Journal of Multivariate Analysis 110 (2012) 171–188

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Nonparametric maximum likelihood estimation for dependent truncation data based on copulas

Takeshi Emura [1], Weijing Wang *

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

ARTICLE INFO

Article history:
Available online 9 April 2012

AMS 2000 subject classifications:
62N02
62G05
62G20

Keywords:
Archimedean copula
Lifetime data
Model selection
Nonparametric maximum likelihood
Truncation
Quasi-independence
Weak convergence

ABSTRACT

Truncation occurs when the variable of interest can be observed only if its value satisfies certain selection criteria. Most existing methods for analyzing such data critically rely on the assumption that the truncation variable is quasi-independent of the variable of interest. In this article, the authors propose a likelihood-based inference approach under the assumption that the dependence structure of the two variables follows a general form of copula model. They develop a model selection method for choosing the best-fitted copula among a broad class of model alternatives, and they derive large-sample properties of the proposed estimators, including the inverse Fisher information matrix. The treatment of ties is also discussed. They apply their methods to the analysis of a transfusion-related AIDS data set and compare the results with existing methods. Simulation results are also provided to evaluate the finite-sample performances of all the competing methods.

1. Introduction

Consider the situation where a pair of failure times (X, Y) is observed only if $X \le Y$. The variable X is said to be right-truncated by Y, while the variable Y is left-truncated by X. For the example of transfusion-related AIDS studied in [12], let T be the calendar time of infection, X be the duration of incubation, and τ be the end of the study in calendar time. An AIDS case can be ascertained only if $T + X \le \tau$. The incubation time X is of major interest, but it is right-truncated by $Y = \tau - T$. An example of left-truncation is the data on age at death in the Channing House Retirement Centre described on pp. 16–17 of [11]. In these data, the age at death Y is left-truncated by the entry age X.

Most methods for analyzing truncation data, including the Lynden–Bell estimator [14] for the distribution/survival function of X, rely on the assumption of quasi-independence between X and Y [17]. To verify the quasi-independence assumption, several statistical tests have been developed in [3,5,15,17]. Interestingly, these tests show that the length of incubation is dependent on the calendar time of infection in the aforementioned AIDS example. To further assess the dependence relationship, [13] proposed the semi-survival copula model defined, for all $x \le y$, by

$$\Pr(X \le x, Y > y \mid X \le Y) = \frac{C_\theta\{F_X(x), S_Y(y)\}}{c(\theta, F_X, S_Y)}, \tag{1}$$

where $C_\theta : [0, 1]^2 \to [0, 1]$ is a copula, i.e., a bivariate distribution function with uniform margins [8], and

$$c(\theta, F_X, S_Y) = \iint_{x \le y} C_\theta^{(1,1)}\{F_X(x), S_Y(y-)\} \mathrm{d}F_X(x)\{-\mathrm{d}S_Y(y)\}$$

* Corresponding author.
E-mail addresses: emura@stat.ncu.edu.tw (T. Emura), wjwang@stat.nctu.edu.tw (W. Wang).

[1] Present address: Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan.

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.03.012

is a normalizing constant defined in terms of $C_\theta^{(1,1)}(u, v) = \partial^2 C_\theta(u, v)/\partial u \partial v$. Note that quasi-independence corresponds to the case where, for all $u, v \in [0, 1]$, $C_\theta(u, v) = \Pi(u, v) \equiv uv$. The model is semiparametric in that the forms of the distribution function F_X and survival function S_Y are unspecified while the form of C_θ is specified up to an unknown parameter $\theta \in \mathbb{R}^p$.

In general, F_X and S_Y may not be the true marginal distribution and survival functions of X and Y, respectively, due to the problem of non-identifiability in the unobservable region for $X > Y$. For instance, if $X \sim \mathcal{E}(1)$, $Y \sim \mathcal{U}(0, 1)$ and $X \perp Y$, then (1) holds for $C_\theta = \Pi$, $F_X(x) = (1 - e^{-x})/(1 - e^{-1})$ and $S_Y(y) = 1 - y$ for $0 \le x \le y < 1$. Hence, F_X is the distribution of X in the observable range $0 \le x < 1$ while S_Y coincides with the true survival function of Y.

If $\Pr(X \le x, Y > y) = C_\theta\{F_X(x), S_Y(y)\}$ holds for all values in $\{(x, y) : 0 \le x, y < \infty\}$, then F_X and S_Y represent the true marginal distribution and survival function of X and Y, respectively. Also, the parameter θ coincides, up to a change in sign, with Kendall's tau for (X, Y), viz.

$$\tau(\theta) = 1 - 4 \int_0^1 \int_0^1 C_\theta(u, v) \mathrm{d}C_\theta(u, v).$$

Semiparametric inference based on a sub-class of model (1) has been considered by [1,6,13]. Specifically, these authors assumed that for all $x \le y$, the following semi-survival Archimedean copula model holds:

$$\Pr(X \le x, Y > y \mid X \le Y) = \frac{\phi_\theta^{-1}[\phi_\theta\{F_X(x)\} + \phi_\theta\{S_Y(y)\}]}{c(\theta, F_X, S_Y)}, \tag{2}$$

where $\phi_\theta : [0, 1] \to [0, \infty]$ is the generating function satisfying $\phi_\theta(1) = 0$, $\phi'_\theta(t) = \partial\phi_\theta(t)/\partial t < 0$, $\phi''_\theta(t) = \partial^2\phi_\theta(t)/\partial t^2 > 0$ for all $t \in (0, 1)$ and $\theta \in \mathbb{R}$. These papers exploited moment properties based on (2) to construct estimating equations.

In this paper, we propose a likelihood-based inference approach based on model (1), which includes more copula choices than the one-parameter Archimedean copula family. We also develop a model selection method for choosing the best-fitted copula among a broad class of model alternatives. The rest of the article is organized as follows. The proposed methodology is presented in Section 2 and large-sample analysis is described in Section 3. Modifications for ties are provided in Section 4. In Section 5, we apply all the competing methods to reanalyze the AIDS data and, in Section 6, we present simulation results. A conclusion and a discussion may be found in Section 7. Technical arguments are relegated to a series of Appendices.

2. Proposed methodology

2.1. Likelihood construction

The proposed methodology was originally motivated in [2,20], where likelihood structures were developed for analyzing transformation models under different data settings. However, to adapt to the nature of truncation, special treatment is needed. Specifically, define $H_X = -\ln(F_X)$ as the reverse-time cumulative hazard function [12], which is a right-continuous decreasing function with $H_X(\infty) = 0$ and $H_X(0) = \infty$. Define $\Lambda_Y = -\ln(S_Y)$ as the cumulative hazard function, which is a right-continuous increasing function with $\Lambda_Y(0) = 0$ and $\Lambda_Y(\infty) = \infty$. Accordingly, model (1) can be re-expressed, for all $x \le y$, as

$$\Pr(X \le x, Y > y \mid X \le Y) = \frac{C_\theta\{e^{-H_X(x)}, e^{-\Lambda_Y(y)}\}}{c(\theta, H_X, \Lambda_Y)}, \tag{3}$$

where

$$c(\theta, H_X, \Lambda_Y) = \iint_{x \le y} \eta\{H_X(x), \Lambda_Y(y-)\}\{-\mathrm{d}H_X(x)\}\mathrm{d}\Lambda_Y(y),$$

and $\eta_\theta(x, y) = e^{-x}e^{-y}C_\theta^{(1,1)}(e^{-x}, e^{-y})$. Under model (3), the density function is given, for all $x \le y$, by

$$\frac{\eta_\theta\{H_X(x), \Lambda_Y(y-)\}\{-\mathrm{d}H_X(x)\}\mathrm{d}\Lambda_Y(y)}{c(\theta, H_X, \Lambda_Y)}. \tag{4}$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identical replicates under model (3); assume that for all $j \in \{1, \ldots, n\}$, the pair (X_j, Y_j) is subject to $X_j \le Y_j$. To simplify the illustration, we temporarily assume that the data contain no ties, i.e., the observed points are all distinct. We will treat H_X and Λ_Y as right-continuous step functions that jump at their own observed values with

$$-\mathrm{d}H_X(X_j) = H_X(X_j-) - H_X(X_j), \qquad \mathrm{d}\Lambda_Y(Y_j) = \Lambda_Y(Y_j) - \Lambda_Y(Y_j-).$$

The log-likelihood is then given by

$$\ell_n(\theta, H_X, \Lambda_Y) = \sum_{j=1}^{n} \left[\ln \eta_\theta\{H_X(X_j), \Lambda_Y(Y_j-)\} + \ln\{\mathrm{d}\Lambda_Y(Y_j)\} + \ln\{-\mathrm{d}H_X(X_j)\} - \ln\{c(\theta, H_X, \Lambda_Y)\}\right], \tag{5}$$

where

$$c(\theta, H_X, \Lambda_Y) = \sum_{i=1}^{n}\left[\sum_{j:Y_j \geq X_i} \eta_\theta\{H_X(X_i), \Lambda_Y(Y_j-)\}\mathrm{d}\Lambda_Y(Y_j)\right]\{-\mathrm{d}H_X(X_i)\}$$

is an integration over a trapezoid

$$\mathbf{T} = \{(x, y) : x \leq y, X_{(1)} \leq x \leq X_{(n)}, Y_{(1)} \leq y \leq Y_{(n)}\},$$

and where $X_{(j)}$ and $Y_{(j)}$ are the jth smallest value of X and Y, respectively. We propose to maximize $\ell_n(\theta, H_X, \Lambda_Y)$ under the additional constraints

$$-\mathrm{d}H_X(X_{(1)}) = 1, \qquad \mathrm{d}\Lambda_Y(Y_{(n)}) = 1.$$

These constraints are necessary for the NPMLE, denoted as $(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)$, to be uniquely determined. Otherwise, there exist infinitely many solutions for $(\hat{H}_X, \hat{\Lambda}_Y)$ for a given θ.

The proposed likelihood-based estimator can be applied to copula models outside the one-parameter Archimedean copula family. These examples include the Plackett, Gaussian, and Student t copulas. Our method can also handle the situation where the association of (X, Y) is induced by a common frailty variable [16] such that

$$\Pr(X > x, Y > y) = \phi_\theta^{-1}[\phi_\theta\{\Pr(X > x)\} + \phi_\theta\{\Pr(Y > y)\}],$$

where ϕ_θ^{-1} is the Laplace transformation of the underlying frailty variable. The truncated distribution $(X, Y) \mid X \leq Y$ does not verify (2), but still verifies (1) with $C_\theta(u, v) = v - \phi_\theta^{-1}\{\phi_\theta(1-u) + \phi_\theta(v)\}$ for all $u, v \in (0, 1)$.

2.2. Score equations

Differentiating $\ell_n(\theta, H_X, \Lambda_Y)$ in Eq. (5) with respect to θ, the score equation is given by

$$0 = \sum_{j=1}^{n}\left[\frac{\dot{\eta}_\theta}{\eta_\theta}\{H_X(X_j), \Lambda_Y(Y_j-)\} - \frac{\dot{c}}{c}(\theta, H_X, \Lambda_Y)\right], \tag{6}$$

where

$$\dot{c}(\theta, H_X, \Lambda_Y) = \sum_{i=1}^{n}\left[\sum_{j:Y_j \geq X_i} \dot{\eta}_\theta\{H_X(X_i), \Lambda_Y(Y_j-)\}\mathrm{d}\Lambda_Y(Y_j)\right]\{-\mathrm{d}H_X(X_i)\}$$

and $\dot{\eta}_\theta(x, y) = \partial\eta_\theta(x, y)/\partial\theta$. Similarly, the score equations for $-\mathrm{d}H_X(X_i)$ and $\mathrm{d}\Lambda_Y(Y_i)$ are

$$-\mathrm{d}H_X(X_i) = \left\{\sum_{j=1}^{n}\Psi_j^{(1,0)}(X_i; \theta, H_X, \Lambda_Y)\right\}^{-1},$$

$$\mathrm{d}\Lambda_Y(Y_i) = \left\{\sum_{j=1}^{n}\Psi_j^{(0,1)}(Y_i; \theta, H_X, \Lambda_Y)\right\}^{-1}, \tag{7}$$

where $\Psi_j^{(1,0)}$ and $\Psi_j^{(0,1)}$ are explicitly defined in Appendix A. It follows from Eq. (7) that

$$H_X(x) = \int_x^\infty \frac{\sum_{j=1}^{n}\mathrm{d}\{\mathbf{1}(X_j \leq u)\}}{\sum_{j=1}^{n}\Psi_j^{(1,0)}(u; \theta, H_X, \Lambda_Y)},$$

$$\Lambda_Y(y) = \int_0^y \frac{\sum_{j=1}^{n}\mathrm{d}\{\mathbf{1}(Y_j \leq u)\}}{\sum_{j=1}^{n}\Psi_j^{(0,1)}(u; \theta, H_X, \Lambda_Y)}. \tag{8}$$

2.3. Numerical algorithms

Note that (7) can be viewed as a self-consistency algorithm. For a fixed value of θ, and under the additional constraints $-\mathrm{d}H_X(X_{(1)}) = 1$ and $\mathrm{d}\Lambda_Y(Y_{(n)}) = 1$, one can update the left-hand sides with current estimates and solve the equations repeatedly until the convergence criterion is reached. The resulting estimators for H_X and Λ_Y are then plugged into (6) to solve for θ. The procedures for estimating $(-\mathrm{d}H_X, \mathrm{d}\Lambda_Y)$ and θ will iterate until the convergence criterion is reached.

Alternatively, one can numerically maximize (5) by adopting a Newton-type algorithm that involves calculating the $(2n-2+p)$-dimensional vector of score and a $(2n-2+p)\times(2n-2+p)$ Hessian matrix in each iteration, where p is the dimension of θ. If $p=1$, the starting values may be set as $\theta\approx 1$ with $\lim_{\theta\to 1} C_\theta=\Pi$, $-\mathrm{d}H_X(X_j)=1/R(X_j)$ and $\mathrm{d}\Lambda_Y(Y_j)=1/R(Y_j)$, where $R(u)=\sum \mathbf{1}(X_j\le u\le Y_j)$. Note that Eqs. (6)–(7) can still be used to set the convergence criteria. The Newton-type algorithm requires less programming effort than the first proposal because it can be implemented using `R` routine `nlm` in which the score vector and Hessian matrix are internally evaluated. The maximization of $(-\mathrm{d}H_X,\mathrm{d}\Lambda_Y)\in[0,\infty)^{2n-2}$ may be done with the log-transformed parameter $(\ln(-\mathrm{d}H_X),\ln(\mathrm{d}\Lambda_Y))\in\mathbb{R}^{2n-2}$ in the unrestricted parameter space. This method is adopted to obtain the numerical results of Sections 5 and 6.

Our numerical studies confirm that any initial choice $-\mathrm{d}H_X(X_{(1)})=p$ and $\mathrm{d}\Lambda_Y(Y_{(n)})=q$ for $0<p,q\le 1$ leads to a unique NPMLE, but the results depend on (p,q). The choice $p=q=1$ seems to be a natural one since the resulting NPMLE is reduced to the Nelson–Aalen type estimates $-\mathrm{d}H_X(x)=1/R(x)$ and $\mathrm{d}\Lambda_Y(y)=1/R(y)$ at $x=X_{(1)}$ and $y=Y_{(n)}$, respectively.

2.4. Copula model selection

We discuss how to choose the best-fitting copula based on the data at hand among several model candidates. First, consider $\theta\in\mathbb{R}$. Let $C^{(0)}=\Pi$ be the independence copula and $C_\theta^{(k)}$ for $k=1,\dots,K$ be K different copula candidates such that, for all $k\in\{1,\dots,K\}$, $C_\theta^{(k)}\to\Pi$ as $\theta\to 1$. We are interested in choosing the best-fitted one among the copulas. For each $k\in\{1,\dots,K\}$, we calculate the deviance measure

$$2\{\ell_n(\hat\theta,\hat H_X,\hat\Lambda_Y)-\ell_n(1,\hat H_X^{\theta=1},\hat\Lambda_Y^{\theta=1})\},$$

where $\hat H_X^{\theta=1}$ and $\hat\Lambda_Y^{\theta=1}$ are the NPMLE under $C^{(0)}$. In other words, what we formulate is a quasi-independence test for $H_0:\theta=1$. Without giving a formal proof, we conjecture that under $C^{(0)}$ the deviance measure follows the chi-square distribution with one degree of freedom. This conjecture can then be applied to compute a p-value based on the observed deviance statistic. If $\theta\in\mathbb{R}^p$, we conjecture that the deviance follows the chi-square distribution with p degree of freedom. To choose the best copula, we first exclude those copulas with p-values larger than the selected significance level (say, 0.05). We then choose the copula that yields the smallest p-value. If all copulas have p-values greater than the significance level, then we choose the independence copula.

3. Asymptotic analysis

In this section, we state our main asymptotic results. Detailed proofs are given in Appendix B. We denote the true parameters by $(\theta^0,H_X^0,\Lambda_Y^0)$ and assume that $(\theta^0,H_X^0,\Lambda_Y^0)\in\Theta$, where Θ is the parameter space for (θ,H_X,Λ_Y). Let $E(\cdot)$ denote expectation with respect to a measure corresponding to the true density in (4), i.e.,

$$\mathrm{d}P^0(x,y)=\frac{\eta_{\theta^0}\{H_X^0(x),\Lambda_Y^0(y-)\}\{-\mathrm{d}H_X^0(x)\}\mathrm{d}\Lambda_Y^0(y)}{c(\theta^0,H_X^0,\Lambda_Y^0)}.$$

Although the basic framework of the asymptotic analysis is similar to that in [20], the technical details of our work are quite different. The following conditions are assumed.

Assumption I. For all θ, C_θ is twice differentiable on $(0,1)^2$ and

$$0<\inf_{u,v\in[0,1]} C_\theta^{(1,1)}(u,v)\le\sup_{u,v\in[0,1]} C_\theta^{(1,1)}(u,v)<\infty.$$

Assumption II. There exists $\varepsilon>0$ such that $\varepsilon\le c(\theta,H_X,\Lambda_Y)$ for all $(\theta,H_X,\Lambda_Y)\in\Theta$.

Assumption III. If $(\theta,H_X,\Lambda_Y)\ne(\theta^0,H_X^0,\Lambda_Y^0)$, then

$$E\left(\ln\left[\frac{\eta_\theta\{H_X(X),\Lambda_Y(Y-)\}}{c(\theta,H_X,\Lambda_Y)}h_X(X)\lambda_Y(Y)\right]\right)<E\left(\ln\left[\frac{\eta_{\theta^0}\{H_X^0(X),\Lambda_Y^0(Y-)\}}{c(\theta^0,H_X^0,\Lambda_Y^0)}h_X^0(X)\lambda_Y^0(Y)\right]\right).$$

Assumption IV. The information operator $\dot W:\Theta\to\mathbb{R}^p\times\{\ell^\infty(Q)\}^2$ defined in Appendix B.2 is continuously invertible.

Assumption I excludes the possibility that the NPMLE diverges to infinity, say, $-\mathrm{d}\hat H_X(X_j)=\infty$ or $\mathrm{d}\hat\Lambda_Y(Y_j)=\infty$. The Frank and Plackett copulas satisfy this condition while the Clayton and Gumbel copulas do not. If the latter happens, one can modify $C_\theta^{(1,1)}$ to be bounded. From our experience, modification by setting $C_\theta^{(1,1)}(u\wedge 0.99,v\wedge 0.99)$ can work quite well when implementing Newton-type algorithms. We will adopt such a modification in the data analysis presented in Section 5.

Assumption II is required for the density function in (4) to be well defined. Clearly, $c(\theta^0, H_X^0, \Lambda_Y^0) \geq \varepsilon > 0$; otherwise nothing is observed. Therefore, the condition can be satisfied by defining the parameter space in the neighborhood of $(\theta^0, H_X^0, \Lambda_Y^0)$.

Assumption III states that the true parameter should be a well-separated point of maximum in terms of the Kullback–Leibler divergence, which is usually imposed for the consistency of M-estimators; see [18, p. 62]. It follows from Jensen's inequality for a function $-\ln(x)$ that, for any $(\theta, H_X, \Lambda_Y) \in \Theta$,

$$E\left(\ln\left[\frac{\eta_\theta\{H_X(X), \Lambda_Y(Y-)\}}{c(\theta, H_X, \Lambda_Y)} h_X(X)\lambda_Y(Y)\right]\right) \leq E\left(\ln\left[\frac{\eta_{\theta^0}\{H_X^0(X), \Lambda_Y^0(Y-)\}}{c(\theta^0, H_X^0, \Lambda_Y^0)} h_X^0(X)\lambda_Y^0(Y)\right]\right).$$

Assumption III also requires that the equality holds if and only if $(\theta, H_X, \Lambda_Y) = (\theta^0, H_X^0, \Lambda_Y^0)$. We examine Assumption III numerically in the Supplemental document.

Assumption IV stipulates the non-singularity of the information operator $\dot{W} : \Theta \to \mathbb{R}^p \times \{\ell^\infty(Q)\}^2$. Since it is difficult to find the inverse of a function that takes a value in a functional space, a convenient alternative is to check the non-singularity of the observed Fisher information matrix, which will be defined in Section 3.2. In the numerical studies of Sections 5 and 6, we have confirmed that the observed Fisher information matrix is always non-singular and thus invertible.

3.1. Consistency

Under Assumptions I, II and III, it is shown in Appendix B.1 that $\hat{\theta} \to \theta^0$, $\hat{H}_X \to H_X^0$, and $\hat{\Lambda}_Y \to \Lambda_Y^0$ almost surely. Here we provide some intuitive explanations for that result. First, note that, if (H_X^0, Λ_Y^0) are known, the model is parametric with dependent truncation [4]. Then, the strong consistency $\hat{\theta} \to \theta^0$ is intuited from the parametric likelihood theory. Next, we assume that θ^0 is known. We show in Appendix A that

$$E\{\Psi_j^{(1,0)}(s;\theta^0, H_X^0, \Lambda_Y^0)\} = \frac{e^{-H_X^0(s)}C_{\theta^0}^{(1,0)}\{e^{-H_X^0(s)}, e^{-\Lambda_Y^0(s-)}\}}{c(\theta^0, H_X^0, \Lambda_Y^0)},$$
$$E\{\Psi_j^{(0,1)}(s;\theta^0, H_X^0, \Lambda_Y^0)\} = \frac{e^{-\Lambda_Y^0(s-)}C_{\theta^0}^{(0,1)}\{e^{-H_X^0(s)}, e^{-\Lambda_Y^0(s-)}\}}{c(\theta^0, H_X^0, \Lambda_Y^0)}.$$

By the Glivenko–Cantelli theorem, the right-hand side of (8) evaluated at $(\theta^0, H_X^0, \Lambda_Y^0)$ converges almost surely to the limit of

$$\tilde{H}_X(x) = \frac{c(\theta^0, H_X^0, \Lambda_Y^0)}{n}\sum_{j=1}^n \int_x^\infty \frac{\mathrm{d}\{\mathbf{1}(X_i \leq u)\}}{e^{-H_X^0(u)}C_{\theta^0}^{(1,0)}\{e^{-H_X^0(u)}, e^{-\Lambda_Y^0(u-)}\}}, \tag{9a}$$

$$\tilde{\Lambda}_Y(y) = \frac{c(\theta^0, H_X^0, \Lambda_Y^0)}{n}\sum_{j=1}^n \int_0^y \frac{\mathrm{d}\{\mathbf{1}(Y_j \leq u)\}}{e^{-\Lambda_Y^0(u-)}C_{\theta^0}^{(0,1)}\{e^{-H_X^0(u)}, e^{-\Lambda_Y^0(u-)}\}}. \tag{9b}$$

Here, $\tilde{H}_X$ and $\tilde{\Lambda}_Y$ are the sum of independent and identically distributed terms with $E\{\tilde{H}_X(x)\} = H_X^0(x)$ and $E\{\tilde{\Lambda}_Y(y)\} = \Lambda_Y^0(y)$. Again, by applying the Glivenko–Cantelli theorem to (9a) and (9b), we obtain $\tilde{H}_X \to H_X^0$ and $\tilde{\Lambda}_Y \to \Lambda_Y^0$ almost surely.

3.2. Asymptotic normality

We study the asymptotic distribution of the random variable

$$n^{1/2}\left[b^\top(\hat{\theta} - \theta^0) + \int_0^\infty w_X(u)\{-\mathrm{d}\hat{H}_X(u) + \mathrm{d}H_X^0(u)\} + \int_0^\infty w_Y(u)\{\mathrm{d}\hat{\Lambda}_Y(u) - \mathrm{d}\Lambda_Y^0(u)\}\right], \tag{10}$$

where $b \in \mathbb{R}^p$ and w_X and w_Y are bounded functions. Define

$$i_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y) = \hat{i}_n = \begin{bmatrix} \hat{i}_{n,11} & \hat{i}_{n,12}^\top \\ \hat{i}_{n,12} & \hat{i}_{n,22} \end{bmatrix},$$

where $\hat{i}_{n,11} : p \times p$, $\hat{i}_{n,12} : (2n-2) \times p$, and $\hat{i}_{n,22} : (2n-2) \times (2n-2)$ are the observed Fisher information matrix which equals the negative of the Hessian matrix of $\ell_n(\theta, H_X, \Lambda_Y)$ with respect to θ, $-\mathrm{d}H_X(X_{(j)})$ for $j \in \{2, \ldots, n\}$, and $\mathrm{d}\Lambda_Y(Y_{(j)})$ for $j \in \{1, \ldots, n-1\}$ evaluated at $(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)$.

Under Assumptions I, II, III and IV, it is shown in Appendix B.2 that the variable defined in (10) converges weakly to a zero mean Gaussian distribution whose asymptotic variance can be estimated by

$$(b^\top, W_X^\top, W_Y^\top)(\hat{i}_n/n)^{-1}(b^\top, W_X^\top, W_Y^\top)^\top, \tag{11}$$

where $W_X^\top = (w_X(X_{(2)}), \ldots, w_X(X_{(n)}))$ and $W_Y^\top = (w_Y(Y_{(1)}), \ldots, w_Y(Y_{(n-1)}))$. Eq. (11) can be used to derive the estimators for the marginal asymptotic variances of $(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)$, respectively. For example, when $p = 1$, the asymptotic variance of $\hat{\theta}$ is obtained by setting $w_X(u) = 0$, $w_Y(u) = 0$ and $b = 1$, which has the form

$$\hat{V}(\hat{\theta}) = (\hat{i}_{n,11} - \hat{i}_{n,12}^\top \hat{i}_{n,22}^{-1} \hat{i}_{n,12})^{-1}.$$

For the purpose of constructing confidence intervals, one may use the log-transformed intervals to improve the normal approximation. For example, when $\theta \in (0, \infty)$, the 95% confidence interval for θ is

$$(\hat{\theta} \exp\{-1.96\, \hat{V}(\hat{\theta})^{1/2}/\hat{\theta}\}, \hat{\theta} \exp\{1.96\, \hat{V}(\hat{\theta})^{1/2}/\hat{\theta}\}).$$

Also, the variance estimator of $\hat{\Lambda}_Y(y)$, denoted by $\hat{V}\{\hat{\Lambda}_Y(y)\}$, is obtained by setting $w_X(u) = 0$, $w_Y(u) = \mathbf{1}(u \le y)$ and $b = 0$. By the Delta Method, the variance estimator for $\hat{S}_Y(y) = e^{-\hat{\Lambda}_Y(y)}$ is $\hat{V}\{\hat{S}_Y(y)\} = \hat{S}_Y(y)^2 \hat{V}\{\hat{\Lambda}_Y(y)\}$, which is useful for constructing the confidence interval.

4. Modifications for ties

In the previous discussions, we assumed for convenience that the data contain no ties. Now we discuss how to handle tied data, which are commonly seen in practical applications. It is worth noting that the methods of [6,13] critically rely on the assumption of no ties; furthermore, the extensions required to handle ties do not seem trivial. The latter paper suggested a tie breaking procedure by adding small noise to the original data. Although the effect of such modification should be modest, it creates unnecessary jumps in the estimated marginal functions; see Section 5 for an example. For additional discussion of the detrimental effects of data jittering in copula modeling contexts, see [8,9].

Modification of the proposed likelihood method for ties is straightforward and also more natural. Specifically let $X^*_{(1)} < \cdots < X^*_{(n_X)}$ and $Y^*_{(1)} < \cdots < Y^*_{(n_Y)}$ be distinct observed values of X and Y, respectively. Then, the likelihood function can still be written as (5), but the definition of $c(\theta, H_X, \Lambda_Y)$ is modified to become

$$c(\theta, H_X, \Lambda_Y) = \sum_{i=1}^{n_X} \left[\sum_{j: Y^*_{(j)} \ge X^*_{(i)}} \eta_\theta\{H_X(X^*_{(i)}), \Lambda_Y(Y^*_{(j)}-)\} \mathrm{d}\Lambda_Y(Y^*_{(j)}) \right] \{-\mathrm{d}H_X(X^*_{(i)})\}.$$

The NPMLE is obtained by maximizing (5) under the constraints $-\mathrm{d}H_X(X^*_{(1)}) = 1$ and $\mathrm{d}\Lambda_Y(Y^*_{(n_Y)}) = 1$. For numerical maximization, we recommend choosing the following starting values:

$$-\mathrm{d}H_X(X^*_{(i)}) = \sum_{k=1}^{n} \mathbf{1}(X_k = X^*_{(i)})/R(X^*_{(i)}),$$

$$\mathrm{d}\Lambda_Y(Y^*_{(j)}) = \sum_{k=1}^{n} \mathbf{1}(Y_k = Y^*_{(j)})/R(Y^*_{(j)}).$$

In Section 5, we will apply this modification in analysis of the AIDS data which contains many tied observations.

5. Data analysis

The proposed method is applied to the transfusion-related AIDS data available in [10]. Let T be the infection time and X be the incubation time from infection to AIDS. Since the total study period was 102 months, any individuals with $T + X \le 102$ could be included in the sample. The incubation time X is right-truncated by $Y = 102 - T$. The sample consists of $(X_1, Y_1), \ldots, (X_{293}, Y_{293})$, subject to $X_j \le Y_j$ for all $j \in \{1, \ldots, 293\}$. Of major interest in the study is the estimation of the marginal distribution of X as well as the association between T and X.

Table 1 lists the results for fitting 12 copulas (2 semi-survival Archimedean copula models, 8 non-semi-survival Archimedean copula models and 2 two-parameter copula models). The deviances of the 12 copulas all yield p-values less than 0.05; hence, we exclude the independence copula. By applying the proposed model selection procedure, the semi-survival Clayton copula provides the best fit among all the competitors as it yields the smallest p-value. Note that the same model was also selected by the method of [1]. However, their choice is based on comparing only three semi-survival Archimedean copula models which do not include the independence copula. Our model selection based on a larger pool of copula choices strengthens the evidence that the semi-survival Clayton copula is indeed suitable for the AIDS data.

Assuming the semi-survival Clayton copula, we compared the proposed NPMLE with the two moment-based estimation methods. The estimates, and the corresponding estimated standard errors in parentheses, are 0.763 (0.033) for the NPMLE,

Table 1
Analysis of the transfusion-related AIDS data.

$C_\theta(u, v)$	Form[a]	$\hat{\theta}$ (SE)	Kendall's τ on (X, Y)	95% CI for θ	Deviance (p-value)[b]
Clayton	Semi-survival	0.763 (0.033)	0.134	(0.701, 0.831)	19.028 (0.000)
	Regular	1.521 (0.172)	0.207	(1.218, 1.898)	8.568 (0.003)
	Survival	1.645 (0.233)	0.244	(1.246, 2.171)	5.228 (0.022)
Frank	Semi-survival	55.72 (42.66)	0.390	(12.43, 249.90)	10.828 (0.001)
Plackett	Semi-survival	0.189 (0.050)	0.356	(0.113, 0.316)	8.068 (0.005)
Normal	Semi-survival	−0.516 (0.083)	0.345	(−0.678,−0.353)	14.341 (0.000)
$t_{(df=10)}$	Semi-survival	−0.520 (0.076)	0.350	(−0.669,−0.371)	9.559 (0.002)
$t_{(df=5)}$	Semi-survival	−0.507 (0.073)	0.344	(−0.650,−0.363)	3.959 (0.047)
Gumbel	Regular	1.459 (0.136)	0.315	(1.257, 1.821)	7.868 (0.005)
	Survival	1.340 (0.120)	0.254	(1.170, 1.678)	6.368 (0.012)
Two-parameter	Regular	$\hat{\theta}$:1.521 (0.400)	0.207	θ:(1.116, 3.348)	8.588 (0.003)
		$\hat{\beta}$:1.000[c]			
	Survival	$\hat{\theta}$:1.344 (0.264)	0.309	θ:(1.076, 2.551)	7.928 (0.019)
		$\hat{\beta}$:1.235 (0.140)		β:(1.073, 1.756)	

[a] A copula C_θ is used to model the distribution of (X, Y) in three different forms: (i) Semi-survival form: $\Pr(X \le x, Y > y | X \le Y) = C_\theta\{e^{-H_X(x)}, e^{-\Lambda_Y(y)}\}/c$. (ii) Regular form: $\Pr(X \le x, Y > y | X \le Y) = [e^{-H_X(x)} - C_\theta\{e^{-H_X(x)}, 1 - e^{-\Lambda_Y(y)}\}]/c$. (iii) Survival form: $\Pr(X \le x, Y > y | X \le Y) = [e^{-\Lambda_Y(y)} - C_\theta\{1 - e^{-H_X(x)}, e^{-\Lambda_Y(y)}\}]/c$. For Frank, Plackett, Gaussian and t-copulas, the three forms lead to the same deviance and Kendall's tau. Hence, only the results for the semi-survival form are reported. For Gumbel and two-parameter copulas, the semi-survival form allows only negative association on (X, Y) and is not suitable for this data. Hence, the results for regular and survival forms are reported.

[b] The p-value is the probability that the chi-squared distribution with one or two (for the two-parameter family) degrees of freedom exceeds the observed deviance.

[c] Maximum of the likelihood function is attained at the parameter boundary $\beta = 1$, and hence the standard error based on the observed Fisher information is not available.

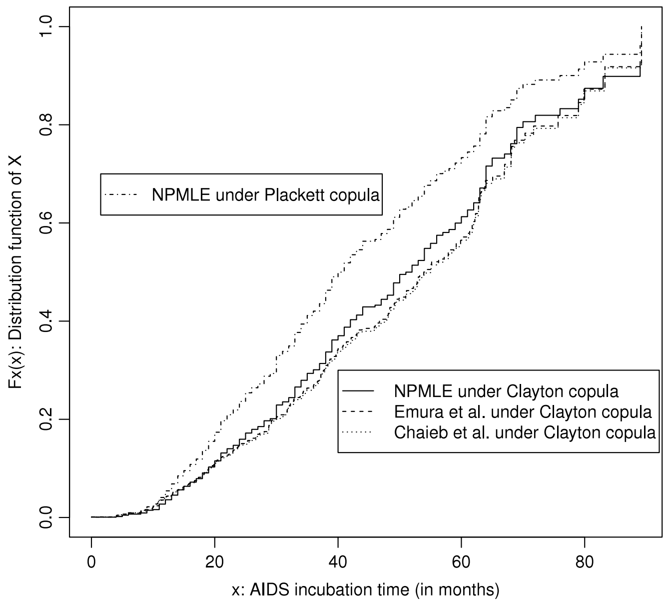

Fig. 1. The estimated cumulative distribution functions of the incubation time of AIDS.

0.816 (0.046) for the method of [6], and 0.823 (0.055) for the method of [13]. The corresponding estimates of Kendall's tau are 0.134 (NPMLE), 0.101 (method [6]) and 0.097 (method [13]). Note that the standard error of the NPMLE uses the observed Fisher information matrix while that of the two competitors use the jackknife method.

All the results show weak negative association between T and X. Fig. 1 displays the estimated distribution functions of X obtained by the three methods. These results imply that, as long as the same copula is fitted, the marginal estimation roughly produces the same result. As mentioned earlier, the marginal estimator by the NPMLE method jumps at observed values. However, the tie-breaking approach adopted by the two competing methods creates unnecessary jumps in the estimated distribution curves of X (see Fig. 1).

We further study the impact of the copula misspecification on the NPMLE under the Plackett copula. The resulting estimated distribution curve of X is somewhat different from the three curves computed under the Clayton copula (Fig. 1).

Table 2
Finite-sample performances of the proposed estimator under the Plackett copula.

Parameter		Mean (Bias)	SE	SEE	95% cov
Spearman's $\rho = 0.25$ ($\theta = 1/2.15$, $\Pr(X \le Y) = 0.79$)					
$\ln(\theta) = -0.765$	$n = 125$	-0.778 (-0.013)	0.407	0.407	0.945
	$n = 250$	-0.697 (0.068)	0.311	0.296	0.965
$H_X(x) = 0.693$	$n = 125$	0.736 (0.043)	0.123	0.121	0.955
	$n = 250$	0.733 (0.040)	0.090	0.086	0.970
$\Lambda_Y(y) = 0.693$	$n = 125$	0.710 (0.017)	0.144	0.139	0.960
	$n = 250$	0.725 (0.032)	0.104	0.102	0.970
Spearman's $\rho = 0.5$ ($\theta = 1/5.11$, $\Pr(X \le Y) = 0.84$)					
$\ln(\theta) = -1.631$	$n = 125$	-1.642 (-0.011)	0.323	0.319	0.965
	$n = 250$	-1.652 (-0.021)	0.231	0.222	0.940
$H_X(x) = 0.693$	$n = 125$	0.726 (0.033)	0.101	0.092	0.910
	$n = 250$	0.716 (0.023)	0.067	0.064	0.920
$\Lambda_Y(y) = 0.693$	$n = 125$	0.704 (0.011)	0.110	0.102	0.960
	$n = 250$	0.701 (0.008)	0.068	0.069	0.950
Spearman's $\rho = -0.25$ ($\theta = 2.15$, $\Pr(X \le Y) = 0.72$)					
$\ln(\theta) = 0.765$	$n = 125$	0.859 (0.094)	0.598	0.554	0.960
	$n = 250$	0.717 (-0.048)	0.342	0.359	0.930
$H_X(x) = 0.693$	$n = 125$	0.809 (0.116)	0.313	0.244	0.960
	$n = 250$	0.717 (0.024)	0.139	0.138	0.935
$\Lambda_Y(y) = 0.693$	$n = 125$	0.793 (0.100)	0.363	0.267	0.960
	$n = 250$	0.699 (0.006)	0.139	0.137	0.930
Spearman's $\rho = -0.5$ ($\theta = 5.11$, $\Pr(X \le Y) = 0.70$)					
$\ln(\theta) = 1.631$	$n = 125$	1.758 (0.127)	0.818	0.598	0.915
	$n = 250$	1.708 (0.077)	0.534	0.386	0.955
$H_X(x) = 0.693$	$n = 125$	0.883 (0.190)	0.582	0.343	0.925
	$n = 250$	0.787 (0.094)	0.374	0.196	0.960
$\Lambda_Y(y) = 0.693$	$n = 125$	0.862 (0.169)	0.624	0.354	0.885
	$n = 250$	0.775 (0.082)	0.404	0.207	0.955

Simulation mean, bias (in parenthesis), standard error (SE), average standard error estimate (SEE) and coverage rates of 95% confidence interval based on 200 runs are reported. The functions $H_X(x)$ and $\Lambda_Y(y)$ are evaluated at $x = 0.462$ and $y = 1.386$, respectively, such that $e^{-H_X(x)} = e^{-\Lambda_Y(y)} = 0.5$.

6. Simulation studies

6.1. Performances under Plackett copula

Here, we evaluate the finite-sample performance of the proposed NPMLE and the variance estimator. We consider the Plackett copula,

$$C_\theta(u, v) = \frac{1}{2(\theta - 1)} + \frac{u + v}{2} - \frac{[\{1 + (\theta - 1)(u + v)\}^2 - 4uv\theta(\theta - 1)]^{1/2}}{2(\theta - 1)},$$

which does not belong to the Archimedean copula family, so that the methods of [6,13] cannot be applied. In the simulations, the marginal functions follow exponential distributions with

$$H_X(x) = -\ln(1 - e^{-\lambda_X x}), \qquad \Lambda_Y(y) = \lambda_Y y.$$

The values of θ are chosen as 1/2.51, 1/5.11, 2.51 and 5.11, which correspond to Spearman's rho taking values 0.25, 0.50, -0.25, -0.50, respectively, for the pre-truncated pair of (X, Y). Note that Kendall's tau for this model does not have a closed form. Also $\theta > 1$ corresponds to a negative association and $0 < \theta < 1$ corresponds to a positive association. Then, truncation data $(X_1, Y_1), \ldots, (X_n, Y_n)$, subject to $X_j \le Y_j$, are generated and the NPMLE is computed over 200 simulation runs.

Table 2 presents the results for estimating θ, $H_X(x)$ and $\Lambda_Y(y)$ under $\lambda_X = 1.5$ and $\lambda_Y = 0.5$. In general, the NPMLE provides accurate estimates of the true parameters. The standard errors decrease as the sample size increases from $n = 125$ to 250 and are close to the average of the estimated standard errors. Also the coverage rates of the confidence intervals are close to the nominal 95% level in most cases. The bias of the NPMLE tends to inflate when data exhibit stronger negative association.

Fig. 2 summarizes the results for estimating $F_X(x) = e^{-H_X(x)}$ at $F_X(x) = 0.2, 0.3, \ldots, 0.8$ under $\lambda_X = 1.5$ and $\lambda_Y = 0.5$. The results show that the estimator generally performs well. The average upper and lower limits in the 95% confidence intervals agree with the 2.5% and 97.5% quantiles of $\hat{F}_X(x)$, respectively. Accordingly, the coverage rates of the confidence interval are close to the nominal 95% in all configurations. However, both the bias and standard deviation of the estimator $\hat{F}_X(x) = \exp\{-\hat{H}_X(x)\}$ gets large as $x \to \infty$. This is due to the lack of information in the upper tail of F_X where a large value of

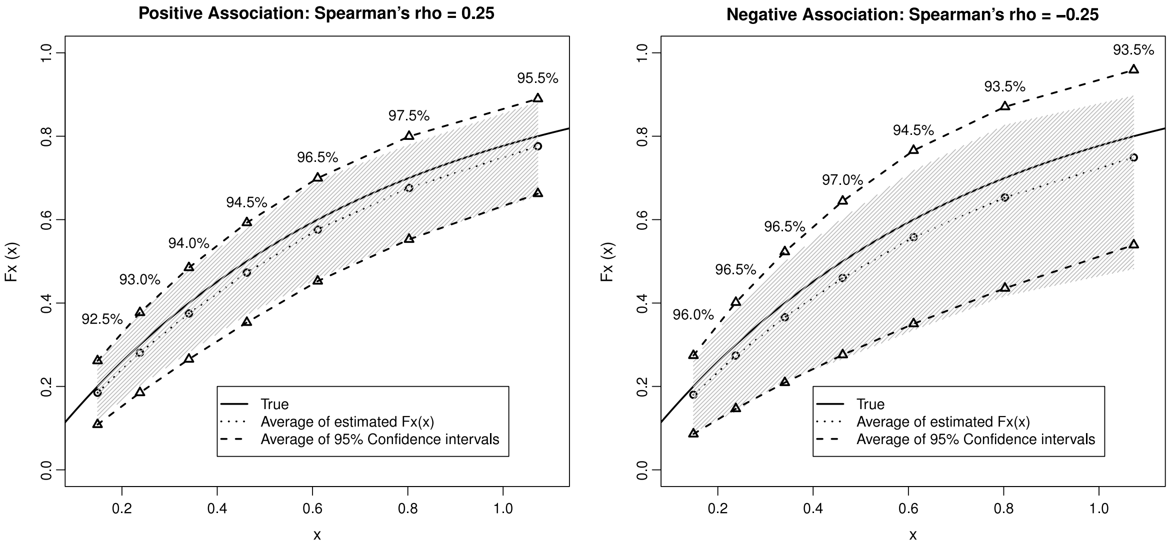

Fig. 2. The performance of the estimator $\hat{F}_X(x)$ based on 200 simulation runs under the Plackett copula model. The solid line (—) draws the true distribution function $F_X(x) = 1 - e^{-1.5x}$ and the dotted line (·····) draws the average of $\hat{F}_X(x)$. The shaded regions are defined by the 2.5th and 97.5th empirical quantiles of $\hat{F}_X(x)$ and the dashed lines (– – – –) are the average of 95% confidence intervals. The numbers near the upper confidence limit is the percentage of the confidence intervals that cover the true value $F_X(x)$.

Table 3
Comparison of the deviance statistic with the chi-squared distribution with one degree of freedom under the Plackett copula.

(λ_X, λ_Y)	n	$\Pr\{\text{Deviance} > \chi^2_{df=1}(1-p)\}$				E (Deviance)
		$p = 0.05$	$p = 0.10$	$p = 0.20$	$p = 0.50$	
(1.5, 0.5)	75	0.065	0.115	0.225	0.495	1.167
	125	0.070	0.095	0.215	0.540	1.178
(1.5, 1.0)	75	0.060	0.140	0.285	0.585	1.267
	125	0.075	0.140	0.230	0.530	1.105
(1.0, 1.0)	75	0.050	0.100	0.185	0.485	0.993
	125	0.055	0.100	0.185	0.480	0.974

Each cell contains the empirical probability $\Pr\{\text{Deviance} > \chi^2_{df=1}(1-p)\}$, where Deviance $= 2\{\ell_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y) - \ell_n(1, \hat{H}_X^{\theta=1}, \hat{\Lambda}_Y^{\theta=1})\}$, and the empirical average E (Deviance) is based on 200 runs.

X is likely to be right-truncated by Y. A similar phenomenon can be found in the Kaplan–Meier estimator for right censored data.

Table 3 summarizes the sampling distribution of the proposed deviance between the independence copula and the Plackett copula. We generated data under the null model (independence copula) with three configurations: $(\lambda_X, \lambda_Y) = (1.5, 0.5)$, $(1.5, 1.0)$, and $(1.0, 1.0)$. It is seen that the empirical probabilities that the deviance exceeds the upper 5% points of the chi-square distribution are between 5% and 7.5%. Also, the averages of the deviance statistics vary between 0.974 and 1.267. Our conjecture that the deviance is approximated by the chi-squared distribution with one degree of freedom seems to be supported by the simulation results.

6.2. Performances under Frank copula

We compare the three competing methods under the semi-survival Frank Archimedean copula model using the parameterization from [7]. That is

$$C_\theta(u, v) = \log_\theta\{1 + (\theta^u - 1)(\theta^v - 1)/(\theta - 1)\}.$$

Note that $\theta > 1$ corresponds to a positive association between X and Y and $\theta \in (0, 1)$ corresponds to a negative association. The values of $\ln(\theta)$ are set to be 2.380, 5.746, −2.380, and −5.746, which correspond to Kendall's tau 0.25, 0.5, −0.25, and −0.5, respectively. The same marginal functions are chosen as in Section 6.1 with $\lambda_X = 1.5$ and $\lambda_Y = 0.5$.

Table 4 summarizes the results. When Kendall's tau are 0.25, 0.5, and −0.25, the three methods produce almost unbiased results. On the other hand, when Kendall's tau is −0.5, the NPMLE remains roughly unbiased but the two competitors become biased. Since these two methods sometimes fail to produce a proper solution due to numerical reasons, we propose some

Table 4
Finite-sample performances of three estimators under the Frank copula.

Parameter	n	NPMLE	Emura et al.	Chaieb et al.
Kendall's $\tau = 0.25$				
$\ln(\theta) = 2.38$	125	−0.0362 (0.9537)	−0.0999 (0.9306)	−0.0995 (0.9317)
	250	−0.0206 (0.6150)	−0.0454 (0.6065)	−0.0451 (0.6072)
$F_X(x) = 0.50$	125	−0.0172 (0.0497)	−0.0145 (0.0503)	−0.0145 (0.0504)
	250	−0.0076 (0.0346)	−0.0056 (0.0353)	−0.0056 (0.0354)
$S_Y(y) = 0.50$	125	−0.0062 (0.0689)	−0.0072 (0.0689)	−0.0071 (0.0689)
	250	0.0013 (0.0426)	0.0011 (0.0427)	0.0011 (0.0427)
Kendall's $\tau = 0.5$				
$\ln(\theta) = 5.746$	125	0.1604 (1.2739)	−0.1829 (0.9508)	−0.1834 (0.9539)
	250	−0.0122 (0.6737)	−0.1078 (0.6524)	−0.1085 (0.6529)
$F_X(x) = 0.50$	125	−0.0051 (0.0519)	−0.0081 (0.0432)	−0.0081 (0.0432)
	250	−0.0049 (0.0301)	−0.0031 (0.0327)	−0.0032 (0.0328)
$S_Y(y) = 0.50$	125	−0.0068 (0.0558)	−0.0031 (0.0433)	−0.0031 (0.0432)
	250	0.0004 (0.0290)	−0.0009 (0.0305)	−0.0009 (0.0305)
Kendall's $\tau = -0.25$				
$\ln(\theta) = -2.38$	125	−0.1216 (1.1843)	0.5180 (1.0491)	0.5521 (0.9709)
	250	−0.2479 (1.0537)	0.1347 (0.9781)	0.1322 (0.9888)
$F_X(x) = 0.50$	125	−0.0202 (0.0952)	0.0586 (0.1129)	0.0613 (0.1113)
	250	−0.0246 (0.0855)	0.0203 (0.0945)	0.0203 (0.0944)
$S_Y(y) = 0.50$	125	−0.0095 (0.1127)	0.0671 (0.1150)	0.0698 (0.1132)
	250	−0.0189 (0.0949)	0.0253 (0.0967)	0.0253 (0.0967)
Kendall's $\tau = -0.5$				
$\ln(\theta) = -5.746$	125	0.6197 (0.9742)	2.2066 (2.1359)	2.2502 (2.0987)
	250	0.3769 (0.9005)	1.9234 (2.1022)	1.9265 (2.0973)
$F_X(x) = 0.50$	125	0.0369 (0.0882)	0.1855 (0.1766)	0.1885 (0.1752)
	250	0.0242 (0.0741)	0.1644 (0.1726)	0.1646 (0.1724)
$S_Y(y) = 0.50$	125	0.0480 (0.0905)	0.1936 (0.1745)	0.1966 (0.1729)
	250	0.0292 (0.0757)	0.1670 (0.1698)	0.1673 (0.1696)

Each cell contains the average bias and standard error (in parenthesis) based on 200 runs. The functions $F_X(x)$ and $S_Y(y)$ are evaluated at $x = 0.462$ and $y = 1.386$, respectively, such that $F_X(x) = S_Y(y) = 0.5$.

modification to handle this problem, which is described in Appendix C. As long as the estimating equations yield proper solutions, the three competing estimators have similar performances.

We have seen that the performance of $\hat{F}_X$ and $\hat{S}_X$ with negative Kendall's tau is poorer than that with positive Kendall's tau (see Fig. 2 and Table 4). To explain this phenomenon, we compare two scatter plots for $(X_1, Y_1), \ldots, (X_{125}, Y_{125})$ between positive and negative Kendall's tau in Fig. 3. Note that, under truncation, the available data are on a trapezoid

$$\mathbf{T} = \{(x, y); x \le y, X_{(1)} \le x \le X_{(n)}, Y_{(1)} \le y \le Y_{(n)}\},$$

which corresponds to the shaded region in Fig. 3. It is shown that, under negative Kendall's tau, the trapezoid did not cover the upper tail of X and lower tail of Y. Hence, the truncation leads to some loss of information in these tails, which carries over the overall performance of $\hat{F}_X$ and $\hat{S}_Y$.

7. Conclusion and discussion

In this article, we proposed a likelihood-based inference procedure to analyze dependent truncation data. In comparison with the two moment-based estimating functions proposed by [6,13], the NPMLE method can fit more copula models and is naturally suitable for tied data. Under strongly negative dependent truncation, the NPMLE may be a more reliable alternative to the two competing approaches that sometimes fail to produce a proper solution. Furthermore, we obtain an analytic variance formula based on the inverse of the observed Fisher information matrix without relying on re-sampling techniques.

The proposed model selection procedure is appealing because many models can be included in the selection pool. Simulation results show that the proposed test statistic, as a deviance measure between the independence model and the imposed copula model, approximately follows the chi-squared distribution with one degree of freedom. However, a formal proof establishing the asymptotic null distribution of the deviance statistics is still missing. Another strategy to model selection is based on the information theoretic criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC). The proposed methods are implemented in the `R depend.truncation` package available at http://cran.r-project.org/.

The proposed NPMLE have some drawbacks, too. The extension of the method to incorporate right censoring appears to be challenging. On the other hand, the moment-based methods of [6,13] have straightforward extensions. Hence none of the three estimators dominate the others. The choice of methods depends on the data at hand. For example, the two

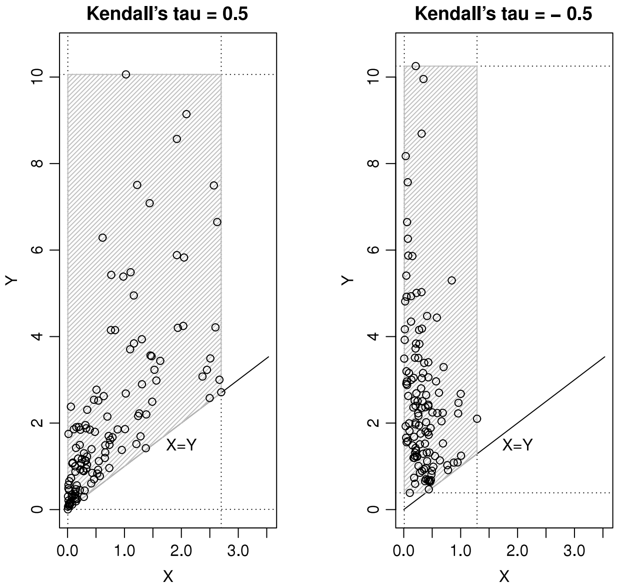

Fig. 3. Scatter plots for simulated data $\{(X_j, Y_j) : j = 1, \ldots, 125\}$ under the Frank copula model with positive association (Kendall's tau = 0.5, left panel) and negative association (Kendall's tau $= -0.5$, right panel). A trapezoidal region $\mathbf{T} = \{(x, y) : x \le y, X_{(1)} \le x \le X_{(n)}, Y_{(1)} \le y \le Y_{(n)}\}$ (shaded region) describes how data provide information for estimating the marginal functions F_X and S_Y. The ranges $[0, 3.53]$ for x-axis and $[0, 10.60]$ for y-axis are chosen such that $F_X(3.53) = 0.995$ and $1 - S_Y(10.60) = 0.995$, respectively.

moment-based methods for estimating θ under the semi-survival Clayton model are computationally appealing since they can be performed without estimating the marginal functions [6,13]. Section 4.2 of [6] also provides a simple explicit formula for the standard error.

Acknowledgments

The authors thank the Editor, Christian Genest, an Associate Editor and two anonymous referees for their helpful comments that improved the paper. They are also grateful to Lin Yi-Chun for her assistance with LaTeX. The second author was financially supported by Research Grant 98-2118-M-009-003-MY2 from the National Science Council of Taiwan, ROC.

Appendix A. Derivation of the score functions

The derivatives of $\ell_n(\theta, H_X, \Lambda_Y)$ in (5) with respect to $-\mathrm{d}H_X(X_i)$ and $\mathrm{d}\Lambda_Y(Y_i)$ are

$$\sum_{j=1}^{n} \mathbf{1}(X_j < X_i)\psi_\theta^{(1,0)}\{H_X(X_j), \Lambda_Y(Y_j-)\} + \frac{1}{-\mathrm{d}H_X(X_i)} - \frac{n}{c(\theta, H_X, \Lambda_Y)} \frac{\partial c(\theta, H_X, \Lambda_Y)}{\partial\{-\mathrm{d}H_X(X_i)\}},$$

$$\sum_{j=1}^{n} \mathbf{1}(Y_j > Y_i)\psi_\theta^{(0,1)}\{H_X(X_j), \Lambda_Y(Y_j-)\} + \frac{1}{\mathrm{d}\Lambda_Y(Y_i)} - \frac{n}{c(\theta,\ H_X, \Lambda_Y)} \frac{\partial c(\theta,\ H_X, \Lambda_Y)}{\partial \mathrm{d}\Lambda_Y(Y_i)},$$

where $\psi_\theta^{(1,0)}(x, y) = \partial \ln \eta_\theta(x, y)/\partial x$, $\psi_\theta^{(0,1)}(x, y) = \partial \ln \eta_\theta(x, y)/\partial y$,

$$\begin{aligned}\frac{\partial c(\theta,\ H_X, \Lambda_Y)}{\partial\{-\mathrm{d}H_X(X_i)\}} &= \sum_{j:X_j<X_i} \left[\sum_{\ell:Y_\ell \ge X_j} \eta_\theta^{(1,0)}\{H_X(X_j), \Lambda_Y(Y_\ell-)\}\mathrm{d}\Lambda_Y(Y_\ell)\right]\{-\mathrm{d}H_X(X_j)\} \\ &\quad + \sum_{\ell:Y_\ell \ge X_i} \eta_\theta\{H_X(X_i), \Lambda_Y(Y_\ell-)\}\mathrm{d}\Lambda_Y(Y_\ell) \\ &= \int_0^{X_i-}\int_{u-}^{\infty} \eta_\theta^{(1,0)}\{H_X(u), \Lambda_Y(v-)\}\mathrm{d}\Lambda_Y(v)\{-\mathrm{d}H_X(u)\} \\ &\quad + \int_{X_i-}^{\infty} \eta_\theta\{H_X(X_i), \Lambda_Y(v-)\}\mathrm{d}\Lambda_Y(v),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial c(\theta,\ H_X,\ \Lambda_Y)}{\partial \mathrm{d}\Lambda_Y(Y_i)} &= \sum_{j:Y_j>Y_i}\left[\sum_{\ell:X_\ell\le Y_j}\eta_\theta^{(0,1)}\{H_X(X_\ell),\ \Lambda_Y(Y_j-)\}\{-\mathrm{d}H_X(X_\ell)\}\right]\mathrm{d}\Lambda_Y(Y_j)\\ &\quad+\sum_{\ell:X_\ell\le Y_i}\eta_\theta\{H_X(X_\ell),\ \Lambda_Y(Y_i-)\}\{-\mathrm{d}H_X(X_\ell)\}\\ &= \int_{Y_i+}^{\infty}\int_0^{v}\eta_\theta^{(0,1)}\{H_X(u),\ \Lambda_Y(v-)\}\{-\mathrm{d}H_X(u)\}\mathrm{d}\Lambda_Y(v)\\ &\quad+\int_0^{Y_i}\eta_\theta\{H_X(u),\ \Lambda_Y(Y_i-)\}\{-\mathrm{d}H_X(u)\},\end{aligned}$$

where $\eta_\theta^{(1,0)}(x,y)=\partial\eta_\theta(x,y)/\partial x$ and $\eta_\theta^{(0,1)}(x,y)=\partial\eta_\theta(x,y)/\partial y$.

Using these formulas, the score equations become

$$1/\{-\mathrm{d}H_X(X_i)\}=\sum_{j=1}^{n}\Psi_j^{(1,0)}(X_i;\theta,\ H_X,\ \Lambda_Y),$$

$$1/\{\mathrm{d}\Lambda_Y(Y_i)\}=\sum_{j=1}^{n}\Psi_j^{(0,1)}(Y_i;\theta,H_X,\ \Lambda_Y),$$

where

$$\begin{aligned}\Psi_j^{(1,0)}(s;\theta,\ H_X,\ \Lambda_Y) &= c(\theta,\ H_X,\ \Lambda_Y)^{-1}\int_0^{s-}\int_{u-}^{\infty}\eta_\theta^{(1,0)}\{H_X(u),\ \Lambda_Y(v-)\}\mathrm{d}\Lambda_Y(v)\{-\mathrm{d}H_X(u)\}\\ &\quad+c(\theta,\ H_X,\ \Lambda_Y)^{-1}\int_{s-}^{\infty}\eta_\theta\{H_X(s),\ \Lambda_Y(v-)\}\mathrm{d}\Lambda_Y(v)\\ &\quad-\mathbf{1}(X_j<s)\psi_\theta^{(1,0)}\{H_X(X_j),\ \Lambda_Y(Y_j-)\},\end{aligned}$$

and

$$\begin{aligned}\Psi_j^{(0,1)}(s;\theta,\ H_X,\ \Lambda_Y) &= c(\theta,\ H_X,\ \Lambda_Y)^{-1}\int_{s+}^{\infty}\int_0^{v}\eta_\theta^{(0,1)}\{H_X(u),\ \Lambda_Y(v-)\}\{-\mathrm{d}H_X(u)\}\mathrm{d}\Lambda_Y(v)\\ &\quad+c(\theta,\ H_X,\ \Lambda_Y)^{-1}\int_0^{s}\eta_\theta\{H_X(u),\ \Lambda_Y(s-)\}\{-\mathrm{d}H_X(u)\}\\ &\quad-\mathbf{1}(Y_j>s)\psi_\theta^{(0,1)}\{H_X(X_j),\ \Lambda_Y(Y_j-)\}.\end{aligned}$$

Now we derive the expectation of $\Psi_j^{(1,0)}$ and $\Psi_j^{(0,1)}$ under the true parameters. Observe that

$$E[\mathbf{1}(X_j<s)\psi_{\theta^0}^{(1,0)}\{H_X^0(X_j),\ \Lambda_Y^0(Y_j-)\}]=\frac{1}{c(\theta^0,H_X^0,\ \Lambda_Y^0)}\int_0^{s-}\int_{u-}^{\infty}\eta_{\theta^0}^{(1,0)}\{H_X^0(u),\ \Lambda_Y^0(v-)\}\mathrm{d}\Lambda_Y^0(v)\{-\mathrm{d}H_X^0(u)\},$$

and

$$E[\mathbf{1}(Y_j>s)\psi_{\theta^0}^{(0,1)}\{H_X^0(X_j),\ \Lambda_Y^0(Y_j-)\}]=\frac{1}{c(\theta^0,H_X^0,\ \Lambda_Y^0)}\int_{s+}^{\infty}\int_0^{v}\eta_{\theta^0}^{(0,1)}\{H_X^0(u),\ \Lambda_Y^0(v-)\}\{-\mathrm{d}H_X^0(u)\}\mathrm{d}\Lambda_Y^0(v).$$

Hence,

$$\begin{aligned}E\{\Psi_j^{(1,0)}(s;\theta^0,H_X^0,\ \Lambda_Y^0)\} &= \frac{\int_{s-}^{\infty}\eta_{\theta^0}\{H_X^0(s),\ \Lambda_Y^0(v-)\}\{\mathrm{d}\Lambda_Y^0(v)\}}{c(\theta^0,H_X^0,\ \Lambda_Y^0)}\\ &= \frac{e^{-H_X^0(s)}C_{\theta^0}^{(1,0)}\{e^{-H_X^0(s)},\ e^{-\Lambda_Y^0(s-)}\}}{c(\theta^0,H_X^0,\ \Lambda_Y^0)},\end{aligned}$$

and

$$\begin{aligned}E\{\Psi_j^{(0,1)}(s;\theta^0,H_X^0,\ \Lambda_Y^0)\} &= \frac{\int_0^{s}\eta_{\theta^0}\{H_X^0(u),\ \Lambda_Y^0(s-)\}\{-\mathrm{d}H_X^0(u)\}}{c(\theta^0,H_X^0,\ \Lambda_Y^0)}\\ &= \frac{e^{-\Lambda_Y^0(s-)}C_{\theta^0}^{(0,1)}\{e^{-H_X^0(s)},\ e^{-\Lambda_Y^0(s-)}\}}{c(\theta^0,H_X^0,\ \Lambda_Y^0)}.\end{aligned}$$

Appendix B. Details of the asymptotic results

B.1. Proof of consistency

We aim to show that $\hat{\theta} \to \theta^0$, $\hat{H}_X(x) \to H_X^0(x)$, and $\hat{\Lambda}_Y(y) \to \Lambda_Y^0(y)$ almost surely as $n \to \infty$. The proof consists of the following three steps.

Step 1: We show that $-\mathrm{d}\hat{H}_X(X_j) < \infty$, $\hat{\Lambda}_Y(Y_j) < \infty$ for all $j \in \{1, \ldots, n\}$. By definition, $-\mathrm{d}\hat{H}_X(X_{(1)}) = 1 < \infty$. Suppose that $-\mathrm{d}\hat{H}_X(X_{(i)}) = \infty$ and that $0 < -\mathrm{d}\hat{H}_X(X_j) < \infty$ for $j \neq i$ and $0 < -\mathrm{d}\hat{\Lambda}_X(X_j) < \infty$ for all j. Then, the $(i-1)$th contribution to the log-likelihood is

$$\begin{aligned}\ell_{j=(i-1)}(\theta, H_X, \Lambda_Y) &= \ln C_\theta^{(1,1)}\{0, e^{-\Lambda_Y(Y_{(i-1)}-)}\} - \Lambda_Y(Y_{(i-1)}-) + \ln \mathrm{d}\Lambda_Y(Y_{(i-1)}) - H_X(X_{(i-1)}) - \ln\{c(\theta, H_X, \Lambda_Y)\}\\ &\leq \ln(M) - \Lambda_Y(Y_{(i-1)}-) + \ln\{\mathrm{d}\Lambda_Y(Y_{(i-1)})\} + \ln\{e^{-H_X(X_{(i-1)})}/c(\theta, H_X, \Lambda_Y)\},\end{aligned}$$

for some $M > 0$. From $\hat{H}_X(X_{(i-1)}) \to \infty$ and Assumption II, one has

$$e^{-\mathrm{d}\hat{H}_X(X_{(i-1)})}/c(\theta, \hat{H}_X, \hat{\Lambda}_Y) \to 0.$$

This implies $\ell_{j=(i-1)}(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y) \leq -\infty$. Therefore, $-\mathrm{d}\hat{H}_X(X_j) = \infty$ cannot be the NPMLE. Thus $-\mathrm{d}\hat{H}_X(X_{(i)}) < \infty$ for all $i \in \{1, \ldots, n\}$. In the same way, one has $0 < \mathrm{d}\hat{\Lambda}_Y(Y_j) < \infty$ for all $j \in \{1, \ldots, n\}$.

Step 2: We show that there exist some convergent subsequences of $\hat{H}_X(x)$, $\hat{\Lambda}_Y(y)$ and $\hat{\theta}$ such that $\hat{H}_X(x) \to H_X^*(x)$, $\hat{\Lambda}_Y(y) \to \Lambda_Y^*(y)$ and $\hat{\theta} \to \theta^*$ almost surely, where $H_X^*(x)$ and $\Lambda_Y^*(y)$ are absolutely continuous and differentiable.

For any bounded sequences $\overline{H}_X^{(n)}(x)$, $\overline{\Lambda}_Y^{(n)}(y)$ and $\overline{\theta}^{(n)}$, it can be shown that $\ell_n(\overline{\theta}^{(n)}, \overline{H}_X^{(n)}, \overline{\Lambda}_Y^{(n)})/n$ is stochastically bounded. It follows that

$$\ell_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)/n \geq O_p(1). \tag{B.1}$$

If $\hat{H}_X(x) \to \infty$ for some subsequence, then for some constant $M > 0$,

$$\begin{aligned}\frac{1}{n}\,\ell_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y) &= \frac{1}{n}\sum_j \ln\left[\frac{C_\theta^{(1,1)}\{e^{-\hat{H}_X(X_j)}, e^{-\hat{\Lambda}_Y(Y_j-)}\}e^{-\hat{H}_X(X_j)}e^{-\hat{\Lambda}_Y(Y_j-)}\mathrm{d}\hat{\Lambda}_Y(Y_j)\{-\mathrm{d}\hat{H}_X(X_j)\}}{c(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)}\right]\\ &\leq \frac{1}{n}\sum_j \ln\left\{\frac{Me^{-\hat{\Lambda}_Y(Y_j-)}\mathrm{d}\hat{\Lambda}_Y(Y_j)}{\varepsilon}\,\frac{\hat{H}_X(X_j)}{e^{\hat{H}_X(X_j)}}\right\} \to -\infty.\end{aligned}$$

This contradicts (B.1). Therefore, $\sup_n \hat{H}_X(x) \to \infty$ holds almost surely. By Helly's Selection Theorem, there exists a subsequence such that $\hat{H}_X(x) \to H_X^*(x)$ almost surely. The same arguments can be applied to prove $\hat{\Lambda}_Y(y) \to \Lambda_Y^*(y)$ and $\hat{\theta} \to \theta^*$.

Step 3: We show $(\theta^*, H_X^*, \Lambda_Y^*) = (\theta^0, H_X^0, \Lambda_Y^0)$. Let $\tilde{H}_X(x)$ and $\tilde{\Lambda}_Y(y)$ be defined as (9a) and (9b), respectively. By the Glivenko–Cantelli Theorem, $\tilde{H}_X(x) \to H_X(x)$ and $\tilde{\Lambda}_Y(y) \to \Lambda_Y(y)$ almost surely. Clearly,

$$\ell_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)/n \geq \ell_n(\theta, \tilde{H}_X, \tilde{\Lambda}_Y)/n.$$

By applying the Strong Law of Large Numbers on both sides, we have

$$E\left(\ln\left[\frac{\eta_\theta\{H_X^*(X), \Lambda_Y^*(Y-)\}}{c(\theta^*, H_X^*, \Lambda_Y^*)}h_X^*(X)\lambda_Y^*(Y)\right]\right) \geq E\left(\ln\left[\frac{\eta_\theta\{H_X^0(X), \Lambda_Y^0(Y-)\}}{c(\theta, H_X^0, \Lambda_Y^0)}h_X^0(X)\lambda_Y^0(Y)\right]\right).$$

From Assumption III, $(\theta^*, H_X^*, \Lambda_Y^*) = (\theta^0, H_X^0, \Lambda_Y^0)$.

B.2. Proof of asymptotic normality

Let P_n be the empirical measure based on the independent and identically distributed random variables, viz.

$$\mathrm{d}P_n(x, y) = \frac{1}{n}\sum_{j=1}^n \mathbf{1}(X_j = x, Y_j = y).$$

Denote the log-likelihood function of a single subject by $\ell(\theta, H_X, \Lambda_Y)$. Let also

$$c^0 = c(\theta^0, H_X^0, \Lambda_Y^0), \qquad \dot{c}^0 = \partial c(\theta, H_X^0, \Lambda_Y^0)/\partial\theta|_{\theta^0}, \qquad \ddot{c}^0 = \partial^2 c(\theta, H_X^0, \Lambda_Y^0)/\partial\theta^2|_{\theta^0}.$$

Define the derivative functions

$$\dot{\ell}(\theta, H_X, \Lambda_Y) = \partial \ell(\theta, H_X, \Lambda_Y)/\partial\theta,$$
$$\ell_{H_X}(\theta, H_X, \Lambda_Y)[\Delta H_X] = \lim_{\varepsilon\to 0}\{\ell(\theta, H_X + \varepsilon\Delta H_X, \Lambda_Y) - \ell(\theta, H_X, \Lambda_Y)\}/\varepsilon,$$
$$\dot{\ell}_{H_X}(\theta, H_X, \Lambda_Y)[\Delta H_X] = \lim_{\varepsilon\to 0}\{\dot{\ell}(\theta, H_X + \varepsilon\Delta H_X, \Lambda_Y) - \dot{\ell}(\theta, H_X, \Lambda_Y)\}/\varepsilon,$$

and

$$\ell_{H_X H_X}(\theta, H_X, \Lambda_Y)[\Delta_1 H_X, \Delta_2 H_X] = \lim_{\varepsilon\to 0}\{\ell_{H_X}(\theta, H_X + \varepsilon\Delta_2 H_X, \Lambda_Y)[\Delta_1 H_X] - \ell_{H_X}(\theta, H_X, \Lambda_Y)[\Delta_1 H_X]\}/\varepsilon.$$

Notations for $\ell_{\Lambda_Y}(\theta, H_X, \Lambda_Y)[\Delta\Lambda_Y]$, $\dot{\ell}_{\Lambda_Y}(\theta, H_X, \Lambda_Y)[\Delta\Lambda_Y]$, $\ell_{\Lambda_Y\Lambda_Y}(\theta, H_X, \Lambda_Y)[\Delta_1\Lambda_Y, \Delta_2\Lambda_Y]$, and $\ell_{H_X\Lambda_Y}(\theta, H_X, \Lambda_Y)[\Delta H_X, \Delta\Lambda_Y]$ are defined similarly.

The above derivatives can be used to characterize the NPMLE as a solution to the likelihood equations. First, we define the derivative of H_X. Suppose that $(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)$ is the NPMLE that maximizes $\ell_n(\theta, H_X, \Lambda_Y)$ in (5). Clearly, for any $\Delta H_X = \Delta H_X(\cdot)$,

$$\ell_n(\hat{\theta}, \hat{H}_X + \varepsilon\Delta\hat{H}_X, \hat{\Lambda}_Y) \le \ell_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y).$$

This implies that $\partial\ell_n(\hat{\theta}, \hat{H}_X + \varepsilon\Delta H_X, \hat{\Lambda}_Y)/\partial\varepsilon|_{\varepsilon=0} = 0$ in any ΔH_X. Let Q be the set of all bounded functions $h : [0, \infty) \to [0, 1]$. For any $p_X \in Q$ and $q_Y \in Q$, define

$$\begin{aligned}
&W_n(\theta, H_X, \Lambda_Y)[p_X, q_Y] \\
&= P_n\begin{bmatrix} \dot{\ell}(\theta, H_X, \Lambda_Y) \\ \ell_{H_X}(\theta, H_X, \Lambda_Y)\left[\int_{\bullet}^{\infty} p_X(u)\{-\mathrm{d}H_X(u)\}\right] \\ \ell_{\Lambda_Y}(\theta, H_X, \Lambda_Y)\left[\int_0^{\bullet} q_Y(u)\{\mathrm{d}\Lambda_Y(u)\}\right] \end{bmatrix} \\
&= \frac{1}{n}\sum_j \begin{bmatrix} \dfrac{\dot{\eta}_\theta}{\eta_\theta}\{H_X(X_j), \Lambda_Y(Y_j-)\} - \dfrac{\dot{c}}{c}(\theta, H_X, \Lambda_Y) \\ \psi_\theta^{(1,0)}\{H_X(X_j), \Lambda_Y(Y_j-)\}\displaystyle\int_{X_j}^{\infty} p_X(u)\{-\mathrm{d}H_X(u)\} + p_X(X_j) - \dfrac{c_{H_X}}{c}(\theta, H_X, \Lambda_Y) \\ \psi_\theta^{(0,1)}\{H_X(X_j), \Lambda_Y(Y_j-)\}\displaystyle\int_0^{Y_j} q_Y(u)\mathrm{d}\Lambda_Y(u) + q_Y(Y_j) - \dfrac{c_{\Lambda_Y}}{c}(\theta, H_X, \Lambda_Y) \end{bmatrix}
\end{aligned}$$

where

$$\begin{aligned}
c_{H_X}(\theta, H_X, \Lambda_Y) \;=\; & \iint_{x\le y}\Bigg[\eta_\theta^{(1,0)}\{H_X(x), \Lambda_Y(y-)\}\int_x^{\infty} p_X(u)\{-\mathrm{d}H_X(u)\} \\
& + \eta_\theta\{H_X(x), \Lambda_Y(y-)\}p_X(x)\Bigg]\{-\mathrm{d}H_X(x)\}\mathrm{d}\Lambda_Y(y), \\
c_{\Lambda_Y}(\theta, H_X, \Lambda_Y) \;=\; & \iint_{x\le y}\Bigg[\eta_\theta^{(0,1)}\{H_X(x), \Lambda_Y(y-)\}\int_0^{y} q_Y(u)\mathrm{d}\Lambda_Y(u) \\
& + \eta_\theta\{H_X(x), \Lambda_Y(y-)\}q_Y(y)\Bigg]\{-\mathrm{d}H_X(x)\}\mathrm{d}\Lambda_Y(y).
\end{aligned}$$

Replacing P_n by P^0 in these expressions, we can define $W(\theta, H_X, \Lambda_Y)[p_X, q_Y]$. Then, it follows that $W(\theta^0, H_X^0, \Lambda_Y^0)[p_X, q_Y] = 0 \in \mathbb{R}^{p+2}$ for any (p_X, q_Y). To see this, consider, e.g., the first p components of $W(\theta, H_X, \Lambda_Y)[p_X, q_Y]$; they are of the form

$$\iint_{x\le y}\frac{\dot{\eta}_\theta}{\eta_\theta}\{H_X(x), \Lambda_Y(y-)\}\mathrm{d}P^0(x, y) - \frac{\dot{c}}{c}(\theta, H_X, \Lambda_Y),$$

and vanish at $(\theta, H_X, \Lambda_Y) = (\theta^0, H_X^0, \Lambda_Y^0)$. Furthermore, by the definition of the NPMLE, $W_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)[p_X, q_Y] = 0 \in \mathbb{R}^{p+2}$ for any (p_X, q_Y). Now we can regard $W_n(\theta, H_X, \Lambda_Y)$ (and $W(\theta, H_X, \Lambda_Y)$) as an element in $\mathbb{R}^p \times \{\ell^\infty(Q)\}^2$, where the argument (p_X, q_Y) is omitted and $\ell^\infty(Q)$ is the set of all uniformly bounded real functions on Q. Consider the following maps:

$$W_n(\theta, H_X, \Lambda_Y) : \Theta \to \mathbb{R}^p \times \{\ell^\infty(Q)\}^2, \qquad W(\theta, H_X, \Lambda_Y) : \Theta \to \mathbb{R}^p \times \{\ell^\infty(Q)\}^2.$$

Then, the NPMLE is regarded as a Z-estimator in $\mathbb{R}^p \times \{\ell^\infty(Q)\}^2$; see p. 309 of [19]. By the Central Limit Theorem, $n^{1/2}\{W_n(\theta^0, H_X^0, \Lambda_Y^0) - W(\theta^0, H_X^0, \Lambda_Y^0)\}$ converges weakly to a Gaussian random element $Z \in \mathbb{R}^p \times \{\ell^\infty(Q)\}^2$, indexed by

(p_X, q_Y). To apply Theorem 3.3.1 of [19], $W(\theta, H_X, \Lambda_Y)$ must be Fréchet-differentiable at $(\theta^0, H_X^0, \Lambda_Y^0)$ with a continuously invertible derivative. This suffices to find a map, called the information operator, $\dot{W} : \Theta \to \mathbb{R}^p \times \{\ell^\infty(Q)\}^2$ so that with the norm in $\mathbb{R}^p \times \{\ell^\infty(Q)\}^2$,

$$\begin{aligned}&\|W(\theta, H_X, \Lambda_Y) - W(\theta^0, H_X^0, \Lambda_Y^0) - \dot{W}(\theta - \theta^0, H_X - H_X^0, \Lambda_Y - \Lambda_Y^0)\|_{\mathbb{R}^p\times\{\ell^\infty(Q)\}^2}\\ &\quad = o(|\theta-\theta^0| + \|H_X - H_X^0\| + \|\Lambda_Y - \Lambda_Y^0\|).\end{aligned}$$

By analytical calculations, it follows that

$$\dot{W}(\theta - \theta^0, H_X - H_X^0, \Lambda_Y - \Lambda_Y^0)[p_X, q_Y] = P^0 \begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{bmatrix}$$

where

$$\begin{aligned}\lambda_1 &= \ddot{\ell}(\theta-\theta^0) + \dot{\ell}_{H_X}[H_X - H_X^0] + \dot{\ell}_{\Lambda_Y}[\Lambda_Y - \Lambda_Y^0],\\
\lambda_2 &= \dot{\ell}_{H_X}\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right](\theta-\theta^0) + \ell_{H_XH_X}\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}, H_X - H_X^0\right]\\
&\quad + \ell_{H_X\Lambda_Y}\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}, \Lambda_Y - \Lambda_Y^0\right],\\
\lambda_3 &= \dot{\ell}_{\Lambda_Y}\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right](\theta-\theta^0) + \ell_{\Lambda_YH_X}\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0, H_X - H_X^0\right]\\
&\quad + \ell_{\Lambda_Y\Lambda_Y}\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0, \Lambda_Y - \Lambda_Y^0\right].\end{aligned}$$

By Assumption IV, $\dot{W} : \Theta \to \mathbb{R}^p \times \{\ell^\infty(Q)\}^2$ is continuously invertible. Therefore, in light of Theorem 3.3.1 in [19], we have

$$n^{1/2}\dot{W}(\hat\theta-\theta^0, \hat{H}_X - H_X^0, \hat\Lambda_Y - \Lambda_Y^0) = n^{1/2}P_n\begin{bmatrix}\dot{\ell}(\theta^0, H_X^0, \Lambda_Y^0)\\ \ell_{H_X}(\theta^0, H_X^0, \Lambda_Y^0)\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right]\\ \ell_{\Lambda_Y}(\theta^0, H_X^0, \Lambda_Y^0)\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right]\end{bmatrix} + o_P(1). \tag{B.2}$$

Define two step functions: p_X that jumps at observed point $X_{(j)}$ with $p_X(X_{(j)}) = p_{(j)}$ and q_Y that jumps at observed point $Y_{(j)}$ with $q_Y(Y_{(j)}) = q_{(j)}$. Let $\vec{\Delta}_X$ be the vector of size $n-1$ consisting of the ordered values of $p_{(j)}\{-\mathrm{d}\hat{H}_X(X_{(j)})\}$ for $j \in \{2, \dots, n\}$. Further let $\vec{\Delta}_Y$ be another vector of size $n-1$ consisting of the ordered values of $q_{(j)}d\hat\Lambda_Y(Y_{(j)})$.

Now let $\tilde{H}_X^0$ and $\tilde\Lambda_Y^0$ be the step functions that jump at X_j and Y_j with $\tilde{H}_X^0(X_j) = H_X^0(X_j)$, $\tilde\Lambda_Y^0(Y_j) = \Lambda_Y^0(Y_j)$, $-\mathrm{d}\tilde{H}_X^0(X_{(1)}) = 1$ and $\mathrm{d}\tilde\Lambda_Y^0(Y_{(n)}) = 1$. Then

$$n^{1/2}\left((\hat\theta-\theta^0)^\top, \left\{-\mathrm{d}(\hat{H}_X - \tilde{H}_X^0)(X_{(j)})\right\}_{j=2}^n, \left\{\mathrm{d}(\hat\Lambda_Y - \tilde\Lambda_Y^0)(Y_{(j)})\right\}_{j=1}^{n-1}\right)\frac{i_n(\hat\theta, \hat{H}_X, \hat\Lambda_Y)}{n}(a^\top, \vec{\Delta}_X^\top, \vec{\Delta}_Y^\top)^\top.$$

This expression further equals

$$\begin{aligned}-n^{1/2}P_n\Bigg(&(\hat\theta-\theta^0)^\top\hat{\ddot{\ell}}a + \hat{\dot{\ell}}_{H_X}[\hat{H}_X - \tilde{H}_X^0]^\top a + \hat{\dot{\ell}}_{\Lambda_Y}[\hat\Lambda_Y - \tilde\Lambda_Y^0]^\top a\\
&+ (\hat\theta-\theta^0)^\top\hat{\dot{\ell}}_{H_X}\left[\int_\bullet^\infty p_X\{-\mathrm{d}\hat{H}_X\}\right] + \hat{\ell}_{H_XH_X}\left[\hat{H}_X - \tilde{H}_X^0, \int_\bullet^\infty p_X\{-\mathrm{d}\hat{H}_X\}\right]\\
&+ \hat{\ell}_{\Lambda_YH_X}\left[\hat\Lambda_Y - \tilde\Lambda_Y^0, \int_\bullet^\infty p_X\{-\mathrm{d}\hat{H}_X\}\right] + (\hat\theta-\theta^0)^\top\hat{\dot{\ell}}_{\Lambda_Y}\left[\int_0^\bullet q_Y\mathrm{d}\hat\Lambda_Y\right]\\
&+ \hat{\ell}_{H_X\Lambda_Y}\left[\hat{H}_X - \tilde{H}_X^0, \int_0^\bullet q_Y\mathrm{d}\hat\Lambda_Y\right] + \hat{\ell}_{\Lambda_Y\Lambda_Y}\left[\hat\Lambda_Y - \tilde\Lambda_Y^0, \int_0^\bullet q_Y\mathrm{d}\hat\Lambda_Y\right]\Bigg),\end{aligned}$$

which in turn can be rewritten as

$$-n^{1/2}P^0\Bigg((\hat{\theta}-\theta^0)^\top\ddot{\ell}a+\dot{\ell}_{H_X}[\hat{H}_X-H_X^0]^\top a+\dot{\ell}_{\Lambda_Y}[\hat{\Lambda}_Y-\Lambda_Y^0]^\top a$$
$$+(\hat{\theta}-\theta^0)^\top\dot{\ell}_{H_X}\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right]+\ell_{H_XH_X}\left[\hat{H}_X-H_X^0,\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right]$$
$$+\ell_{\Lambda_YH_X}\left[\hat{\Lambda}_Y-\Lambda_Y^0,\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right]+(\hat{\theta}-\theta^0)^\top\dot{\ell}_{\Lambda_Y}\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right]$$
$$+\ell_{H_X\Lambda_Y}\left[\hat{H}_X-H_X^0,\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right]+\ell_{\Lambda_Y\Lambda_Y}\left[\hat{\Lambda}_Y-\Lambda_Y^0,\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right]\Bigg)+o_P(1)$$

and finally simplifies to

$$n^{1/2}(a^\top,\ 1,\ 1)\dot{W}(\hat{\theta}-\theta^0,\hat{H}_X-H_X^0,\hat{\Lambda}_Y-\Lambda_Y^0)+o_P(1),$$

where $\hat{\ddot{\ell}}\equiv\ddot{\ell}(\hat{\theta},\hat{H}_X,\hat{\Lambda}_Y)$, $\hat{\dot{\ell}}_{H_X}[\cdot]\equiv\dot{\ell}_{H_X}(\hat{\theta},\hat{H}_X,\hat{\Lambda}_Y)[\cdot]$ and so on. Together with (B.2),

$$n^{1/2}\left((\hat{\theta}-\theta^0)^\top,\left\{-\mathrm{d}(\hat{H}_X-\tilde{H}_X^0)(X_{(j)})\right\}_{j=2}^n,\left\{\mathrm{d}(\hat{\Lambda}_Y-\tilde{\Lambda}_Y^0)(Y_{(j)})\right\}_{j=1}^{n-1}\right)\frac{i_n(\hat{\theta},\hat{H}_X,\hat{\Lambda}_Y)}{n}(a^\top,\vec{\Delta}_X^\top,\vec{\Delta}_Y^\top)^\top$$

equals

$$(a^\top,1,1)n^{1/2}P_n\begin{bmatrix}\dot{\ell}(\theta^0,H_X^0,\Lambda_Y^0)\\ \ell_{H_X}(\theta^0,H_X^0,\Lambda_Y^0)\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right]\\ \ell_{\Lambda_Y}(\theta^0,H_X^0,\Lambda_Y^0)\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right]\end{bmatrix}+o_P(1).$$

For $b\in\mathbb{R}^p$, $W_X=(w_X(X_{(2)}),\ldots,w_X(X_{(n)}))^\top$, and $W_Y=(w_Y(Y_{(1)}),\ldots,w_Y(Y_{(n-1)}))^\top$, we choose $a\in\mathbb{R}^p$, p_X, and q_Y such that

$$\frac{i_n(\hat{\theta},\hat{H}_X,\hat{\Lambda}_Y)}{n}(a^\top,\vec{\Delta}_X^\top,\vec{\Delta}_Y^\top)^\top=(b^\top,W_X^\top,W_Y^\top)^\top.$$

Therefore,

$$n^{1/2}\left[b^\top(\hat{\theta}-\theta^0)+\int_0^\infty w_X(u)\{-\mathrm{d}\hat{H}_X(u)+\mathrm{d}H_X^0(u)\}+\int_0^\infty w_Y(u)\{\mathrm{d}\hat{\Lambda}_Y(u)-\mathrm{d}\Lambda_Y^0(u)\}\right]$$

equals

$$n^{1/2}\left[b^\top(\hat{\theta}-\theta^0)+\int_0^\infty w_X(u)\{-\mathrm{d}\hat{H}_X(u)+\mathrm{d}\tilde{H}_X^0(u)\}+\int_0^\infty w_Y(u)\{\mathrm{d}\hat{\Lambda}_Y(u)-\mathrm{d}\tilde{\Lambda}_Y^0(u)\}\right]+o_P(1).$$

This expression further equals

$$n^{1/2}\left[b^\top(\hat{\theta}-\theta^0)+\sum_{j=2}^n w_X(X_j)\{-\mathrm{d}\hat{H}_X(X_j)+\mathrm{d}\tilde{H}_X^0(X_j)\}+\sum_{j=1}^{n-1}w_Y(Y_j)\{\mathrm{d}\hat{\Lambda}_Y(Y_j)-\mathrm{d}\tilde{\Lambda}_Y^0(Y_j)\}\right]+o_P(1)$$
$$=n^{1/2}\left((\hat{\theta}-\theta^0)^\top,\left\{-\mathrm{d}(\hat{H}_X-\tilde{H}_X^0)(X_{(j)})\right\}_{j=2}^n,\left\{\mathrm{d}(\hat{\Lambda}_Y-\tilde{\Lambda}_Y^0)(Y_{(j)})\right\}_{j=1}^{n-1}\right)\times(b^\top,W_X^\top,W_Y^\top)^\top+o_P(1)$$
$$=(a^\top,1,1)n^{1/2}P_n\begin{bmatrix}\dot{\ell}(\theta^0,H_X^0,\Lambda_Y^0)\\ \ell_{H_X}(\theta^0,H_X^0,\Lambda_Y^0)\left[\int_\bullet^\infty p_X\{-\mathrm{d}H_X^0\}\right]\\ \ell_{\Lambda_Y}(\theta^0,H_X^0,\Lambda_Y^0)\left[\int_0^\bullet q_Y\mathrm{d}\Lambda_Y^0\right]\end{bmatrix}+o_P(1).$$

From (B.2), the above random variable converges weakly to a zero mean Gaussian distribution with asymptotic variance

$$(a^{\top}, 1, 1)P^{0}\begin{bmatrix} \dot{\ell}(\theta^{0}, H_X^{0}, \Lambda_Y^{0}) \\ \ell_{H_X}(\theta^{0}, H_X^{0}, \Lambda_Y^{0})\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}\right] \\ \ell_{\Lambda_Y}(\theta^{0}, H_X^{0}, \Lambda_Y^{0})\left[\int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}\right] \end{bmatrix}^{\otimes 2}\begin{bmatrix} a \\ 1 \\ 1 \end{bmatrix}.$$

The latter can be expressed as

$$-P^{0}\Bigg(a^{\top}\ddot{\ell}a + a\dot{\ell}_{H_X}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}\right] + a\dot{\ell}_{\Lambda_Y}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}\right] + a\dot{\ell}_{H_X}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}\right]$$
$$+ \ell_{H_X H_X}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}, \int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}\right] + \ell_{\Lambda_Y H_X}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}, \int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}\right]$$
$$+ a\dot{\ell}_{\Lambda_Y}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}\right] + \ell_{H_X \Lambda_Y}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}H_X^{0}\}, \int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}\right] + \ell_{\Lambda_Y \Lambda_Y}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}, \int_{0}^{\bullet} q_Y \mathrm{d}\Lambda_Y^{0}\right]\Bigg).$$

This is estimated by

$$-P_n\Bigg(a^{\top}\hat{\ddot{\ell}}a + \hat{\dot{\ell}}_{H_X}^{\top}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}\hat{H}_X\}\right]a + \hat{\dot{\ell}}_{\Lambda_Y}^{\top}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\hat{\Lambda}_Y\right]a$$
$$+ a^{\top}\hat{\dot{\ell}}_{H_X}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}\hat{H}_X\}\right] + \hat{\ell}_{H_X H_X}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}\hat{H}_X\}, \int_{\bullet}^{\infty} p_X\{-\mathrm{d}\hat{H}_X\}\right]$$
$$+ \hat{\ell}_{\Lambda_Y H_X}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\hat{\Lambda}_Y, \int_{\bullet}^{\infty} p_X\{-\mathrm{d}\hat{H}_X\}\right] + a^{\top}\hat{\dot{\ell}}_{\Lambda_Y}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\hat{\Lambda}_Y\right]$$
$$+ \hat{\ell}_{H_X \Lambda_Y}\left[\int_{\bullet}^{\infty} p_X\{-\mathrm{d}\hat{H}_X\}, \int_{0}^{\bullet} q_Y \mathrm{d}\hat{\Lambda}_Y\right] + \hat{\ell}_{\Lambda_Y \Lambda_Y}\left[\int_{0}^{\bullet} q_Y \mathrm{d}\hat{\Lambda}_Y, \int_{0}^{\bullet} q_Y \mathrm{d}\hat{\Lambda}_Y\right]\Bigg).$$

This expression can also be written as

$$(a^{\top}, \vec{\Delta}_X^{\top}, \vec{\Delta}_Y^{\top})\left\{\frac{1}{n} i_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)\right\}(a^{\top}, \vec{\Delta}_X^{\top}, \vec{\Delta}_Y^{\top}) = (b^{\top}, W_X^{\top}, W_Y^{\top})\left\{\frac{1}{n} i_n(\hat{\theta}, \hat{H}_X, \hat{\Lambda}_Y)\right\}^{-1}(b^{\top}, W_X^{\top}, W_Y^{\top})^{\top}.$$

Appendix C. Numerical modifications for the two competing estimators

In our simulations, the estimating equations for (θ, c) used in [6,13] sometimes fail to produce a proper solution when data are generated from the semi-survival Frank model with negative associations. This occurs as the estimating equations do not have a zero solution for any value of (θ, c). By adopting an idea of [6], the estimating equation for (θ, c) is modified as

$$U_c(\theta, c) = \sum_{j:\, t_1 < x_j}\left[\phi_\theta\left\{c\,\frac{\tilde{R}(x_j)}{n}\right\} - \phi_\theta\left\{c\,\frac{\tilde{R}(x_j) - 1}{n}\right\}\right]\mathbf{1}\{\tilde{R}(x_j) \geq bn^{a}\} + \phi_\theta\left(\frac{c}{n}\right),$$

where $t_1 = X_{(1)}$, and $0 < a < 1$ and $b > 0$ are arbitrary tuning parameters. The tuning parameters can be set as $a = 1/10$ and $b = 1$ for usual cases. When the estimating equations do not yield any solution, we set $a = 1.5/10$ and $b = 1$. In this way, the estimating equations always produce a solution. Before the modification, the percentages lacking a solution are 11.5% ($n = 125$) and 6.5% ($n = 250$) for $\ln(\theta) = -2.38$; they are 33.25% ($n = 125$) and 58.5% ($n = 250$) for $\ln(\theta) = -5.746$. The modified estimating equation still produces unbiased results when $\ln(\theta) = -2.380$, but it is somewhat biased when $\ln(\theta) = -5.746$.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2012.03.012.

References

[1] D. Beaudoin, L. Lakhal Chaieb, Archimedean copula model selection under dependent truncation, Stat. Med. 27 (2008) 4440–4454.
[2] Y.-H. Chen, Semiparametric marginal regression analysis for dependent competing risks under an assumed copula, J. R. Stat. Soc. Ser. B 72 (2010) 235–251.
[3] C.-H. Chen, W.-Y. Tsai, W.-H. Chao, The product-moment correlation coefficient and linear regression for truncated data, J. Amer. Statist. Assoc. 91 (1996) 1181–1186.
[4] T. Emura, Y. Konno, A goodness-of-fit test for parametric models based on dependently truncated data, Comput. Statist. Data Anal. 56 (2012) 2237–2250.
[5] T. Emura, W. Wang, Testing quasi-independence for truncation data, J. Multivariate Anal. 101 (2010) 223–239.
[6] T. Emura, W. Wang, H.-N. Hung, Semi-parametric inference for copula models for truncated data, Statist. Sinica 21 (2011) 349–367.
[7] C. Genest, Frank's family of bivariate distributions, Biometrika 74 (1987) 549–555.
[8] C. Genest, J. Nešlehová, A primer on copulas for count data, Astin Bull. 37 (2007) 475–515.
[9] C. Genest, J. Nešlehová, M. Ruppert, Comment on "Statistical models and methods for dependence in insurance data", by S. Haug, C. Klüppelberg, L. Peng, J. Korean Statist. Soc. 40 (2011) 141–148.
[10] J.D. Kalbfleisch, J.F. Lawless, Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS, J. Amer. Statist. Assoc. 84 (1989) 360–372.
[11] J.P. Klein, M.L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, second ed., Springer, Berlin, 2003.
[12] S.W. Lagakos, L.M. Barraj, V. de Gruttola, Non-parametric analysis of truncated survival data, with application to AIDS, Biometrika 75 (1988) 515–523.
[13] L. Lakhal Chaieb, L.-P. Rivest, B. Abdous, Estimating survival under a dependent truncation, Biometrika 93 (2006) 655–669.
[14] D. Lynden-Bell, A method of allowing for known observational selection in small samples applied to 3RC quasars, Mon. Not. R. Astron. Soc. 155 (1971) 95–118.
[15] E.C. Martin, R.A. Betensky, Testing quasi-independence of failure and truncation via conditional Kendall's tau, J. Amer. Statist. Assoc. 100 (2005) 484–492.
[16] D. Oakes, Bivariate survival models induced by frailties, J. Amer. Statist. Assoc. 84 (1989) 487–493.
[17] W.-Y. Tsai, Testing the assumption of independence of truncation time and failure time, Biometrika 77 (1990) 169–177.
[18] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 1998.
[19] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Process, Springer, New York, 1996.
[20] D. Zeng, D.Y. Lin, Efficient estimation of semiparametric transformation models for counting processes, Biometrika 93 (2006) 627–640.