (2003), and Romano and Wolf (2005), the finite sample power of this

test may be improved by standardization. In our empirical application, we have computed both “Monte-Carlo RC” p-value using

$$T_{nS} = \max_{i=2, \dots, M} \left\{ \frac{n^{1/2} LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})}{\sqrt{\hat{\sigma}_{ii}}} G_b(\hat{\sigma}_{ii}) \right\},$$

and “bootstrap RC” p-value based on

$$T_{nI} = \max \left[\max_{i=2, \dots, M} \left\{ \frac{n^{1/2} LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})}{\sqrt{\hat{\sigma}_{ii}}} G_b(\hat{\sigma}_{ii}) \right\}, 0 \right],$$

where $\hat{\sigma}_{ii}$ is a consistent estimator of σ_{ii} such as the one given in (4.1), $b = b_n \rightarrow 0$ as $n \rightarrow \infty$, and $G_b(\cdot)$ is a smoothed trimming function which trims out small $\hat{\sigma}_{ii}$. The particular trimming function being used in our empirical study is

$$G_b(x) = \int_{-\infty}^x g_b(z) dz = \begin{cases} 0, & x < b \\ \int_{-\infty}^x g_b(z) dz, & b \leq x \leq 2b \\ 1, & x > 2b. \end{cases}$$

where $g_b(x) = b^{-1}g(b^{-1}x - 1)$ and $g(z) = B(a + 1)^{-1}z^a(1 - z)^a$, $z \in [0, 1]$ for some positive integer $a \geq 1$, where $B(a) = \Gamma(a)^2/\Gamma(2a)$ is the beta function and $\Gamma(a)$ is the Euler gamma function.

We note that the standardized tests T_{nS} and T_{nI} proposed here allow that some candidate models are generalized nested with the benchmark model, since the trimming $G_b(\hat{\sigma}_{ii})$ in T_{nS} and T_{nI} removes the effect of generalized nested models (with the benchmark model) on its limiting distribution. By a minor modification of the proof of Theorem 7 in Chen and Fan (2005), we immediately obtain the following result:

Proposition 4.4 *Suppose all conditions of Proposition 4.2 are satisfied. If $b \rightarrow 0$ and $nb \rightarrow \infty$, then under the null hypothesis H_0 , the limiting distribution of T_{nI} is given by that of $\max_{i \in S_{NB}} (Z_i/\sqrt{\sigma_{ii}}, 0)$, where*

$$S_{NB} = \left\{ \begin{array}{l} i \in \{2, \dots, M\} : \sigma_{ii} > 0 \text{ and} \\ n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] = 0 \end{array} \right\}.$$

Proposition 4.4 implies that the asymptotic null distribution of T_{nI} depends on models that are generalized non-nested with the benchmark and satisfy

$$n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] = 0,$$

and hence is unknown. We propose the following bootstrap procedure to approximate the asymptotic null distribution of T_{nI} :

Step 1. Generate a bootstrap sample by random draws with replacement from a consistent nonparametric estimator of the unknown joint distribution of (X_{1t}, X_{2t}) that takes into account the censoring scheme. Denote $(\tilde{F}_1^*, \tilde{F}_2^*, \hat{\alpha}_{in}^*, \hat{\alpha}_{1n}^*)$ as the bootstrap analogs of $(\tilde{F}_1, \tilde{F}_2, \hat{\alpha}_{in}, \hat{\alpha}_{1n})$.

Step 2. Compute the bootstrap value $T_{in}^* \equiv LR_n(\tilde{F}_1^*, \tilde{F}_2^*; \hat{\alpha}_{in}^*, \hat{\alpha}_{1n}^*)$ of $T_{in} \equiv LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})$, $i = 2, \dots, M$, and define its recentered value as $T_{inC}^* = T_{in}^* - T_{in}I(T_{in} \geq -a_n)$, where $a_n \rightarrow 0$ is a small positive (possibly random) number such that $\sqrt{n}a_n \rightarrow \infty$.

Step 3. Compute the bootstrap value of T_{nI} as

$$T_{nI}^* = \max_{2 \leq i \leq M} \left\{ \frac{\sqrt{n} T_{inC}^*}{\sqrt{\hat{\sigma}_{ii}}} G_b(\hat{\sigma}_{ii}), 0 \right\};$$

Step 4. Repeat Steps 1–3 for a large number of times and use the empirical distribution function of the resulting values T_{nI}^* to approximate the null distribution of T_{nI} .

We note that the above bootstrap procedure is very similar to that proposed in Chen and Fan (2005), except that in Step 1 we generate bootstrap samples from a consistent nonparametric estimator of the joint distribution that takes account of the censoring. For example, for bivariate random right censoring, we could sample from the bivariate Kaplan-Meier estimator; see Dabrowska (1989). See Davison and Kinkley (1997, page 85) for additional ways to generate bootstrap sample for censored data. The consistency of this standardized bootstrap RC test T_{nI}^b could be established by a minor modification of the proof of Theorem 8 in Chen and Fan (2005).

Remark 4.2: Recall that

$$PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) = LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - \frac{Pen(p_i, n) - Pen(p_1, n)}{n}.$$

If $\max_{i=2, \dots, M} [Pen(p_i, n) - Pen(p_1, n)]/\sqrt{n} \rightarrow 0$ (which is automatically satisfied with AIC and BIC), then

$$PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) = LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) + o_p(n^{-1/2}) \text{ for } i = 2, \dots, M.$$

Therefore, penalization could be incorporated in the tests. Define

$$T_n^P = \max_{i=2, \dots, M} [n^{1/2} PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})],$$

$$T_{nS}^P = \max_{i=2, \dots, M} \left\{ \frac{n^{1/2} PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})}{\sqrt{\hat{\sigma}_{ii}}} G_b(\hat{\sigma}_{ii}) \right\},$$

and

$$T_{nI}^P = \max \left[\max_{i=2, \dots, M} \left\{ \frac{n^{1/2} PLR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})}{\sqrt{\hat{\sigma}_{ii}}} G_b(\hat{\sigma}_{ii}) \right\}, 0 \right]. \quad (4.2)$$

Then we can conduct the test using T_n^P (or T_{nS}^P or T_{nI}^P) instead of T_n (or T_{nS} or T_{nI}).

5 An empirical application

In this section, we illustrate our testing procedure for selection of multiple copula-based survival functions by using insurance company data on losses and ALAEs. The particular data set we use were collected by the US Insurance Services Office and have been analyzed in some detail in Frees and Valdez (1998), Klugman and Parsa (1999), and Denuit et al. (2004).

Two alternative approaches have been used in the literature to model multivariate survival data; that of multivariate distribution function and that of multivariate survival function. It is important to realize that in the context of semiparametric copula-based models, the copula in a semiparametric copula-based distribution function corresponds to its survival copula in the corresponding semiparametric survival function. To be specific, consider bivariate case. Let (X_1, X_2) be the survival variables of interest with a joint survival function $F^o(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2)$ and marginal survival functions $F_j^o(\cdot)$, $j = 1, 2$. Let $H(x_1, x_2)$ denote the corresponding joint cumulative distribution function (cdf) with marginal distributions $H_j(\cdot)$, $j = 1, 2$. Assume that $H_1 \equiv 1 - F_1^o$ and $H_2 \equiv 1 - F_2^o$ are continuous. By the Sklar's (1959) theorem, there exists a unique copula function C_h such that $H(x_1, x_2) \equiv C_h(H_1(x_1), H_2(x_2))$, which in turn implies the representation

$$F^o(x_1, x_2) \equiv C^o(F_1^o(x_1), F_2^o(x_2)),$$

holds where

$$C^o(u_1, u_2) \equiv u_1 + u_2 - 1 + C_h(1 - u_1, 1 - u_2)$$

is itself a copula function, known as a survival copula. Hence the bivariate distribution function $C_h(H_1(x_1), H_2(x_2))$ and the bivariate survival function $C^o(F_1^o(x_1), F_2^o(x_2))$, where $F_j^o(\cdot)$ is the survival function of $H_j(\cdot)$ and C^o is the survival copula of C_h , represent the same model.

In Frees and Valdez (1998) and Klugman and Parsa (1999), fully parametric modelling of the joint distribution of the loss and ALAE has been examined; using various model selection techniques including AIC/BIC, Frees and Valdez (1998) select Pareto marginals and Gumbel copula, while Klugman and Parsa (1999) select inverse paralogistic for loss, inverse Burr for ALAE and Frank copula. Denuit et al. (2004) adopt a semiparametric distribution framework in which the marginal distributions of loss and ALAE are left unspecified, but their copula is modelled parametrically via a one-parameter Archimedean copula. Their model selection procedure is the same as that in Wang and Wells (2000) except that the joint distributions of loss and ALAE are estimated differently. They examined four one-parameter Archimedean copulas: Gumbel, Clayton, Frank and Joe, and select the same Gumbel copula as Frees and Valdez (1998). Compared with Denuit et al. (2004), we do not restrict the parametric copulas to be Archimedean. In addition, our test takes into account the randomness of the selection criterion. Chen and Fan (2005) have also studied this data set, but since their model selection test is applicable to uncensored data only, they restrict their analysis to the subset of 1466 complete data. We now apply our proposed test to the original censored data with 1500 data points.

The scatterplots for loss and ALAE presented in Frees and Valdez (1998) and Denuit et al. (2004) reveal positive right tail dependence between loss and ALAE: large losses tend to be associated with large ALAE's. This is because expensive claims generally need some time to be settled and induce considerable costs for the insurance company. Actuaries therefore expect positive dependence between large losses and large ALAE's. On the other hand, these plots do not reveal any visible left tail dependence between the two variables. As a result, it is not surprising that Gumbel copula is chosen in Frees and Valdez (1998) and Denuit et al. (2004). To shed some light on the robustness of this result to the set of copula families being considered, we add three more copula families to the set considered in Denuit et al. (2004): Gaussian copula, survival Clayton, mixture of Clayton and Gumbel copulas; see Appendix B for expressions of these seven copulas and their partial derivatives. Survival Clayton has right tail dependence and the mixture of Clayton and Gumbel exhibits both left tail and right tail dependence unless the weights are degenerate. Gaussian copula does not have tail dependence and is thus expected to fit poorly. They are included here in the set of copulas to see if the power of the test is adversely affected by the presence of poor copula candidates

in the selection set.³

To facilitate comparison, we also apply our tests to the subset of 1466 complete data. The results of the ‘‘Monte Carlo RC’’ test T_n^P (using AIC penalization factor) for the original censored data are presented in Table 1 and those for the subset of 1466 complete data are presented in Table 2, with 500,000 number of Monte Carlo repetitions. For each copula, we estimated its parameter(s) by the two-step procedure and computed the value of AIC. To apply our model selection test we need to choose a benchmark model. In view of the existing results, we first use Gumbel copula as the benchmark. For the Gumbel benchmark, we found the p-value of the test to be 1 with or without taking into account censoring. This provides strong evidence that none of the other six copulas performs significantly better than the Gumbel copula for the loss-ALAE data. This is consistent with the selection result based on comparing the values of AIC only; Gumbel followed by mixture of Clayton and Gumbel, then by survival Clayton and then by Joe. The parameter estimates for the mixture of Clayton and Gumbel provide additional evidence in favor of the Gumbel copula; the estimates of the weight on Clayton are only 0.0003 when censoring is taken into account and 0.0002 when censoring is not taken into account. In addition, the estimates of the parameter in the Gumbel copula obtained by fitting the mixture of Clayton and Gumbel are very close to the estimates obtained by fitting the Gumbel copula alone for both the subset of complete data and the original censored data. To see if the test is sensitive to the choice of the benchmark model, we also used each of the remaining six copulas as the benchmark.

For each of the Tables 1 and 2, we present two versions of the Monte Carlo tests based on the non-standardized test, T_n^P , and the standardized test, T_{nS}^P , as described in Remark 4.2.⁴ Comparing the first two columns in Tables 1 and 2, we see that both tests yield similar high p-values when the benchmark is either Gumbel or the mixture of Clayton and Gumbel; for all the other cases, the standardized test T_{nS}^P yields significantly lower p-values than those of T_n^P . This indicates that the standardized version of the test is generally more powerful than the original non-standardized test.

Additionally, we present a bootstrap version of the test based on T_{nI}^P (using AIC penalization factor). We generate bootstrap sample by random draws with replacement from

³Since our test is developed for semiparametric copula-based survival functions instead of distribution functions, we use the survival copulas of these seven copula functions in implementing our test. However, we present our empirical results in terms of copulas of the corresponding semiparametric distribution functions in order to compare our results with existing results just cited.

⁴When computing the test statistic T_{nS}^P , we have used $a = 1$ and $b_n = 10/n^2$.

a consistent nonparametric estimator of the bivariate joint distribution that takes into account the censoring scheme. For this loss-ALAE data set, we could draw bootstrap samples either from the bivariate Kaplan-Meier estimator of Dabrowska (1989), or from the estimator of Akritas (1994) and Denuit et al. (2004). Let $T_{nI}^{*,P}$ be the counterpart of T_{nI}^* for one bootstrap iteration, we write the re-centered bootstrap test statistic as $T_{nIC}^{*,P} = T_{nI}^{*,P} - T_{nI}^P \times I\{T_{nI}^P \geq -a_n\}$, where for simplicity we use the same parameter values $(a, b_n, a_n) = (1, n^{-1/2}, 0.025n^{-1/2} \log \log n)$ as those in Chen and Fan (2005). In this empirical application we use 100 bootstrap repetitions. The bootstrap p-values in Tables 3 and 4 overwhelmingly support the conclusion that the Gumbel copula fits the loss-ALAE data the best among the seven copulas we considered. This finding is consistent with existing results in the literature. The fact that the results in Tables 3 and 4 are so close to each other confirms the statement in Denuit et al. (2004) that the limited amount of censored points present in this Loss-ALAE data does not seem to affect the copula selection result.

Finally, by comparing the bootstrap p-values in Tables 3 and 4 with the Monte Carlo p-values in Tables 1 and 2, we notice that the standardized “bootstrap RC” test is in general more powerful than the standardized “Monte Carlo RC” test, which in turn is more powerful than the non-standardized “Monte Carlo RC” test. Nevertheless, it is noteworthy that the standardized “bootstrap RC” test is computationally much more intensive than the standardized “Monte Carlo RC” test. For an AMD Athlon(tm) 64 Processor, 1.18 GHz and 384 Mb of RAM, for each benchmark case, the standardized “bootstrap RC” test (with 100 bootstrap replications) takes about 10500 computer seconds, whereas the standardized “Monte Carlo RC” test (with 500,000 Monte Carlo repetitions) only takes about 350 computer seconds. Moreover, we are happy to see that the standardized “Monte Carlo RC” test and the standardized “bootstrap RC” test yield very similar rankings and lead to the same conclusion that the Gumbel copula fits the loss-ALAE data the best.

Benchmark	p-value of T_n^P	p-value of T_{nS}^P	AIC	2-step Estimator
Gumbel	1.0000	0.9980	-0.1447	1.4428
Clayton	0.0015	0.0004	-0.0000	0.5152
Frank	0.0688	0.0394	-0.1009	0.0473
Joe	0.3968	0.2533	-0.1263	1.6466
Gaussian	0.1692	0.0724	-0.1125	0.4668
Survival Clayton	0.6295	0.4298	-0.1380	0.7825
Mix Clayton & Gumbel	0.9469	0.9794	-0.1420	(0.1505,1.4433,0.0003)

Table 1: Monte Carlo p-values of the test for the original dataset subject to censoring

Benchmark	p-value of T_n^P	p-value of T_{nS}^P	AIC	2-step Estimator
Gumbel	1.0000	0.9940	-0.2560	1.4254
Clayton	0.0037	0.0008	-0.1203	0.5098
Frank	0.1197	0.0834	-0.2160	0.0494
Joe	0.3530	0.1643	-0.2384	1.6105
Gaussian	0.2499	0.1442	-0.2286	0.4604
Survival Clayton	0.5570	0.3412	-0.2472	0.7440
Mix Clayton & Gumbel	0.9382	0.9590	-0.2530	(0.1572,1.4256,0.0002)

Table 2: Monte Carlo p-values of the test for the subset without censoring

Benchmark	p-value of T_{nI}^P	AIC	Two-step estimate
Gumbel	1.0000	-0.1447	1.4428
Clayton	0.0000	-0.0000	0.5152
Frank	0.0000	-0.1009	0.0473
Joe	0.1010	-0.1263	1.6466
Gaussian	0.0517	-0.1125	0.4668
Survival Clayton	0.1414	-0.1380	0.7825
Mix Clayton & Gumbel	0.9900	-0.1420	(0.1505,1.4433,0.0003)

Table 3: Bootstrap p-values of the test for the original dataset subject to censoring

Benchmark	p-value of T_{nI}^P	AIC	Two-step estimate
Gumbel	1.0000	-0.2560	1.4254
Clayton	0.0000	-0.1203	0.5098
Frank	0.0000	-0.2160	0.0494
Joe	0.1052	-0.2384	1.6105
Gaussian	0.0202	-0.2286	0.4604
Survival Clayton	0.0909	-0.2472	0.7440
Mix Clayton & Gumbel	0.9963	-0.2530	(0.1572,1.4256,0.0002)

Table 4: Bootstrap p-values of the test for the subset without censoring

6 Conclusion

Many models of semiparametric multivariate survival functions are characterized by nonparametric marginal survival functions and parametric copula functions, where different copulas imply different dependence structures. In this paper, we first establish large sample properties of the two-step estimator of copula dependence parameter when the parametric copula function may be misspecified and when data may be subject to an independent but otherwise general right censorship. We then provide a penalized pseudo-likelihood ratio test for

selecting among multiple semiparametric copula models for multivariate survival data. An empirical application to the famous Loss-ALAE insurance data set indicates the usefulness of our theoretical results.

Although our theoretical results allow for general right censoring scheme, we still assume that the data is independent and is subject to independent censoring. In some economic and financial applications, data could be serially dependent and may be subject to dependent censorship. The two-step estimator and its large sample properties have been extended to time series settings in Chen and Fan (2006a, 2006b), but their results do not allow for any censoring. We shall extend the results in this paper to allow for time series and/or dependent censoring in another paper.

Appendix A. Technical Proofs

We first introduce additional notation: $N_{jt}(x) = \delta_{jt}I(\tilde{X}_{jt} \leq x)$, $J_{jt}(u) = I(\tilde{X}_{jt} \geq u)$, $M_{jt}(x) = N_{jt}(x) - \int_{-\infty}^x J_{jt}(u)d\Lambda_j(u)$ and $\Lambda_j(u) = -\log F_j^o(u)$ the marginal cumulative hazard function of X_j , $j = 1, 2$.

Lemma A.1 *Suppose that Conditions C1 and C5 are satisfied. Then: (i) the marginal Kaplan-Meier estimators are uniformly strongly consistent: $\sup_{x \leq \tau_j} |\tilde{F}_j(x) - F_j^o(x)| \rightarrow 0$ a.s. for $j = 1, 2$; (ii) they can be expressed as martingale integrals:*

$$\begin{aligned} \tilde{F}_j(x) - F_j^o(x) &= -F_j^o(x) \int_{-\infty}^x \frac{\tilde{F}_j(u-)}{F_j^o(u)} \frac{\sum_{t=1}^n dM_{jt}(u)}{\sum_{t=1}^n I(\tilde{X}_{jt} \geq u)} \\ &= -F_j^o(x) \int_{-\infty}^x \frac{\sum_{t=1}^n dM_{jt}(u)}{F_j^o(u) \sum_{t=1}^n G_{jt}(u)} + o_p(n^{-1/2}), \end{aligned}$$

where $o_p(\cdot)$ is uniform in $x \in [0, \tau_j]$, for $j = 1, 2$.

Proof of Lemma A.1. Because of Condition C5, the risk set size in $(-\infty, \tau_j]$ is of order n . Consequently, the uniform strong consistency is a special case of Theorem 3 of Lai and Ying (1991). The martingale integral approximation follows from formula (3.2.13) of Gill (1980) and the consistency of the Kaplan-Meier estimator.

Lemma A.2 *Let $\hat{x}_j = \inf\{x : \tilde{F}_j(x) < 1\}$, $j = 1, 2$. There exists $\tau_0 > 0$ such that for every $\epsilon > 0$, there is an $\eta > 0$ such that*

$$\liminf_{n \rightarrow \infty} P \left(\inf_{\hat{x}_j \leq x \leq \tau_0} \frac{1 - \tilde{F}_j(x)}{1 - F_j^o(x)} \geq \eta \right) > 1 - \epsilon, \quad j = 1, 2.$$

Proof of Lemma A.2. For notational convenience, subscript $j = 1, 2$ will be omitted. By definition,

$$\tilde{F}(x) = \prod_{t: \tilde{X}_t \leq x} \left(1 - \frac{\delta_t}{\sum_{k=1}^n J_k(\tilde{X}_t)} \right) \leq \exp \left\{ - \int_{-\infty}^x \frac{\sum_{k=1}^n dN_k(u)}{\sum_{k=1}^n J_k(u)} \right\}.$$

The right-hand side is bounded by $1 - \frac{2}{3} \int_{-\infty}^x \frac{\sum_{k=1}^n dN_k(u)}{\sum_{k=1}^n J_k(u)}$, $x \leq \tau_0$ for suitably chosen τ_0 , provided that $\int_{-\infty}^{\tau_0} \frac{\sum_{k=1}^n dN_k(u)}{\sum_{k=1}^n J_k(u)} < -\log(2/3)$, which holds for all large n . Thus,

$$1 - \tilde{F}(x) + \frac{1}{n} \geq \frac{2}{3} \left\{ \sum_{t=1}^n I(C_t \geq \tau_0) I(X_t \leq x) + \frac{1}{n} \right\}. \quad (\text{A.1})$$

By a theorem of van Zuijlen (1978, Theorem 1.1), for any $\epsilon > 0$, there exists η such that

$$P \left\{ \sum_{t=1}^n I(C_t \geq \tau_0) I(X_t \leq x) + \frac{1}{n} \geq \eta \sum_{t=1}^n I(C_t \geq \tau_0) F^o(x) \right\} > 1 - \epsilon. \quad (\text{A.2})$$

Since $\liminf n^{-1} \sum_{t=1}^n I(C_t \geq \tau_0) > 0$, it follows from (A.1), (A.2) and the fact that $1 - \tilde{F}(x) \geq n^{-1}$ for all $x \geq \hat{x}$ that the lemma holds.

Proof of Proposition 3.1. The main ideas here are to use the uniform consistency of the Kaplan-Meier estimator and the identifiability Condition C2. Write

$$\begin{aligned} & n^{-1} \sum_{t=1}^n \{ \ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) - E^0[\ell(F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}); \alpha)] \} \\ = & n^{-1} \sum_{t=1}^n \{ \ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) - \ell(U_{1t}, U_{2t}; \alpha) \} \\ & + n^{-1} \sum_{t=1}^n \{ \ell(U_{1t}, U_{2t}; \alpha) - E^0[\ell(U_{1t}, U_{2t}; \alpha)] \}. \end{aligned} \quad (\text{A.3})$$

We first show that the first term on the right-hand side of (A.3) is of order $o_p(1)$, uniformly in $\alpha \in \mathcal{A}$. Under Condition C5, $\tilde{F}_j(x) \geq \tilde{F}_j(\tau_j)$, $j = 1, 2$, are bounded away from 0. By continuity of $\ell(\cdot)$ on $(0, 1) \times (0, 1) \times \mathcal{A}$ and Lemma A.1, the first term, with summation over t such that both $\tilde{F}_1(\tilde{X}_{1t})$ and $\tilde{F}_2(\tilde{X}_{2t})$ are bounded away from 0, is of order $o_p(1)$, uniformly in $\alpha \in \mathcal{A}$. i.e., for every $\eta > 0$,

$$\lim_{n \rightarrow \infty} \sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n |\ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) - \ell(U_{1t}, U_{2t}; \alpha)| I(\tilde{X}_{1t} \wedge \tilde{X}_{2t} \geq \eta) = 0. \quad (\text{A.4})$$

It remains to show that for every $\epsilon > 0$, there exists $\eta > 0$ such that

$$P \left\{ \sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n |\ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) I(\tilde{X}_{jt} \leq \eta)| \geq \epsilon \right\} \leq \epsilon, \quad j = 1, 2, \quad (\text{A.5})$$

and

$$P \left\{ \sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n |\ell(U_{1t}, U_{2t}; \alpha) I(\tilde{X}_{jt} \leq \eta)| \geq \epsilon \right\} \leq \epsilon, \quad j = 1, 2. \quad (\text{A.6})$$

By Lemma A.2 and Condition C4(iii), there exists $K > 0$ such that

$$P \left\{ \sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n |\ell(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) I(\tilde{X}_{jt} \leq \eta)| > \frac{\epsilon}{3}, \right. \\ \left. K \sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n |\ell(U_{1t}, U_{2t}; \alpha) I(\tilde{X}_{jt} \leq \eta)| \right\} < \frac{\epsilon}{3}.$$

Therefore, to show (A.5) and (A.6), it suffices to show that for any $\epsilon^* > 0$, there exists η such that

$$P \left\{ |n^{-1} \sum_{t=1}^n L_t I(\tilde{X}_{jt} \leq \eta)| \geq \epsilon^* \right\} \leq \frac{2}{3} \epsilon, \quad j = 1, 2. \quad (\text{A.7})$$

By Condition C4(ii) and the Markov inequality, to show (A.7), we only need to show that for any $K^* > 0$, there exists η such that

$$P\{|n^{-1} \sum_{t=1}^n L_t I(L_t < K^*) I(\tilde{X}_{jt} \leq \eta)| \geq \epsilon^*\} \leq \frac{1}{3}\epsilon, \quad j = 1, 2. \quad (\text{A.8})$$

But, again by the Markov inequality, the left-hand side of (A.8) is bounded by

$$\frac{K^*}{\epsilon^*} n^{-1} \sum_{t=1}^n P\{\tilde{X}_{jt} \leq \eta\}$$

which can be made arbitrarily small by Condition C1.

We next show that the second term is also of order $o_p(1)$. By Condition C4(ii), it suffices to show that for every $K > 0$,

$$n^{-1} \sum_{t=1}^n \left\{ \ell(U_{1t}, U_{2t}; \alpha) I(\max\{L_t, L_{\alpha t}\} \leq K) - E^0[\ell(U_{1t}, U_{2t}; \alpha) I(\max\{L_t, L_{\alpha t}\} \leq K)] \right\}$$

converges to 0 uniformly in $\alpha \in \mathcal{A}$. But this sequence converges to 0 a.s. for every α and has uniformly bounded derivatives over the compact set \mathcal{A} , and, therefore, the convergence must be uniform.

Proof of Proposition 3.2. The proof can be done by essentially combining the techniques of Shih and Louis (1995) and Chen and Fan (2005). A critical part is how to appropriately control the tail behavior.

By the mean-value theorem, we can linearly expand the pseudo-likelihood score function at α_n^* to get

$$\hat{\alpha}_n - \alpha_n^* = \tilde{B}_n^{-1} \frac{1}{n} \sum_{t=1}^n \ell_{\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha_n^*), \quad (\text{A.9})$$

where $\tilde{B}_n = \frac{1}{n} \sum_{t=1}^n \ell_{\alpha\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \tilde{\alpha}_n)$ for some $\tilde{\alpha}_n$ on the line segment between α_n^* and $\hat{\alpha}_n$. Under Condition A4, we can apply the same argument for proving (A.5) to show that $\sup_{\alpha \in \mathcal{A}} n^{-1} \sum_{t=1}^n |\ell_{\alpha\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha) I(\tilde{X}_{jt} \leq \eta)|$ is asymptotically negligible as $\eta \rightarrow 0$. This in conjunction with Condition A2 and the consistency of \tilde{F}_j and $\hat{\alpha}_n$, implies that $\tilde{B}_n B_n^{-1} \rightarrow I_p$ in probability as $n \rightarrow \infty$.

Again by the mean-value theorem,

$$\begin{aligned} \sum_{t=1}^n \ell_{\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha_n^*) &= \sum_{t=1}^n \ell_{\alpha}(F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}); \alpha_n^*) \\ &+ \sum_{j=1}^2 \sum_{t=1}^n \ell_{\alpha j}(\tilde{U}_{1t}, \tilde{U}_{2t}; \alpha_n^*) \{ \tilde{F}_j(\tilde{X}_{jt}) - F_j^o(\tilde{X}_{jt}) \} = D_{1n} + D_{2n}, \end{aligned} \quad (\text{A.10})$$

where $(\tilde{U}_{1t}, \tilde{U}_{2t})$ lies on the line segment between $(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}))$ and $(F_1^o(\tilde{X}_{1t}), F_2^o(\tilde{X}_{2t}))$.

By Lemma A.1,

$$D_{2n} = - \sum_{j=1}^2 \sum_{t=1}^n \ell_{\alpha_j}(\tilde{U}_{1t}, \tilde{U}_{2t}; \alpha_n^*) F_j^o(\tilde{X}_{jt}) \int_{-\infty}^{\tilde{X}_{jt}} \frac{\tilde{F}_j(u-)}{F_j^o(u-)} \frac{\sum_s dM_{js}(u)}{\sum_s I(\tilde{X}_{js} \geq u)}. \quad (\text{A.11})$$

Let $D_{2n}(\eta) = \sum_{j=1}^2 D_{2n,j}(\eta)$ denote the right-hand side of (A.11) with the summation restricted to those terms such that $\tilde{X}_{jt} \leq \eta$. We next show that for some $\epsilon_j > 0$,

$$|n^{-1/2} D_{2n,j}(\eta)| = O_p(1)(1 - F_j^o(\eta))^{\epsilon_j}, \quad j = 1, 2, \quad (\text{A.12})$$

where $O_p(1)$ is uniform over $\eta > 0$. For any $\xi \in (0, 1)$, since

$$\begin{aligned} & E^0 \left\{ (1 - F_j^o(x))^{-\xi/2} \int_{-\infty}^x \frac{\tilde{F}_j(u-)}{F_j^o(u-)} \frac{\sum_s dM_{js}(u)}{\sum_s I(\tilde{X}_{js} \geq u)} \right\}^2 \\ & \leq E^0 \left\{ \int_{-\infty}^x \left\{ \frac{\tilde{F}_j(u-)}{F_j^o(u-)} \right\}^2 \frac{\{1 - F^o(u)\}^{-\xi}}{\sum_s I(\tilde{X}_{js} \geq u)} I(\max_t \tilde{X}_{jt} \geq u) d\Lambda_j^o(u) \right\}, \end{aligned}$$

it follows from Lengart's inequality (Gill, 1980, Theorem 2.4.2) that

$$\begin{aligned} & E^0 \left\{ (1 - F_j^o(x))^{-\xi/2} \int_{-\infty}^x \frac{\tilde{F}_j(u-)}{F_j^o(u-)} \frac{\sum_s dM_{js}(u)}{\sum_s I(\tilde{X}_{js} \geq u)} \right\}^2 \\ & = O_p(1) \int_{-\infty}^x \left\{ \frac{\tilde{F}_j(u-)}{F_j^o(u-)} \right\}^2 \frac{\{1 - F^o(u)\}^{-\xi}}{\sum_s I(\tilde{X}_{js} \geq u)} I(\max_t \tilde{X}_{jt} \geq u) d\Lambda_j^o(u) \\ & = O_p(1) n^{-1} \int_{-\infty}^x \{1 - F^o(u)\}^{-\xi} d\{1 - F^o(u)\} \\ & = O_p(1) n^{-1} \{1 - F^o(x)\}^{1-\xi}, \end{aligned} \quad (\text{A.13})$$

where $O_p(1)$ is uniform in x and the second equality follows from Lemma A.2 and van Zuijlen (1978, Theorem 1.1). From (A.13), Lemma A.2 (with $\xi = 2\xi_j$) and Condition A3, we have, ignoring the right tail,

$$|n^{-1/2} D_{2n,j}(\eta)| = O_p(1) n^{-1} \sum_{t=1}^n U_{jt}^{\xi_j - b_j} U_{kt}^{-a_k} (1 - F_j^o(\eta))^{1-2\xi_j} = O_p(1) (1 - F_j^o(\eta))^{1-2\xi_j}.$$

Hence, (A.12) holds with $\epsilon_j = 1 - 2\xi_j$, $j = 1, 2$.

In view of (A.12), we can essentially pretend that ℓ_{α_j} in (A.10) does not blow up at the tail. Therefore, (A.11) implies that for $j = 1, 2$,

$$\begin{aligned} D_{2n,j} & = - \sum_{s=1}^n \int \frac{n^{-1} \sum_{t=1}^n \ell_{\alpha_j}(\tilde{U}_{1t}, \tilde{U}_{2t}; \alpha_n^*) U_{jt} I(\tilde{X}_{jt} \geq u)}{n^{-1} \sum_{t=1}^n I(\tilde{X}_{jt} \geq u)} dM_{js}(u) \\ & = - \sum_{s=1}^n \int \frac{n^{-1} \sum_{t=1}^n E^0 \{ \ell_{\alpha_j}(U_{1t}, U_{2t}; \alpha_n^*) U_{jt} I(\tilde{X}_{jt} \geq u) \}}{P_{nj}(u)} dM_{js}(u) + o_p(n^{1/2}) \end{aligned} \quad (\text{A.14})$$

From (A.9), (A.10), (A.11) and (A.14), we see that $\hat{\alpha}_n - \alpha_n^*$ is asymptotically a sum of independent zero-mean random vectors. Given Condition A1, Proposition 3.2 now follows from the standard multivariate central limit theorem for independent but non-identically distributed random variables.

Proof of Propositions 3.3. The consistency of the variance estimator clearly follows from the laws of large numbers, the consistency of the Kaplan-Meier estimator and of $\hat{\alpha}_n$, when the possible “tail instability” is ignored. To control the tail behavior, we can apply the same techniques as in the proofs of Propositions 3.1 and 3.2. The details are omitted.

Proof of Proposition 4.2. For $i = 1, \dots, M$, by the definition of $\hat{\alpha}_{in}$, we have

$$\sum_{t=1}^n \ell_{i,\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{in}) = 0.$$

Hence,

$$\begin{aligned} \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha_{in}^*) &= \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{in}) \\ &\quad + \frac{1}{2}(\alpha_{in}^* - \hat{\alpha}_{in})' \sum_{t=1}^n \ell_{i,\alpha\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \bar{\alpha}_{in})(\alpha_{in}^* - \hat{\alpha}_{in}), \end{aligned}$$

where $\bar{\alpha}_{in}$ is between α_{in}^* and $\hat{\alpha}_{in}$. By conditions C2–C5, A1–A4 and Proposition 3.2, we have

$$\begin{aligned} \frac{1}{2n}(\alpha_{in}^* - \hat{\alpha}_{in})' \sum_{t=1}^n \ell_{i,\alpha\alpha}(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \bar{\alpha}_{in})(\alpha_{in}^* - \hat{\alpha}_{in}) \\ = -\frac{1}{2}(\alpha_{in}^* - \hat{\alpha}_{in})' B_{in}(\alpha_{in}^* - \hat{\alpha}_{in}) + o_p(1/n). \end{aligned}$$

Hence,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \hat{\alpha}_{in}) \\ = \frac{1}{n} \sum_{t=1}^n \ell_i(\tilde{F}_1(\tilde{X}_{1t}), \tilde{F}_2(\tilde{X}_{2t}); \alpha_{in}^*) + \frac{1}{2}(\alpha_{in}^* - \hat{\alpha}_{in})' B_{in}(\alpha_{in}^* - \hat{\alpha}_{in}) + o_p(1/n). \end{aligned}$$

As a result, we get for all $i = 2, \dots, M$,

$$LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - \frac{1}{n} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] = A_{i,n} + D_{i,n} + o_p(1/n),$$

where

$$A_{i,n} \equiv LR_n(\tilde{F}_1, \tilde{F}_2; \alpha_{in}^*, \alpha_{1n}^*) - \frac{1}{n} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)]$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{t=1}^n \left\{ [\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] - E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right\} \\
&\quad + \sum_{j=1}^2 \frac{1}{n} \sum_{t=1}^n \left\{ [\ell_{i,j}(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_{1,j}(U_{1t}, U_{2t}; \alpha_{1n}^*)] \{ \tilde{F}_j(\tilde{X}_{jt}) - F_j^o(\tilde{X}_{jt}) \} \right\} + o_p(1/\sqrt{n}),
\end{aligned}$$

$$D_n \equiv \frac{1}{2}(\alpha_{in}^* - \hat{\alpha}_{in})' B_{in}(\alpha_{in}^* - \hat{\alpha}_{in}) - \frac{1}{2}(\alpha_{1n}^* - \hat{\alpha}_{1n})' B_{1n}(\alpha_{1n}^* - \hat{\alpha}_{1n}).$$

By Proposition 3.2, we have $D_n = O_p(n^{-1})$.

For generalized non-nested models, Using the proof similar to that of Proposition 3.2, we obtain:

$$\sqrt{n} \times A_{i,n} = \frac{1}{\sqrt{n}} \sum_{t=1}^n [e_{it} - E^0(e_{it})] + o_p(1) = O_p(n^{-1/2}),$$

hence $\sqrt{n} \times A_{i,n}$ converges in distribution to a $\mathcal{N}(0, \sigma_{ii})$. Therefore,

$$\sqrt{n} \left[LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - \frac{1}{n} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right]$$

converges in distribution to a $\mathcal{N}(0, \sigma_{ii})$.

For generalized nested models, the term $A_{i,n}$ becomes zero almost surely, we have

$$LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - \frac{1}{n} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] = D_{i,n} + o_p(1/n),$$

where by Proposition 3.2, $2nD_{i,n}$ is distributed as a weighted sum of independent $\chi_{[1]}^2$ random variables.

Proof of Proposition 4.3. Note that

$$\begin{aligned}
T_n &= \max_{i=2, \dots, M} [n^{1/2} LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n})] \\
&= \max_{i=2, \dots, M} \left[n^{1/2} \left\{ LR_n(\tilde{F}_1, \tilde{F}_2; \hat{\alpha}_{in}, \hat{\alpha}_{1n}) - n^{-1} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right\} \right. \\
&\quad \left. + n^{-1/2} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right] \\
&\rightarrow \max_{i=2, \dots, M} \left[Z_i + \lim_{n \rightarrow \infty} \left\{ n^{-1/2} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right\} \right].
\end{aligned}$$

Let

$$K_{in} = \lim_{n \rightarrow \infty} \left\{ n^{-1/2} \sum_{t=1}^n E^0[\ell_i(U_{1t}, U_{2t}; \alpha_{in}^*) - \ell_1(U_{1t}, U_{2t}; \alpha_{1n}^*)] \right\}.$$

Then

$$\begin{aligned}
P(T_n > Z_\alpha) &\rightarrow P\left(\max_{i=2, \dots, M} \{Z_i + K_{in}\} > Z_\alpha \right) \\
&\geq P\left(\max_{i=2, \dots, M} Z_i + \max_{i=2, \dots, M} K_{in} > Z_\alpha \right).
\end{aligned}$$

For fixed alternatives, $\max_{i=2,\dots,M} K_{in} = +\infty$ and so $P(T_n > Z_\alpha) \rightarrow 1$. For local alternatives such that $\max_{i=2,\dots,M} K_{in} > 0$,

$$P\left(\max_{i=2,\dots,M} Z_i + \max_{i=2,\dots,M} K_{in} > Z_\alpha\right) > P\left(\max_{i=2,\dots,M} Z_i > Z_\alpha\right) = \alpha.$$

Hence $\lim_{n \rightarrow \infty} P(T_n > Z_\alpha) > \alpha$.

Appendix B. Expressions of Copulas and Their Derivatives

In the Appendix B we describe the seven copulas and their derivatives that we have used in the empirical application Section 5.⁵ Let (X_1, X_2) be the lifetime variables of interest with joint survival function $F^o(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2)$ and continuous marginal survival functions $F_j^o(\cdot)$, $j = 1, 2$. Let $H(x_1, x_2)$ denote the corresponding joint cumulative distribution function (cdf) with marginal distributions $H_j \equiv 1 - F_j^o$, $j = 1, 2$. By the Sklar's (1959) theorem, there exists a unique copula function C_h on $[0, 1]^2$ such that

$$H(x_1, x_2) \equiv C_h(H_1(x_1), H_2(x_2)),$$

or equivalently

$$F^o(x_1, x_2) \equiv C^o(F_1^o(x_1), F_2^o(x_2)),$$

holds with

$$C^o(u_1, u_2) \equiv u_1 + u_2 - 1 + C_h(1 - u_1, 1 - u_2), \quad (\text{B.1})$$

where the copula function $C^o()$ is sometimes called survival copula (of C_h).

It is easy to see that, for any $j \in \{1, 2\}$

$$\frac{\partial C^o}{\partial u_j}(u_1, u_2) = 1 - \frac{\partial C_h}{\partial u_j}(1 - u_1, 1 - u_2), \quad (\text{B.2})$$

in fact, for any partial derivative of order k higher than 2 we have that

$$\frac{\partial^k C^o}{\partial u_{j_1} \dots \partial u_{j_k}}(u_1, u_2) = (-1)^k \frac{\partial^k C_h}{\partial u_{j_1} \dots \partial u_{j_k}}(1 - u_1, 1 - u_2), \quad (\text{B.3})$$

where $j_i \in \{1, 2\}$. Note that this last equation implies that

$$c^o(u_1, u_2) = c_h(1 - u_1, 1 - u_2), \quad (\text{B.4})$$

where c^o and c_h are the copula densities associated to C^o and C_h , respectively.

Using relations (B.2), (B.3) and (B.4), by replacing $v_j = 1 - u_j$ in the expressions of partial derivatives of a copula C_h and its density c_h , we immediately obtain the expressions for the partial derivatives of the survival copula C^o and its density c^o . Therefore, in the

⁵In the empirical application we have used both analytical derivatives and numerical derivatives, while the results based on analytical derivatives perform slightly better. Since these analytical derivatives for copulas are tedious to compute, we include them in this Appendix B so that readers could use them in other applications as well.

following we only provide expressions for the partial derivatives of several copula functions C_h and their densities c_h that we have used in the empirical application.

Gumbel Copula. The Gumbel copula and its density are given by

$$C_h(v_1, v_2) = \exp\{-((-\log(v_1))^\alpha + (-\log(v_2))^\alpha)^{1/\alpha}\}, \quad (\text{B.5})$$

and

$$c_h(v_1, v_2) = \frac{1}{C_h} \frac{\partial C_h}{\partial v_1} \frac{\partial C_h}{\partial v_2} T_1, \quad \text{with } T_1 = ((\alpha - 1)(-\log(C_h))^{-1} + 1).$$

Following Frees and Valdez (1998), we can express the partial derivative of C_h with respect to v_j , $j = 1, 2$, as

$$\frac{\partial C_h}{\partial v_j}(v_1, v_2) = \frac{C_h(v_1, v_2)}{v_j} (\log(v_j) / \log(C_h(v_1, v_2)))^{\alpha-1}. \quad (\text{B.6})$$

Hence, a little algebra implies⁶

$$\begin{aligned} \frac{\partial^2 C_h}{\partial v_j^2} &= (\log(v_j) / \log(C_h))^{\alpha-1} \left(\frac{\partial C_h}{\partial v_j} - \frac{C_h}{v_j^2} \right) + (\alpha - 1) \frac{C_h}{\log(C_h) v_j^2} (\log(v_j) / \log(C_h))^{\alpha-2} \\ &\quad - (\alpha - 1) \frac{\log(v_j) (\log(v_j) / \log(C_h))^{\alpha-2}}{v_j \log(C_h)^2} \frac{\partial C_h}{\partial v_j}. \end{aligned}$$

The partial derivative of the copula density c_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial c_h}{\partial v_j} = -\frac{1}{C_h^2} \left(\frac{\partial C_h}{\partial v_j} \right)^2 \frac{\partial C_h}{\partial v_i} T_1 + \frac{1}{C_h} \frac{\partial^2 C_h}{\partial v_j^2} \frac{\partial C_h}{\partial v_i} T_1 + \frac{1}{C_h} \frac{\partial C_h}{\partial v_j} c_h T_1 + \frac{1}{C_h} \frac{\partial C_h}{\partial v_j} \frac{\partial C_h}{\partial v_i} \frac{\partial T_1}{\partial v_j},$$

where

$$\frac{\partial T_1}{\partial v_j} = \frac{(\alpha - 1)(-\log(C_h))^{-2}}{C_h} \frac{\partial C_h}{\partial v_j}.$$

Clayton Copula. The Clayton copula and its density are given by

$$C_h(v_1, v_2) = (v_1^{-\alpha} + v_2^{-\alpha} - 1)^{-1/\alpha},$$

and

$$c_h(v_1, v_2) = (1 + \alpha) v_1^{-(1+\alpha)} v_2^{-(1+\alpha)} (v_1^{-\alpha} + v_2^{-\alpha} - 1)^{-(1/\alpha+2)}.$$

Hence the second order partial derivative of C_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial^2 C_h}{\partial v_j^2} = (1 + \alpha) \left(\frac{1}{C_h} \left(\frac{\partial C_h}{\partial v_j} \right)^2 - \frac{1}{v_j} \frac{\partial C_h}{\partial v_j} \right),$$

⁶We leave the dependence on (v_1, v_2) implicit, to ease the notational burden.

where

$$\frac{\partial C_h}{\partial v_j} = v_j^{-(1+\alpha)}(v_1^{-\alpha} + v_2^{-\alpha} - 1)^{-1/\alpha-1}.$$

The first order partial derivative of the copula density c_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial c_h}{\partial v_j} = c_h \left(\frac{(1+2\alpha)}{C_h} \frac{\partial C_h}{\partial v_j} - (1+\alpha)/v_j \right).$$

Frank Copula. The Frank copula and its density are given by

$$C_h(v_1, v_2) = \frac{1}{\log(\alpha)} \log \left(1 - \frac{(1-\alpha^{v_1})(1-\alpha^{v_2})}{1-\alpha} \right),$$

and

$$c_h(v_1, v_2) = \log(\alpha^{-1}) \frac{\alpha^{v_1} \alpha^{v_2}}{1-\alpha} \left(1 - \frac{(1-\alpha^{v_1})(1-\alpha^{v_2})}{1-\alpha} \right)^{-2}.$$

After some algebra, the second order partial derivative of C_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial^2 C_h}{\partial v_j^2} = \log(\alpha) \left(\frac{\partial C_h}{\partial v_j} - \left(\frac{\partial C_h}{\partial v_j} \right)^2 \right)$$

where

$$\frac{\partial C_h}{\partial v_j} = \left(1 - \frac{(1-\alpha^{v_1})(1-\alpha^{v_2})}{1-\alpha} \right)^{-1} \frac{(1-\alpha^{v_i})\alpha^{v_j}}{1-\alpha},$$

and the first order derivative of the copula density c_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial c_h}{\partial v_j} = \log(\alpha) \left(1 - 2 \frac{\partial C_h}{\partial v_j} \right) c_h.$$

Joe Copula. The Joe copula and its density are given by

$$C_h(v_1, v_2) = 1 - (\bar{v}_1^\alpha + \bar{v}_2^\alpha - \bar{v}_1^\alpha \bar{v}_2^\alpha)^{1/\alpha}$$

and

$$c_h(v_1, v_2) = \bar{v}_1^{\alpha-1} \bar{v}_2^{\alpha-1} T_2^{1/\alpha-1} \left(\alpha - \frac{(1-\alpha)(1-\bar{v}_1^\alpha)(1-\bar{v}_2^\alpha)}{T_2} \right),$$

where $\bar{v}_j = 1 - v_j$ and $T_2 = \bar{v}_1^\alpha + \bar{v}_2^\alpha - \bar{v}_1^\alpha \bar{v}_2^\alpha$.

The second order partial derivative of the copula C_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial^2 C_h}{\partial v_j^2} = (1-\alpha) \bar{v}_j^{\alpha-1} (1-\bar{v}_i^\alpha) T_2^{1/\alpha-1} \left(\frac{\bar{v}_j^{\alpha-1} (1-\bar{v}_i^\alpha)}{T_2} - \frac{1}{\bar{v}_j} \right).$$

After some tedious algebra, the first order partial derivative of the copula density c_h with respect to v_j , $j = 1, 2$, is given by

$$\begin{aligned} \frac{\partial c_h}{\partial v_j} &= (\alpha - 1)\bar{v}_i^{\alpha-1}\bar{v}_j^{\alpha-2}T_2^{1/\alpha-1}(-1 + \bar{v}_j^\alpha(1 - \bar{v}_i^\alpha)T_2^{-1})(\alpha - \frac{(1 - \alpha)(1 - \bar{v}_1^\alpha)(1 - \bar{v}_2^\alpha)}{T_2}) \\ &+ (1 - \alpha)\alpha\bar{v}_i^{\alpha-1}\bar{v}_j^{2\alpha-2}(1 - \bar{v}_i^\alpha)T_2^{1/\alpha-3}(-T_2 - (1 - \bar{v}_j^\alpha)(1 - \bar{v}_i^\alpha)). \end{aligned}$$

Gaussian copula. The Gaussian copula and its density are given by

$$C_h(v_1, v_2) = \Phi_\alpha(\Phi^{-1}(v_1), \Phi^{-1}(v_2)),$$

where Φ_α is the bivariate standard normal distribution with correlation α , Φ is the scalar standard normal distribution, and

$$c_h(v_1, v_2) = \frac{\phi_\alpha(\Phi^{-1}(v_1), \Phi^{-1}(v_2))}{\phi(\Phi^{-1}(v_1))\phi(\Phi^{-1}(v_2))},$$

where ϕ is the density function of Φ , and ϕ_α is the density function of Φ_α .

The second order partial derivative of the copula C_h with respect to v_j , $j = 1, 2$, is given by

$$\begin{aligned} \frac{\partial^2 C_h}{\partial v_j^2} &= -\left[\int_{-\infty}^{\Phi^{-1}(v_i)} \frac{\Phi^{-1}(v_j) - \alpha s}{2\pi(1 - \alpha^2)^{3/2}} \exp\left(-\frac{1}{2} \frac{\Phi^{-1}(v_j)^2 - 2\alpha\Phi^{-1}(v_j)s + s^2}{1 - \alpha^2}\right) ds\right] \phi(\Phi(v_j)^{-1})^{-2} \\ &+ \frac{\partial C_h}{\partial v_j} \Phi(v_j)^{-1} \phi(\Phi(v_j)^{-1})^{-1}. \end{aligned}$$

The first order partial derivative of the copula density c_h with respect to v_j , $j = 1, 2$, is given by

$$\frac{\partial c_h}{\partial v_j} = (\Phi(v_j)^{-1} - \frac{\Phi(v_j)^{-1} - \Phi(v_i)^{-1}\alpha}{1 - \alpha^2}) \phi(\Phi(v_j)^{-1})^{-1} c_h.$$

Mixture copula. A mixture copula $C_h(v_1, v_2; \alpha)$, with its parameter $\alpha = (\alpha_1, \alpha_2, \lambda)$, is simply given by

$$C_h(v_1, v_2; \alpha) = \lambda C_h^1(v_1, v_2; \alpha_1) + (1 - \lambda) C_h^2(v_1, v_2; \alpha_2), \quad 0 \leq \lambda \leq 1,$$

where $C_h^1(v_1, v_2; \alpha_1)$ is one copula (such as Clayton copula in our application) with its parameter α_1 , and $C_h^2(v_1, v_2; \alpha_2)$ is another copula (such as Gumbel copula in our application) with its parameter α_2 . Then it is clear that the partial derivatives of C_h is simply the linear combination of the partial derivatives of the two copulas:

$$\frac{\partial^k C_h}{\partial v_j^k} = \lambda \frac{\partial^k C_h^1}{\partial v_j^k} + (1 - \lambda) \frac{\partial^k C_h^2}{\partial v_j^k}, \quad j = 1, 2.$$

References

- Akritas, M., 1994. Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics* 22, 1299-1327.
- Chen, X., Fan, Y., 2005. Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. *The Canadian Journal of Statistics* 33, 389-414.
- Chen, X., Fan, Y., 2006a. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* 135, 125-154.
- Chen, X., Fan, Y., 2006b. Semiparametric estimation of copula-based time series models 130, 307-335.
- Chen, X., Fan, Y., 2007. A model selection test for bivariate failure-time data. *Econometric Theory*, forthcoming.
- Chen, X., Y. Fan, Tsyrennikov, V., 2006. Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association* 101, 1229-1240.
- Dabrowska, D., 1989. Kaplan-Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *Journal of Multivariate Analysis* 29, 308-325.
- Davison, A.C., Hinkley, D.V., 1997. *Bootstrap Methods and Their Application*. Cambridge University Press.
- Denuit, M., Purcaru, O., Van Keilegom, I., 2004. Bivariate Archimedean copula modelling for loss-ALAE data in non-life insurance. Working paper, UCL.
- Frees, E., Valdez, E., 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2, 1-25.
- Genest, C., Ghoudi, K., Rivest, L., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82, 543-552.
- Genest, C., Werker, B., 2002. Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. *Proceedings of the*

- Conference on Distributions with Given Marginals and Statistical Modelling, C. M. Cuadras and J. A. Rodríguez Lallena (eds).
- Gill, R., 1980. Censoring and Stochastic Integrals. Math. Centre Tracts 124. Mathematisch Centrum, Amsterdam.
- Hansen, R. P. 2003. A test for superior predictive ability. Manuscript, Brown University.
- Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall/CRC, London.
- Klaassen, C., Wellner, J., 1997. Efficient estimation in the bivariate Normal copula model: Normal margins are least-favourable. *Bernoulli* 3, 55-77.
- Klugman, S., Parsa, R., 1999. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics* 24, 139-148.
- Lai, T., Ying, Z., 1991. Estimating a distribution function with truncated and censored data. *Annals of Statistics* 19, 417-442.
- Li, D., 2000. On default correlation: a copula function approach. *Journal of Fixed Income*, 43-54.
- Nelsen, R., 1999. An Introduction to Copulas. Springer, New York.
- Oakes, D., 1989. Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 84, 487-493.
- Oakes, D., 1994. Multivariate survival distributions. *Journal of Nonparametric Statistics* 3, 343-354.
- Romano, J. P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73, 1237-1282.
- Shih, J., Louis, T., 1995. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* 51, 1384-1399.
- Sin, C., White, H., 1996. Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics* 71, 207-225.

- Sklar, A., 1959. Fonctions de r'epartition 'a n dimensionset leurs marges. Publ. Inst. Statis. Univ. Paris 8, 229-231.
- van Zuijlen, M.C.A., 1978. Properties of the empirical distribution function for independent nonidentically distributed random variables. *Annals of Probability*, 6, 250-266.
- Wang, W., Wells, M., 2000. Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association* 95, 62-76.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. Cambridge University Press.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68, 1097-1126.