Stat Papers
DOI 10.1007/s00362-017-0947-z

REGULAR ARTICLE

# Parametric likelihood inference and goodness-of-fit for dependently left-truncated data, a copula-based approach

**Takeshi Emura**[1] · **Chi-Hung Pan**[1]

Received: 21 April 2017 / Revised: 13 August 2017


**Abstract** Traditionally, the literature on statistical inference with left-truncated samples assumes the independence of truncation variable on lifetime. Alternatively, this paper considers an approach of using a copula for dependent truncation. When considering maximum likelihood estimation and goodness-of-fit procedures, key challenges are the absence of the explicit form of the inclusion probability and truncated distribution functions. This paper shows that, under the copula model, the inclusion probability and truncated distribution functions are expressed as univariate integrals of some functions. With aid of these expressions, we propose computational algorithms to maximize the log-likelihood and to perform goodness-of-fit tests. Simulations are conducted to examine the performance of the proposed method. Real data from a field reliability study on the brake pad lifetimes are analyzed for illustration. Relevant computational programs are made available in the R package "depend.truncation".

**Keywords** Bivariate life distribution · Goodness-of-fit test · Newton–Raphson algorithm · Reliability · Survival analysis · Left-truncation · Weibull distribution

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00362-017-0947-z) contains supplementary material, which is available to authorized users.

---

✉ Takeshi Emura
takeshiemura@gmail.com

Chi-Hung Pan
cool54333@gmail.com

[1] Graduate Institute of Statistics, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan City 32001, Taiwan

Published online: 04 September 2017

## 1 Introduction

This paper considers the situation that a pair of lifetime variables $(L, X)$ can be available in the samples only if $L \leq X$ holds. Nothing is available if $L > X$. The variable $X$ is said to be *left-truncated* by $L$. Owing to this sampling scheme, the available samples tend to exhibit larger values of $X$ than the completely random samples (i.e., length-biased sampling). Such left-truncation phenomena occur in studies of education, economics, engineering, and epidemiology. Statistical methods for left-truncated data are typically found in textbooks of survival analysis, such as Klein and Moeschberger (2003) and Lawless (2003).

Traditionally, the literature on the truncated data analysis considers statistical inference by assuming that $L$ and $X$ are independent. For instance, the Lynden-Bell nonparametric estimator for $S_X(x) = \Pr(X > x)$ is derived under the independence (Lynden-Bell 1971). Kalbfleisch and Lawless (1992) assumed the independence in analysis of their brake pads data, and then developed parametric likelihood inference. Using the same data, Gardes and Stupfler (2014) derived their quantile estimator under the independence. Standard textbooks of survival analysis, such as Klein and Moeschberger (2003) and Lawless (2003), introduce methods for analyzing left-truncated lifetime data under the independence.

The independence assumption may often be questionable and can be tested by several different nonparametric tests (Tsai 1990; Uña-Álvarez 2012; Martin and Betensky 2005; Emura and Wang 2010; Strzalkowska-Kominiak and Stute 2013). Bakoyannis and Touloumi (2012, 2017) raised concerns about the effect posed by dependent truncation under the competing risks setting. Recently, Emura and Wang (2016) reported the bias of the traditional regression analysis caused by ignoring the effect of dependent truncation. These previous works underscore the need to appropriately handle the dependence if it exists.

Statistical inference for dependently truncated data was initiated by Chaieb et al. (2006) in which they suggested a copula model between $L$ and $X$. A parametric inference for dependently truncated data under the bivariate normal distribution was subsequently considered by Emura and Konno (2012a). Similarly, Emura and Konno (2012b) considered parametric inference under the bivariate Poisson model, along with the goodness-of-fit tests. Chaieb et al. (2006) proposed a semi-parametric estimation procedure for their copula model in which the marginal models are unspecified. Under the same copula model, Emura and Wang (2012) and Emura and Murotani (2015) proposed different semiparametric estimation schemes. Ding (2012) verified the identifiability of the model used in Chaieb et al. (2006). Another proposal is a copula-based nonparametric association study of Strzalkowska-Kominiak and Stute (2013).

We follow the aforementioned approaches to model the dependence between the truncation time $L$ and lifetime $X$ via copulas. In particular, we develop the maximum likelihood estimator (MLE) under a copula model between $L$ and $X$, where the marginal distributions of $L$ and $X$ are parametrically specified. Our work is regarded as a more flexible version of the results of Emura and Konno (2012a, b) who considered specific cases: bivariate normal and bivariate Poisson models. From a different view, our work is similar to the copula-based parametric inference of Escarela and

Carriere (2003) that was developed for dependent censoring (not for dependent truncation). However, the problem is more challenging in dependent truncation, where the inference involves the complicated inclusion probability $c(\boldsymbol{\theta}) = \Pr(L \leq X)$.

In stress-strength models, for any bivariate random variables $(X, Y)$, the probability $R = \Pr(Y < X)$ is called "reliability". In the assessment of $R$, most existing literature imposes the independence assumption between $X$ and $Y$ (Greco and Ventura 2011; Cortese and Ventura 2013, and references therein). The calculation of $R$ under a copula model is only recently considered by Domma and Giordano (2013). They focused on the Farlie–Gumbel–Morgenstern copula and Frank copula, but did not discuss about statistical inference. Here our paper works on general copulas, and also discusses likelihood inference under dependent truncation.

The paper is organized as follows. Section 2 describes the motivating example. Section 3 proposes a copula-based likelihood inference method and goodness-of-fit procedure under dependent truncation. Section 4 presents an example under the Clayton copula. Section 5 presents simulation studies. Section 6 considers a real data application. Section 7 discuss the issue of double-truncation. Section 8 concludes the paper and mentions future work. Supplementary Material includes detailed mathematical formulas and additional numerical results.

## 2 Motivating example

Figure 1 displays an example of left-truncated data that appear in a field reliability study on the lifetimes of brake pads of automobiles (Kalbfleisch and Lawless 1992; Lawless 2003). The lifetime of the brake pads was defined as the number of kilometers driven before the pads fail (wear that requires replacement of the pads). The manufacturer selected random samples of cars which were sold over the preceding 12 months. The samples become available when cars have the brake pads still working at the time of sampling. Thus, the sample inclusion criterion is $L \leq X$, where $L$ is the number

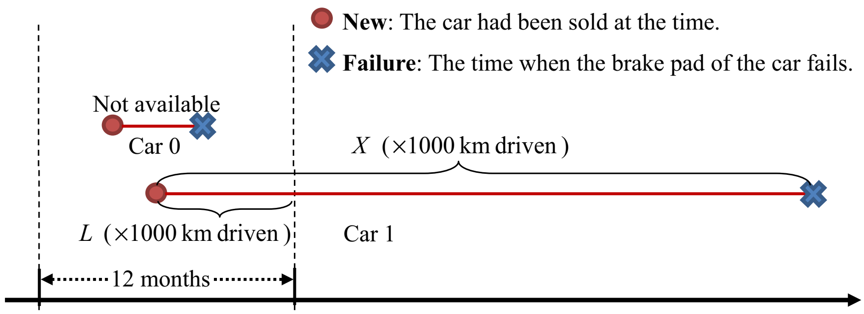


**Fig. 1** Left-truncated lifetimes on the brake pads of automobiles (Kalbfleisch and Lawless 1992; Lawless 2003), where $X$ is the number of kilometers of the car driven at failure, and $L$ is the number of kilometers of the car driven at the sampling point. *Car 1* is sold before the sampling starts, and it fails after the sampling time. Hence, *Car 1* is available in the samples. *Car 0* is not available in the samples as the brake pad is broken too early (before the sampling time)

of kilometers driven for brake pads at the time of sampling, and $X$ is the number of kilometers driven for brake pads at the time of failure.

The aforementioned brake pads data are written as $(L_j, X_j)$, $L_j \le X_j$, $j = 1, 2, \ldots, n$ with $n = 98$. Let $f_{L,X}(l, x) = f_L(l) f_X(x)$ be the joint density of $(L, X)$ under the independence assumption. Following Wang (1991), the *full likelihood* is expressed as

$$
\begin{aligned}
L_n &= \prod_{j=1}^{n} \frac{f_{L,X}(L_j, X_j)}{\iint_{l \le x} f_{L,X}(l, x) dx dl} \\
&= \prod_{j=1}^{n} \frac{f_{L,X}(L_j, X_j)}{\int_{L_j \le x} f_{L,X}(L_j, x) dx} \times \frac{\int_{L_j \le x} f_{L,X}(L_j, x) dx}{\iint_{l \le x} f_{L,X}(l, x) dx dl} \\
&= \prod_{j=1}^{n} \frac{f_X(X_j)}{\int_{L_j \le x} f_X(x) dx} \times \prod_{j=1}^{n} \frac{\int_{L_j \le x} f_{L,X}(L_j, x) dx}{\iint_{l \le x} f_{L,X}(l, x) dx dl} \\
&\equiv L_{1,n} \times L_{2,n}.
\end{aligned}
$$

Traditional statistical methods for left-truncated data are constructed by maximizing $L_{1,n}$ while ignoring $L_{2,n}$. For instance, the nonparametric estimator of Lynden-Bell (1971) is regarded as the maximizer of $L_{1,n}$. The above likelihood decomposition avoids the calculation of a complicated double integral in $L_{2,n}$.

Kalbfleisch and Lawless (1992) and Lawless (2003) analyzed the brake bad data by maximizing $L_{1,n}$ under the lognormal distribution on $f_X$. It is easy to see that the likelihood is

$$
L_{1,n}(\mu, \sigma) = \prod_{j=1}^{n} \frac{1}{\sqrt{X_j}\sigma} \frac{\phi[\{\log(X_j) - \mu\}/\sigma]}{1 - \Phi[\{\log(L_j) - \mu\}/\sigma]}.
$$

All the first and second derivatives of $\log L_{1,n}(\mu, \sigma)$ with respect to $(\mu, \sigma)$ are analytically obtained, and hence, the Newton–Raphson (NR) algorithm can be employed to maximize $L_{1,n}(\mu, \sigma)$. Our numerical analysis results in $(\hat{\mu}, \hat{\sigma}) = (4.109, 0.421)$ which are exactly the same values as those reported in Lawless (2003).

Clearly, the likelihood decomposition is meaningless if the independence assumption $f_{L,X}(l, x) = f_L(l) f_X(x)$ is violated. The problem of dependent truncation necessarily involves the full-likelihood. Here, a key computational challenge is to obtain the expressions of the double integral $c(\boldsymbol{\theta}) = \iint_{l \le x} f_{L,X}(l, x) dx dl$, where $\boldsymbol{\theta}$ is a vector of parameters for the joint density $f_{L,X}$. Furthermore, *the truncated distribution function*,

$$
F_{L \le X}(l, x; \boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}(L \le l, X \le x, L \le X) = \iint_{s \le l,\, s \le t \le x} f_{L,X}(s, t; \boldsymbol{\theta}) ds dt,
$$

is useful in goodness-of-fit assessment, which again requires a double integral. This becomes a serious burden when computing these integrals during the iterations of the NR algorithm and hundreds of the bootstrap replications.

This paper proposes novel expressions of $c(\boldsymbol{\theta})$ and $F_{L\le X}(l,\ x;\ \boldsymbol{\theta})$ in terms of univariate integrals of some functions on the unit interval (Theorems 1 and 2). Beside the analysis of dependently left-truncated data, obtaining such expressions would facilitate the analysis of reliability $R = \Pr(Y < X)$ (Sect. 1) and the analysis of doubly-truncated data (Sect. 7). With aid of the new expressions, we develop computational procedures in maximum likelihood estimation, including the NR algorithm, the formulas of standard errors (SEs), and the formal goodness-of-fit tests.

## 3 Proposed method

### 3.1 Copula models for dependent truncation

A bivariate copula is a bivariate distribution function for a pair of two uniform random variables on [0, 1] (Nelsen 2006). Let $(U_1,\ U_2)$ be a pair of uniform random variables on $[0,\ 1]^2$ following a copula $C_\alpha\colon[0,\ 1]^2 \mapsto [0,\ 1]$, where $\alpha \in R$ is a dependence parameter. For notational simplicity, we define the conditional distribution function for $U_1|U_2 = u_2$ as

$$h_\alpha\left(u_1,\ u_2\right) \equiv \Pr\left(U_1 \le u_1 | U_2 = u_2\right) = C_\alpha^{[0,1]}\left(u_1,\ u_2\right) = \frac{\partial C_\alpha(u_1,\ u_2)}{\partial u_2}.$$

The function $h_\alpha\colon[0,\ 1]^2 \to [0,\ 1]$ is called "$h$-function" (Schepsmeier and Stöber 2014). The density function of $(U_1,\ U_2)$ is

$$C_\alpha^{[1,1]}\left(u_1,\ u_2\right) = \frac{\partial^2 C_\alpha(u_1,\ u_2)}{\partial u_1 \partial u_2}.$$

Let $F_L(l;\ \boldsymbol{\theta}_L)$ and $F_X(x;\ \boldsymbol{\theta}_X)$ be the marginal distribution functions of $L$ and $X$, respectively. Here $\boldsymbol{\theta}_L$ is a $k$-variate vector of parameters and $\boldsymbol{\theta}_X$ is a $m$-variate vector of parameters. We assume that the inverse functions of $F_L(l;\ \boldsymbol{\theta}_L)$ and $F_X(x;\ \boldsymbol{\theta}_X)$ exist. Then the probability integral transformations $U_1 = F_L(L;\ \boldsymbol{\theta}_L)$ and $U_2 = F_X(X;\ \boldsymbol{\theta}_X)$ yield a joint distribution function

$$\Pr_{\boldsymbol{\theta}}(L \le l,\ X \le x) = C_\alpha\left[F_L\left(l;\ \boldsymbol{\theta}_L\right),\ F_X\left(x;\ \boldsymbol{\theta}_X\right)\right],$$

where $\alpha$ represents the degree of dependence between $L$ and $X$. Let $\boldsymbol{\theta} = (\alpha,\ \boldsymbol{\theta}_L, \boldsymbol{\theta}_X) \in \Theta$ be a $(k + m + 1)$-variate vector of parameters and $\Theta \subset R^{k+m+1}$ be a parameter space. Then the density function of $(L,\ X)$ is

$$f_{L,X}(l,\ x;\ \boldsymbol{\theta}) = C_\alpha^{[1,1]}\left[F_L\left(l;\ \boldsymbol{\theta}_L\right),\ F_X\left(x;\ \boldsymbol{\theta}_X\right)\right] f_L\left(l;\ \boldsymbol{\theta}_L\right) f_X\left(x;\ \boldsymbol{\theta}_X\right),$$

where $f_L(l;\ \boldsymbol{\theta}_L) = dF_L(l;\ \boldsymbol{\theta}_L)/dl$ and $f_X(x;\ \boldsymbol{\theta}_X) = dF_X(x;\ \boldsymbol{\theta}_X)/dx$. The above parametric model is different from the semiparametric model of Chaieb et al. (2006) in which the marginal densities are unspecified.

### 3.2 Likelihood construction

Given the observed data $\{(L_j,\ X_j);\ j=1,\ 2,\ldots,n\}$ subject to $L_j \le X_j$, the likelihood function has the form

$$L_n(\boldsymbol{\theta}) = c(\boldsymbol{\theta})^{-n} \sum_{j=1}^{n} f_{L,X}\left(L_j,\ X_j;\ \boldsymbol{\theta}\right),$$

where

$$c(\boldsymbol{\theta}) = \Pr(L \le X) = \iint_{l \le x} f_{L,X}(l,\ x;\ \boldsymbol{\theta})dxdl.$$

As mentioned in Sect. 2, the form of $c(\boldsymbol{\theta})$ is necessary to perform likelihood inference. Indeed, Emura and Konno (2012a, b) developed likelihood inference under the bivariate normal and bivariate Poisson models where the simple forms of $c(\boldsymbol{\theta})$ are available. The Monte Carlo method to approximate the numerical value of $c(\boldsymbol{\theta})$ is not a suitable option since the likelihood inference requires repeated evaluation of $c(\boldsymbol{\theta})$ at different values of $\boldsymbol{\theta}$.

To obtain the form of $c(\boldsymbol{\theta})$, we define *H-function*

$$H(u;\ \boldsymbol{\theta}) = h_\alpha\left[F_L\left\{F_X^{-1}\left(u;\ \boldsymbol{\theta}_X\right);\ \boldsymbol{\theta}_L\right\},\ u\right].$$

**Theorem 1** *The inclusion probability is written as the univariate integral*

$$c(\boldsymbol{\theta}) = \Pr(L \le X) = \int_0^1 H(u;\ \boldsymbol{\theta})du.$$

*Proof* By straight forward calculations, we have

$$\begin{aligned} c(\boldsymbol{\theta}) &= \Pr(L \le X) = \Pr\left\{F_L\left(L;\ \boldsymbol{\theta}_L\right) \le F_L\left(X;\ \boldsymbol{\theta}_L\right)\right\} \\ &= \Pr\left[U_1 \le F_L\left\{F_X^{-1}\left(U_2;\ \boldsymbol{\theta}_X\right);\ \boldsymbol{\theta}_L\right\}\right] \\ &= \mathrm{E}\left(\Pr\left[U_1 \le F_L\left\{F_X^{-1}\left(U_2;\ \boldsymbol{\theta}_X\right);\ \boldsymbol{\theta}_L\right\} | U_2\right]\right) \\ &= \int_0^1 h_\alpha\left[F_L\left\{F_X^{-1}\left(u;\ \boldsymbol{\theta}_X\right);\ \boldsymbol{\theta}_L\right\} u\right]du. \end{aligned}$$

□

Theorem 1 is quite general, which works on almost all types of bivariate continuous lifetime models. The univariate integral in Theorem 1 can be accurately approximated by numerical integration routines in software packages, much more accurately than the Monte Carlo method.

By Theorem 1, the log-likelihood is written as

$$\ell_n(\boldsymbol{\theta}) = -n \log\left\{\int_0^1 H(u;\, \boldsymbol{\theta})du\right\} + \sum_{j=1}^{n} \log f_L\left(L_j;\, \boldsymbol{\theta}_L\right) + \sum_{j=1}^{n} \log f_X\left(X_j;\, \boldsymbol{\theta}_X\right) + \sum_{j=1}^{n} \log C_\alpha^{[1,1]}\left[F_L\left(L_j;\, \boldsymbol{\theta}_L\right),\, F_X\left(X_j;\, \boldsymbol{\theta}_X\right)\right].$$

Then, the MLE is defined as

$$\hat{\boldsymbol{\theta}} = \left(\hat{\alpha},\, \hat{\boldsymbol{\theta}}_L,\, \hat{\boldsymbol{\theta}}_X\right) = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max}\, \ell_n(\boldsymbol{\theta}).$$

### 3.3 Score vector and Hessian matrix

In order to perform likelihood inference, one needs the first- and second-order derivatives of the log-likelihood $\ell_n(\boldsymbol{\theta})$, such as $\partial\ell_n(\boldsymbol{\theta})/\partial\alpha$ and $\partial^2\ell_n(\boldsymbol{\theta})/\partial\alpha^2$. These derivatives comprise the score function and Hessian matrix whose expressions are given in Supplementary Material. They involves the derivatives of copulas

$$C_\alpha^{[i,j]}(u_1,\, u_2) = \frac{\partial^{(i+j)} C_\alpha(u_1,\, u_2)}{\partial u_1^i \partial u_2^j},$$

where $(i,\, j) \in \{0,\, 1,\, 2,\, 3\} \times \{0,\, 1,\, 2,\, 3\}$. We provides all these derivatives of the Clayton copula and the Joe copula (1993) in Supplementary Material. The Clayton copula has the lower tail dependence while the Joe copula has the upper tail dependence.

For R users, one can use the R package: *VineCopula* (Schepsmeier et al. 2015) to calculate the derivatives of copulas. We have checked the correctness of the package by comparing between our derivative formulas and the package outputs (Supplementary Material).

The following lemma facilitates the subsequent calculus.

**Lemma 1** (Khuri 2003, p. 301) *Let* $D = \{(u,\, \theta_1,\, \theta_2, \ldots, \theta_p)\, |0 \le u \le 1,\, a_i \le \theta_i \le b_i,\, i = 1, \ldots, p\}$, *where* $a_i$ *and* $b_i$ *be real numbers with* $a_i < b_i$, $i = 1, 2, \ldots, p$. *Let* $D_i = \{(u,\, \theta_i)\, |0 \le u \le 1,\, a_i \le \theta_i \le b_i\}$ *for fixed* $\theta_j$ *with* $j \ne i$. *If* $H$ *and* $\partial H/\partial\theta_i$ *are continuous in* $D_i$, *then*

$$\frac{\partial}{\partial\theta_i}\int_0^1 H(u;\, \boldsymbol{\theta})du = \int_0^1 \frac{\partial H(u;\, \boldsymbol{\theta})}{\partial\theta_i}du, \quad i = 1, \ldots, p.$$

If $H(u;\ \boldsymbol{\theta})$ in Theorem 1 satisfies the condition of Lemma 1, all the first- and second-order derivatives of $c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\alpha,\ \boldsymbol{\theta}_L,\ \boldsymbol{\theta}_X)$ are again univariate integrals of some functions on the interval [0, 1]. In particular,

$$\begin{bmatrix} c_\alpha(\boldsymbol{\theta}) \\ c_{\boldsymbol{\theta}_L}(\boldsymbol{\theta}) \\ c_{\boldsymbol{\theta}_X}(\boldsymbol{\theta}) \end{bmatrix} \equiv \begin{bmatrix} \partial c(\boldsymbol{\theta})/\partial\alpha \\ \partial c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_L \\ \partial c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_X \end{bmatrix} = \int_0^1 \begin{bmatrix} \partial H(u;\ \boldsymbol{\theta})/\partial\alpha \\ \partial H(u;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}_L \\ \partial H(u;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}_X \end{bmatrix} du,$$

$$\begin{bmatrix} c_{\alpha\alpha}(\boldsymbol{\theta}) & c^{\mathrm{T}}_{\alpha\boldsymbol{\theta}_L}(\boldsymbol{\theta}) & c^{\mathrm{T}}_{\alpha\boldsymbol{\theta}_X}(\boldsymbol{\theta}) \\ c_{\alpha\boldsymbol{\theta}_L}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_L\boldsymbol{\theta}_L}(\boldsymbol{\theta}) & c^{\mathrm{T}}_{\boldsymbol{\theta}_L\boldsymbol{\theta}_X}(\boldsymbol{\theta}) \\ c_{\alpha\boldsymbol{\theta}_X}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_L\boldsymbol{\theta}_X}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_X\boldsymbol{\theta}_X}(\boldsymbol{\theta}) \end{bmatrix}$$

$$\equiv \int_0^1 \begin{bmatrix} \partial^2 H(u;\ \boldsymbol{\theta})/\partial\alpha^2 & \partial^2 H(u;\ \boldsymbol{\theta})/\partial\alpha\partial\boldsymbol{\theta}_L^{\mathrm{T}} & \partial^2 H(u;\ \boldsymbol{\theta})/\partial\alpha\partial\boldsymbol{\theta}_X^{\mathrm{T}} \\ \partial^2 H(u;\ \boldsymbol{\theta})/\partial\alpha\partial\boldsymbol{\theta}_L & \partial^2 H(u;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}_L\partial\boldsymbol{\theta}_L^{\mathrm{T}} & \partial^2 H(u;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}_L\partial\boldsymbol{\theta}_X^{\mathrm{T}} \\ \partial^2 H(u;\ \boldsymbol{\theta})/\partial\alpha\partial\boldsymbol{\theta}_X & \partial^2 H(u;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}_X\partial\boldsymbol{\theta}_L^{\mathrm{T}} & \partial^2 H(u;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}_X\partial\boldsymbol{\theta}_X^{\mathrm{T}} \end{bmatrix} du.$$

The derivative expressions of $H(u;\ \boldsymbol{\theta})$ are available in closed forms. With the above expression, the first and second derivatives of the log-likelihood are obtained, which are given in Supplementary Material.

### 3.4 Standard error and confidence interval

Under the regularity conditions stated in Emura and Konno (2012b), the asymptotic theory holds for the MLE. The SE is obtained from the negative of the inverse Hessian matrix of the log-likelihood (the observed Fisher information matrix), and the confidence interval (CI) is formed by the normal approximation.

The SE of $\hat{\theta}_j$ is

$$SE\left(\hat{\theta}_j\right) = \sqrt{\left[\left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\ell_n(\hat{\boldsymbol{\theta}})\right\}^{-1}\right]_{jj}},$$

where $[A]_{jj}$ is the $j$th diagonal element of a matrix $A$, and $\theta_j$ is the $j$th component of $\boldsymbol{\theta}$.

The $(1-\beta)\times 100\%$ CI for $\theta_j$ utilizes the $p$th upper quantile for $N(0,\ 1)$, denoted as $Z_p$, If the range of parameter $\theta_j$ is unrestricted, one can use the linear CI

$$\hat{\theta}_j \pm Z_{\beta/2} \times SE\left(\hat{\theta}_j\right).$$

For a positive parameter $\theta_j > 0$, one may use the log-transformed CI

$$\hat{\theta}_j \exp\left\{\pm Z_{\beta/2} \times SE\left(\hat{\theta}_j\right)/\hat{\theta}_j\right\}.$$

Let $g(\hat{\boldsymbol{\theta}})$ be an estimate of $g(\boldsymbol{\theta})$, where $g$ is a differentiable function. By the delta method, we obtain the SE of $g(\hat{\boldsymbol{\theta}})$:

$$SE\{g(\hat{\boldsymbol{\theta}})\} = \sqrt{\left\{\frac{\partial}{\partial \boldsymbol{\theta}} g(\hat{\boldsymbol{\theta}})\right\}^{\mathrm{T}} \times \left\{-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \ell_n(\hat{\boldsymbol{\theta}})\right\}^{-1} \times \frac{\partial}{\partial \boldsymbol{\theta}} g(\hat{\boldsymbol{\theta}})}.$$

A quantity of interest is the mean lifetime $g(\boldsymbol{\theta}) = E(X) = \int x f_X(x;\ \boldsymbol{\theta}_X)dx$. Since the mean lifetime is positive, one may use the log-transformed CI.

*Remark 1* Another quantity of interest is the CI for the marginal distribution function $g(\boldsymbol{\theta}) = F_X(x;\ \boldsymbol{\theta}_X)$. For this case, the quantile-based transformation may be recommended for constricting the CI, especially for small samples (Hong et al. 2008).

### 3.5 Goodness-of-fit test

Goodness-of-fit tests provide a formal way to test the validity of the parametric form of a given model. Following Emura and Konno (2012a, b), we consider a goodness-of-fit test based on the distance between the empirical distribution function $\hat{F}_{L,X}(l,\ x) = \sum_j \mathbf{I}(L_j \le l,\ X_j \le x)/n$ and its parametric counterpart $F_{L\le X}(l,\ x;\ \hat{\boldsymbol{\theta}})/c(\hat{\boldsymbol{\theta}})$, where $\mathbf{I}(\cdot)$ is the indicator function and

$$F_{L\le X}(l,\ x;\ \boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}(L \le l,\ X \le x,\ L \le X) = \iint_{s\le l,\, s\le t\le x} f_{L,X}(s,\ t;\ \boldsymbol{\theta})dsdt,$$

is the truncated distribution function. Emura and Konno (2012a, b) suggested the Kolmogorov–Smirnov type statistics

$$K = \sup_{x,\ y} \left| \hat{F}_{L,X}(l,\ x) - F_{L\le X}(l,\ x;\ \hat{\boldsymbol{\theta}})/c(\hat{\boldsymbol{\theta}}) \right|,$$

and the Cramér–von Mises type statistic

$$\begin{aligned} C &= \iint_{l\le x} n \left\{ \hat{F}_{L,X}(l,\ x) - F_{L\le X}(l,\ x;\ \hat{\boldsymbol{\theta}})/c(\hat{\boldsymbol{\theta}}) \right\}^2 d\hat{F}(l,\ x) \\ &= \sum_j \left\{ \hat{F}_{L,X}\left(L_j,\ X_j\right) - F_{L\le X}\left(L_j,\ X_j;\ \hat{\boldsymbol{\theta}}\right)/c(\hat{\boldsymbol{\theta}}) \right\}^2. \end{aligned}$$

Note that the truncated distribution function $F_{L\le X}(l,\ x;\ \boldsymbol{\theta})$ is not equal to the untruncated distribution function $\Pr_{\boldsymbol{\theta}}(L \le l,\ X \le x) = C_\alpha[F_L(l;\ \boldsymbol{\theta}_L),\ F_X(x;\ \boldsymbol{\theta}_X)]$; the former involves the double integral while the latter does not. The form of $F_{L\le X}(l,\ x;\ \boldsymbol{\theta})$ was previously obtained only under specific distributional forms of the bivariate normal and bivariate Poisson model (Emura and Konno 2012a, b). We will generalize the previous results.

Similar to the idea of Theorem 1, we derive the following theorem to reduce the computational cost of the double-integral:

**Theorem 2** *The truncated distribution function is expressed as the univariate integrals*

$$F_{L\leq X}(l,\ x;\ \boldsymbol{\theta}) = \Pr(L \leq l,\ X \leq x,\ L \leq X) = \int_0^{F_X(l)} H(v;\ \boldsymbol{\theta})dv + \int_{F_X(l)}^{F_X(x)} h_\alpha\left\{F_L(l),\ v\right\}dv.$$

*Proof* By straight forward calculations, we have

$$\begin{aligned}\Pr(L \leq l,\ X \leq x,\ L \leq X) &= \int_0^l \Pr(L \leq t,\ X = t) + \int_l^x \Pr(L \leq l,\ X = t)\\ &= \int_0^l \Pr\left(U \leq F_L(t)|V = F_X(t)\right) f_X(t)dt\\ &\quad + \int_l^x \Pr\left(U \leq F_L(l)|V = F_X(t)\right) f_X(t)dt\\ &= \int_0^l h_\alpha\left\{F_L(t),\ F_X(t)\right\} f_X(t)dt\\ &\quad + \int_l^x h_\alpha\left\{F_L(l),\ F_X(t)\right\} f_X(t)dt.\end{aligned}$$

The desired result is obtained by transformations to the last two integrals. □

Obviously, Theorem 2 with $l = x \to \infty$ agrees with Theorem 1.

For a given dataset, one can numerically evaluate the statistics $K$ and $C$ with aid of Theorem 1. In the R depend.truncation package (Emura 2017), we have implemented routines for computing $K$ and $C$ under three models; the Clayton copula with Weibull margins, the Clayton copula with exponential margins, and the Gaussian copula with lognormal margins. The P-value of the goodness-of-fit tests is obtained by a parametric bootstrap (Emura and Konno 2012a). However, since each bootstrap replications involve re-computations for the MLE, $K$, and $C$, a huge number of integrations must be evaluated. This is a real burden without aids of Theorems 1 and 2.

In conjunction with the test results, a graphical diagnostic procedure is useful by plotting $F_{L\leq X}(L_j,\ X_j;\ \hat{\boldsymbol{\theta}})/c(\hat{\boldsymbol{\theta}})$ against $\hat{F}_{L,X}(L_j,\ X_j)$. If the plot bend away from the diagonal line, this indicates evidence that the fitted model is not a good choice.

### 3.6 Software

All aspects of numerical computations developed in this paper are made available in our R package, depend.truncation (Emura 2017). The three functions (PMLE.Clayton.Weibull, PMLE.Clayton.Exponential, and PMLE.Normal) allow users to calculate the MLE and estimates of the mean lifetimes with their SEs and CIs. They also calculate goodness-of-fit statistics and output model-diagnostic plots. For instance, the model with the Clayton copula and Weibull lifetimes is fitted by a function PMLE.Clayton.Weibull.

## 4 An example under the Clayton copula

We demonstrate the method developed in a general copula model (Sect. 3) by using specific models. Consider the Clayton copula

$$C_\alpha(u_1, u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-1/\alpha}, \quad \alpha > 0.$$

Also, consider the Weibull lifetime models, defined as $F_L(l;\, \lambda_L,\, \nu_L) = 1 - \exp(-\lambda_L l^{\nu_L})$ and $F_X(x;\, \lambda_X,\, \nu_X) = 1 - \exp(-\lambda_X x^{\nu_X})$, where $\lambda_L > 0$ and $\lambda_X > 0$ are scale parameters and $\nu_L > 0$ and $\nu_X > 0$ are shape parameters. Let the parameter space for $\boldsymbol{\theta} = (\alpha,\, \lambda_L,\, \lambda_X,\, \nu_L,\, \nu_X)$ be $\Theta = [\varepsilon,\, M]^5$, where $\varepsilon > 0$ is a small number and $M > 0$ is a large number.

### 4.1 Computation of $c(\boldsymbol{\theta})$

By Theorem 1, the inclusion probability is

$$c(\boldsymbol{\theta}) = \Pr(L \le X) = \int_0^1 H(u;\, \boldsymbol{\theta})du, \tag{1}$$

where

$$H(u;\, \boldsymbol{\theta}) = u^{-\alpha-1} B(u,\, \boldsymbol{\theta})^{-1/\alpha-1},$$
$$B(u;\, \boldsymbol{\theta}) = \left(1 - \exp\left[-\lambda_L \left\{-\lambda_X^{-1} \log(1-u)\right\}^{\nu_L/\nu_X}\right]\right)^{-\alpha} + u^{-\alpha} - 1.$$

On the other hand, $c(\boldsymbol{\theta})$ can be computated by the double-integral of the joint density such that

$$c(\boldsymbol{\theta}) = \lambda_L \lambda_X \nu_L \nu_X (1+\alpha) \times \iint_{l \le x} \frac{l^{\nu_L-1} x^{\nu_X-1} \exp(-\lambda_L l^{\nu_L}) \exp(-\lambda_X x^{\nu_X}) [\{1-\exp(-\lambda_L l^{\nu_L})\}\{1-\exp(-\lambda_X x^{\nu_X})\}]^{-\alpha-1}}{(\{1-\exp(-\lambda_L l^{\nu_L})\}^{-\alpha} + \{1-\exp(-\lambda_X x^{\nu_X})\}^{-\alpha} - 1)^{1/\alpha+2}} dl dx. \tag{2}$$

We provide the R codes for computing Eqs. (1) and (2) below:

```
-----------------------------------------------------------------------------------------------------
lambda_L = 2; lambda_X = 1; nu_L=1; nu_X=1; alpha = 1


H_func = function(u){
  E = exp(-lambda_L*(-1/lambda_X*log(1-u))^(nu_L/nu_X))
  RR = (E == rep(1,length(E)))
  E[RR] = 1-10^-8
  B=(1-E)^-alpha+u^-alpha-1
  u^(-alpha-1)*B^(-1/alpha-1)
}
integrate(H_func, lower = 0, upper = 1)


f_func=function(x){
  inner_func=function(l){
  u1=1-exp(-lambda_L*l^nu_L)
  u2=1-exp(-lambda_X*x^nu_X)
  fL=lambda_L*nu_L*l^(nu_L-1)*exp(-lambda_L*l^nu_L)
  fX=lambda_X*nu_X*x^(nu_X-1)*exp(-lambda_X*x^nu_X)
  A=u1^(-alpha)+u2^(-alpha)-1
  fL*fX*(1+alpha)*(u1*u2)^(-alpha-1)/A^(1/alpha+2)
  }
  integrate(inner_func,0,x)$value
}
integrate(Vectorize(f_func),0,Inf)
-----------------------------------------------------------------------------------------------------------
```

By implementing the codes, one can see that Eqs. (1) and (2) yield exactly the same value 0.7363999.

Table 1 evaluates the computation time required to calculate $c(\boldsymbol{\theta})$ under selected values of $\boldsymbol{\theta} = (\alpha,\ \lambda_L,\ \lambda_X,\ \nu_L,\ \nu_X)$. In all parameter values, the computing time based on Eq. (1) is much shorter than that based on Eq. (2). The reduction of computation time is more remarkable when the value of $c(\boldsymbol{\theta})$ is small.

**Table 1** Computation times of computing $c(\boldsymbol{\theta})$ based on the two methods

| Parameters | $c(\boldsymbol{\theta})$ | $B$ | The number of computations: $N = B \times \text{AI}$ | Computing time (sec.): Eq. (1) | Computing time (sec.): Eq. (2) |
|---|---|---|---|---|---|
| $\alpha = 1,\ \lambda_L = 2, \lambda_X = 1, \nu_L = 1, \nu_X = 1$ | 0.736 | 100 | 660 | 0.02 | 1.98 |
| | | 200 | 1280 | 0.05 | 3.80 |
| | | 300 | 1920 | 0.08 | 5.64 |
| $\alpha = 1,\ \lambda_L = 1, \lambda_X = 1, \nu_L = 1, \nu_X = 1$ | 0.500 | 100 | 900 | 0.02 | 2.62 |
| | | 200 | 1560 | 0.05 | 4.48 |
| | | 300 | 2340 | 0.08 | 6.72 |
| $\alpha = 1,\ \lambda_L = 1, \lambda_X = 1, \nu_L = 2, \nu_X = 1$ | 0.416 | 100 | 3270 | 0.09 | 10.01 |
| | | 200 | 6100 | 0.18 | 18.89 |
| | | 300 | 8130 | 0.23 | 25.04 |

Thu number of computations is defined as $N = B \times \text{AI}$, where $B$ is the number of bootstraps, and AI is the average number of iterations in the Newton–Raphson algorithm obtained from Supplementary Material

## 4.2 Computation of the MLE

To obtain the score function, we need the first-order derivatives of $H(u;\ \boldsymbol{\theta})$ as:

$$\begin{aligned}
\frac{\partial H(u;\ \boldsymbol{\theta})}{\partial \alpha} &= u^{-\alpha-1} B(u,\ \boldsymbol{\theta})^{-1/\alpha-1} \\
&\quad \times \left[ -\log(u) + \frac{\log\{B(u;\ \boldsymbol{\theta})\}}{\alpha^2} + \left( -\frac{1}{\alpha} - 1 \right) \frac{B_\alpha(u;\ \boldsymbol{\theta})}{B(u;\ \boldsymbol{\theta})} \right], \\
\partial H(u;\ \boldsymbol{\theta})/\partial \lambda_L &= (-1/\alpha - 1) u^{-\alpha-1} B(u;\ \boldsymbol{\theta})^{-1/\alpha-2} B_{\lambda_L}(u;\ \boldsymbol{\theta}), \\
\partial H(u;\ \boldsymbol{\theta})/\partial \lambda_X &= (-1/\alpha - 1) u^{-\alpha-1} B(u;\ \boldsymbol{\theta})^{-1/\alpha-2} B_{\lambda_X}(u;\ \boldsymbol{\theta}), \\
\partial H(u;\ \boldsymbol{\theta})/\partial \nu_L &= (-1/\alpha - 1) u^{-\alpha-1} B(u;\ \boldsymbol{\theta})^{-1/\alpha-2} B_{\nu_L}(u;\ \boldsymbol{\theta}), \\
\partial H(u;\ \boldsymbol{\theta})/\partial \nu_X &= (-1/\alpha - 1) u^{-\alpha-1} B(u;\ \boldsymbol{\theta})^{-1/\alpha-2} B_{\nu_X}(u;\ \boldsymbol{\theta}),
\end{aligned}$$

where by defining $Q_X(u) \equiv \{-\lambda_X^{-1} \log(1-u)\}^{1/\nu_X}$,

$$\begin{aligned}
B_\alpha(u;\ \boldsymbol{\theta}) &\equiv -\frac{\log(1 - \exp[-\lambda_L Q_X(u)^{\nu_L}])}{(1 - \exp[-\lambda_L Q_X(u)^{\nu_L}])^\alpha} + \frac{-\log(u)}{u^\alpha}, \\
B_{\lambda_L}(u;\ \boldsymbol{\theta}) &\equiv -\frac{\alpha Q_X(u)^{\nu_L} \exp[-\lambda_L Q_X(u)^{\nu_L}]}{(1 - \exp[-\lambda_L Q_X(u)^{\nu_L}])^{\alpha+1}}, \\
B_{\lambda_X}(u;\ \boldsymbol{\theta}) &\equiv -\frac{\alpha \lambda_L \nu_L \exp[-\lambda_L Q_X(u)^{\nu_L}]}{\lambda_X^2 \nu_X (1 - \exp[-\lambda_L Q_X(u)^{\nu_L}])^{\alpha+1}} \\
&\quad \times \left\{ -\lambda_X^{-1} \log(1-u) \right\}^{\nu_L/\nu_X - 1} \log(1-u),
\end{aligned}$$

$$B_{\nu_L}(u;\ \boldsymbol{\theta}) \equiv -\frac{\alpha\lambda_L Q_X(u)^{\nu_L}\exp[-\lambda_L Q_X(u)^{\nu_L}]}{\nu_X(1-\exp[-\lambda_L Q_X(u)^{\nu_L}])^{\alpha+1}} \times \log\left\{-\lambda_X^{-1}\log(1-u)\right\},$$
$$B_{\nu_X}(u;\ \boldsymbol{\theta}) \equiv \frac{\alpha\lambda_L \nu_L Q_X(u)^{\nu_L}\exp[-\lambda_L Q_X(u)^{\nu_L}]}{\nu_X^2(1-\exp[-\lambda_L Q_X(u)^{\nu_L}])^{\alpha+1}} \times \log\left\{-\lambda_X^{-1}\log(1-u)\right\}.$$

As $H$ and $\partial H/\partial\boldsymbol{\theta}$ are continuous with respect to $(u,\ \boldsymbol{\theta})$ in $[0,\ 1]\times\Theta$, by Lemma 1,

$$\begin{bmatrix} c_\alpha(\boldsymbol{\theta}) \\ c_{\lambda_L}(\boldsymbol{\theta}) \\ c_{\lambda_X}(\boldsymbol{\theta}) \\ c_{\nu_L}(\boldsymbol{\theta}) \\ c_{\nu_X}(\boldsymbol{\theta}) \end{bmatrix} \equiv \begin{bmatrix} \partial c(\boldsymbol{\theta})/\partial\alpha \\ \partial c(\boldsymbol{\theta})/\partial\lambda_L \\ \partial c(\boldsymbol{\theta})/\partial\lambda_X \\ \partial c(\boldsymbol{\theta})/\partial\nu_L \\ \partial c(\boldsymbol{\theta})/\partial\nu_X \end{bmatrix} = \int_0^1 \begin{bmatrix} \partial H(u;\ \boldsymbol{\theta})/\partial\alpha \\ \partial H(u;\ \boldsymbol{\theta})/\partial\lambda_L \\ \partial H(u;\ \boldsymbol{\theta})/\partial\lambda_X \\ \partial H(u;\ \boldsymbol{\theta})/\partial\nu_L \\ \partial H(u;\ \boldsymbol{\theta})/\partial\nu_X \end{bmatrix} du.$$

The score function and Hessian matrix are available in Supplementary Material.

With the availability of the score function and Hessian matrix, we apply the NR algorithm to obtain the MLE. Due to the large number of parameters, the NR algorithm is sensitive to the initial values (Knight 2000). Thus, we apply the following the randomized NR algorithm as previously developed (Hu and Emura 2015).

***Algorithm 1: randomized NR algorithm for the Weibull model***

**Step 1** Set initial values $\alpha^{(0)} = 2\hat{\tau}/(1-\hat{\tau})$, $\lambda_L^{(0)} = 1/\bar{L}$, $\lambda_X^{(0)} = 1/\bar{X}$, $\nu_L^{(0)} = 1$ and $\nu_X^{(0)} = 1$, where $\hat{\tau}$ is the sample Kendall tau, $\bar{L} = \sum_{i=1}^n L_i/n$ and $\bar{X} = \sum_{i=1}^n X_i/n$.

**Step 2** Repeat the following iteration for $k = 0,\ 1,\ 2,\ldots$:

$$\begin{bmatrix} \alpha^{(k+1)} \\ \lambda_L^{(k+1)} \\ \lambda_X^{(k+1)} \\ \nu_L^{(k+1)} \\ \nu_X^{(k+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(k)} \\ \lambda_L^{(k)} \\ \lambda_X^{(k)} \\ \nu_L^{(k)} \\ \nu_X^{(k)} \end{bmatrix} - \left[\begin{bmatrix} \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\alpha^2} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_X\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_X\partial\alpha} \\ \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L^2} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L\partial\lambda_X} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L\partial\nu_L} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L\partial\nu_X} \\ \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_X\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_L\partial\lambda_X} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\lambda_X^2} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\lambda_X} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_X\partial\lambda_X} \\ \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\lambda_L} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\lambda_X} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L^2} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\nu_X} \\ \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_X\partial\alpha} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_X\partial\lambda_L} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_X\partial\lambda_X} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_L\partial\nu_X} & \frac{\partial^2\ell_n(\boldsymbol{\theta})}{\partial\nu_X^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\alpha} \\ \frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\lambda_L} \\ \frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\lambda_X} \\ \frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\nu_L} \\ \frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\nu_X} \end{bmatrix}\right]_{\substack{\alpha=\alpha^{(k)} \\ \lambda_L=\lambda_L^{(k)},\ \lambda_X=\lambda_X^{(k)} \\ \nu_L=\nu_L^{(k)},\ \nu_X=\nu_X^{(k)}}}.$$

- If $\max\{|\alpha^{(k+1)}-\alpha^{(k)}|,\ |\lambda_L^{(k+1)}-\lambda_L^{(k)}|,\ |\lambda_X^{(k+1)}-\lambda_X^{(k)}|,\ |\nu_L^{(k+1)}-\nu_L^{(k)}|,\ |\nu_X^{(k+1)}-\nu_X^{(k)}|\} < \varepsilon$ and the Hessian matrix is negative definite, then the values $(\alpha^{(k+1)},\ \lambda_L^{(k+1)},\ \lambda_X^{(k+1)},\ \nu_L^{(k+1)},\ \nu_X^{(k+1)})$ are the MLE of $(\alpha,\ \lambda_L,\ \lambda_X,\ \nu_L,\ \nu_X)$.
- If $\max\{|\alpha^{(k+1)}-\alpha^{(k)}|,\ |\lambda_L^{(k+1)}-\lambda_L^{(k)}|,\ |\lambda_X^{(k+1)}-\lambda_X^{(k)}|,\ |\nu_L^{(k+1)}-\nu_L^{(k)}|,\ |\nu_X^{(k+1)}-\nu_X^{(k)}|\} > Err$, $\alpha^{(k+1)} > \alpha_{\max}$, $\alpha^{(k+1)} < \alpha_{\min}$, $\min\{\lambda_L^{(k+1)},\ \lambda_X^{(k+1)},\ \nu_L^{(k+1)},\ \nu_X^{(k+1)}\} < 10^{-8}$ or $k = \{100,\ 200,\ 300,\ldots\}$ holds, then replace the initial values $(\alpha^{(0)},\ \lambda_L^{(0)},\ \lambda_X^{(0)},\ \nu_L^{(0)},\ \nu_X^{(0)})$ with$\{\alpha^{(0)}\times\exp(u_1),\ \lambda_L^{(0)}\times\exp(u_2),\ \lambda_X^{(0)}\times\exp(u_3),\ \nu_L^{(0)}\times\exp(u_4),\ \nu_X^{(0)}\times\exp(u_5)\}$, where $u_i \sim U(-r_i,\ r_i)$ for a radius $r_i > 0,\ i = 1,\ldots,5$. Then, restart Step 2.

*Remark 2* We used $Err = 2$, $\alpha_{\min} = 10^{-4}$, $\alpha_{\max} = 20$, $r_1 = 1$ and $r_2 = \cdots = r_5 = 0.5$ for numerical analysis. We randomize the initial value when $k = \{100, 200, \ldots\}$ to escape infinite roops occurring during the iterations.

The exponential distribution emerges as the case of $\nu_L = \nu_X = 1$, which however requires a separate development (the details are given in Supplementary Material).

Copula models allow one to consider a "differing marginal model", such as the Clayton copula with the Weibull model for $X$ and the lognormal model for $L$, and vice versa. However, such models are more difficult to interpret as the two models may have different meanings of parameters. Indeed, many medical and engineering applications of the copula-based bivariate survival models apply the same martinal forms (Escarela and Carriere 2003; Hsu et al. 2016; Emura and Michimae 2017).

## 5 Simulation

Monte Carlo simulations were performed to examine the performance of the proposed method. Data were generated from the exponential and Weibull lifetime models under the Clayton copula (detailed in Supplementary Material). Then, we checked the following factors: (1) convergence of the NR algorithm, (2) closeness of the estimate to the true value, (3) closeness of the SE to the sample standard deviation, (4) correctness of the coverage probability of the CI.

Our simulation results are given in Supplementary Material, which demonstrate that the NR algorithm converges quickly and robustly. The simulations also show that the estimators are nearly unbiased and the CIs provide desirable coverage performance. Overall, the proposed method had sound performance on the factors examined.

## 6 Data analysis

The proposed method is illustrated using the field reliability data of the brake pad lifetimes (Kalbfleisch and Lawless 1992; Lawless 2003). The data is introduced in Sect. 2.

### 6.1 Data preprocessing

We first extracted the data from Table 1 of Kalbfleisch and Lawless (1992). They defined

- $L =$ the number of kilometers driven for brake pads at sampling point,
- $X =$ the number of kilometers driven for brake pads at failure.

The lifetime $X$ is left-truncated by $L$ as explained in Sect. 2.

We decided to remove one outlying data point ($L = 6.951$, $X = 53.926$) since the presence of this point made all the candidate parametric models fit poorly (i.e., all models rejected by our goodness-of-fit tests). This data point has the smallest left-truncation point, which is far away from other left-truncation points (extreme data point).

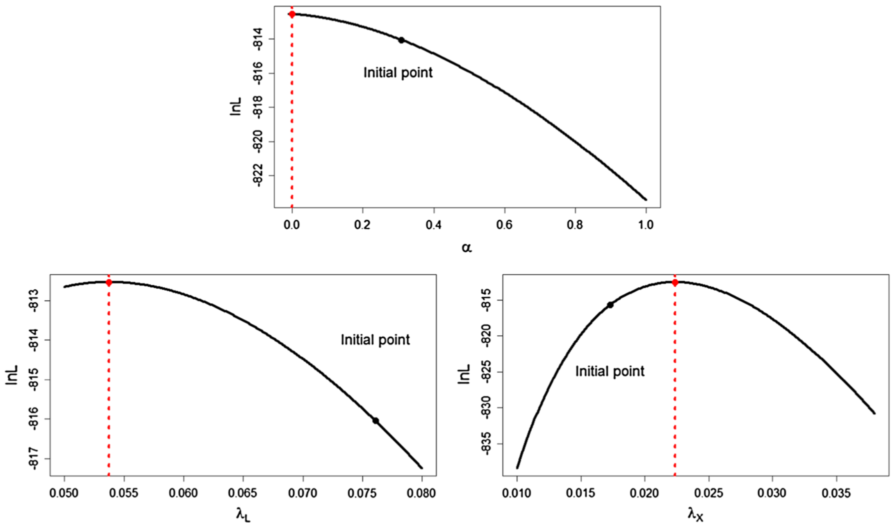


**Fig. 2** Profile plots for $\ell_n(\alpha, \lambda_L, \lambda_X)$ under the Clayton copula with exponential margins based on the automobile brake pads data. The *vertical lines* signify the MLEs $\hat{\alpha} = 0.000$, $\hat{\lambda}_L = 0.0537$, and $\hat{\lambda}_X = 0.0224$

In the presence of left-truncation, the identifiable region of the model for $X$ is $[u_{\min}, \infty)$, where $l_{\min} = \inf\{l;\ F_L(l) > 0\}$ is the lower support for the left-truncation variable (Woodroofe 1985). Thus, what we can estimate is the conditional distribution $X|X \geq l_{\min}$. The value $l_{\min}$ is unknown, often replaced by the smallest left-truncation point (Kalbfleisch and Lawless 1992; Lawless 2003). Thus, we set $l_{\min} = 10$. In the case of dependent truncation models, what we can actually estimate is the conditional distribution $(L, X)|L \geq l_{\min}, X \geq l_{\min}$. This corresponds to the model fitted after subtracting $l_{\min}$ from both $L$ and $X$.

Kalbfleisch and Lawless (1992) performed a parametric analysis under the independence between $L$ and $X$. They chose the log-normal distribution for $X$. Here in our paper, we analyze the data by fitting copula models allowing for the dependence between $L$ and $X$.

### 6.2 Numerical result

Under the exponential lifetime model with the Clayton copula, the NR algorithm (Algorithm A in Supplementary Material) ascertained the MLEs $\hat{\alpha} = 0.000$, $\hat{\lambda}_L = 0.0537$, and $\hat{\lambda}_X = 0.0224$. Figure 2 shows that the MLE attained the maximum of the log-likelihood function. The goodness-of-fit tests yield significant evidence against the model; $K = 0.189$ (P-value <0.000) and $C = 1.066$ (P-value <0.000).

Under the Weibull lifetime model with the Clayton copula, the randomized NR algorithm (Algorithm 1) ascertained the MLEs $\hat{\alpha} = 0.242$, $\hat{\lambda}_L = 0.0192$, $\hat{\lambda}_X = 0.000126$, $\hat{\nu}_L = 1.457$ and $\hat{\nu}_X = 2.163$. Figure 3 displays that the MLEs attained the maximum of the log-likelihood function. The goodness-of-fit tests did not yield significant evidence against the model; $K = 0.079$ (P-value$=$0.200) and $C = 0.095$ (P-value$=$0.113).

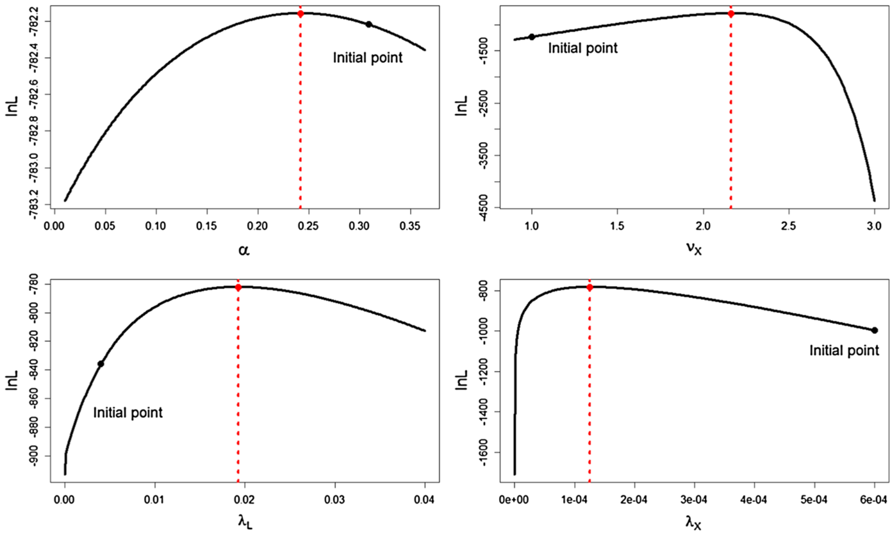


**Fig. 3** Profile plots for $\ell_n(\alpha,\ \lambda_L,\ \lambda_X,\ \nu_L,\ \nu_X)$ under the Clayton copula with Weibull margins based on the automobile brake pads data. The *vertical lines* signify the MLEs $\hat{\alpha} = 0.242$, $\hat{\lambda}_L = 0.0192$, $\hat{\lambda}_X = 0.000126$, $\hat{\nu}_L = 1.457$ (not shown) and $\hat{\nu}_X = 2.163$

We also fitted the lognormal lifetime model with the normal copula which can be implemented by fitting log-transformed lifetimes to the bivariate normal model (Emura and Konno 2012a). The estimates were obtained by *PMLE.Normal* routine in R *depend.truncation* package and were given in Table 2. The Cramér–von Mises goodness-of-fit test yielded some evidence against the model ($C = 0.099$, P-value = 0.094), but the Kolmogorov–Smirnov goodness-of-fit test did not ($K = 0.066$, P-value = 0.494).

Table 2 compares the results of the three fitted models. Both the Weibull lifetime model and the lognormal lifetime model showed weak positive dependence between $L$ and $X$. For the Weibull model, $\hat{\alpha} = 0.242$ and 95% CI = [0.049, 1.193], and for the lognormal model, $\hat{\rho} = 0.209$ and 95% CI = [−0.011, 0.429]. The exponential lifetime model yield $\hat{\alpha} = 0$ at the boundary of parameter space and hence 95% CI is not available.

From the goodness-of-fit results (Fig. 4), the Weibull lifetime model and the lognormal lifetime model were very competitive. The lognormal lifetime model reached a significance level (P-value $<0.10$) by the Cramér–von Mises test while the Weibull model did not (P-value = 0.113). In general, the Cramér–von Mises test measures a more global fit than the Kolmogorov–Smirnov test. The goodness-of-fit test yields strong evidence against the exponential model. Note that information theoretic criteria, such as AIC and BIC, may not be straightforwardly applied for model comparison in the presence of truncation. Taking these into consideration, we finally chose the Weibull model for the subsequent analysis. In addition, the Weibull model may be more suitable for the brake pad example, where the study of aging properties is of great importance for industrial manufacturers.

**Table 2** The result of analysing the brake pad lifetime data (Kalbfleisch and Lawless 1992)

| | Weibull lifetime; Clayton copula | Exponential lifetime; Clayton copula | Lognormal lifetime; normal copula |
|---|---|---|---|
| Copula parameter | $\hat{\alpha} = 0.242$ (SE=0.197) | $\hat{\alpha} = 0.000$ (SE=0.006) | $\hat{\rho} = 0.209$ (SE=0.112) |
| | 95% CI [0.049, 1.193] | 95% CI not available | 95% CI [−0.011, 0.429] |
| Parameters for $L$ | $\hat{\lambda}_L = 0.019$ (SE=0.007) | $\hat{\lambda}_L = 0.054$ (SE=0.008) | $\hat{\mu}_L = 2.38$ (SE=0.097) |
| | $\hat{\nu}_L = 1.457$ (SE=0.116) | $\nu_L = 1$ (fixed) | $\sigma_L^2 = 0.722$ (SE= 0.124) |
| Parameters for $X$ | $\hat{\lambda}_X = 0.00013$ (SE=0.0001) | $\hat{\lambda}_X = 0.022$ (SE=0.002) | $\hat{\mu}_X = 3.92$ (SE=0.054) |
| | $\hat{\nu}_X = 2.163$ (SE=0.186) | $\nu_X = 1$ (fixed) | $\sigma_X^2 = 0.266$ (SE=0.039) |
| KS test | $K = 0.079$ (P=0.200) | $K = 0.189$ (P=0.000) | $K = 0.066$ (P= 0.494) |
| CvM test | $C = 0.095$ (P=0.113) | $C = 1.066$ (P=0.000) | $C = 0.099$ (P= 0.094) |

*SE* standard error, *KS test* Kolmogorov–Smirnov test for goodness-of-fit, *CvM test* Cramér–von Mises test for goodness-of-fit, *P* P-value based on the parametric bootstrap, *95% CI* 95% confidence interval (it is not available if $\hat{\alpha} = 0$)

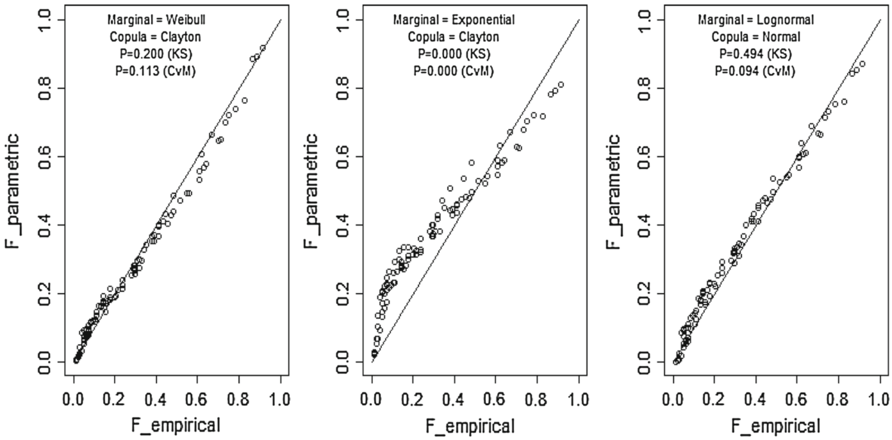


**Fig. 4** Goodness-of-fit tests for the brake pad lifetime data (Kalbfleisch and Lawless 1992). *KS* Kolmogorov–Smirnov test, *CvM* Cramér–von Mises test, *P* P-value based on the parametric bootstrap

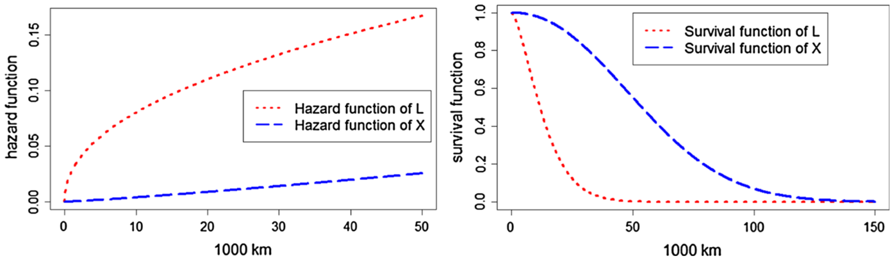


**Fig. 5** Hazard functions and survival functions under the Weibull model

Under the Weibull model, we assess the mean lifetime for the brake pads based on the estimate $\hat{E}(X) = \Gamma(1 + 1/\hat{\nu}_X)/\hat{\lambda}_X^{1/\hat{\nu}_X} = 56.4$ (SE=2.9). It is actually the mean residual lifetime $X - l_{\min}|X \geq l_{\min}$, where $l_{\min} = 10$. Hence, on average, the brake pad needs to be replaced after the car drives about $66.4 \times 1000$ km. We present the fitted hazard functions and fitted survival functions of $L$ and $X$ in Fig. 5. The hazard for $X$ increases with time, which is the natural aging process of brake pads after long drive. Note that this hazard plot is not possible by the semiparametric approaches in which the survival curve is a step function.

## 7 Extension to double-truncation

Suppose that the sample is available only if $L \leq X \leq R$ holds, where $R$ is right-truncation time. This is the setting of doubly-truncated data, where $X$ is subject to both left- and right- truncations. For much the same reason as left-truncated data, finding the form of $c = \Pr(L \leq X \leq R)$ is crucial to conduct likelihood inference (Emura and Konno 2012a).

Suppose that the distribution of $(L, X, R)$ follows the model,

$$\Pr(L \le l, X \le x, R \le r) = C\left[F_L(l), F_X(x), F_R(r)\right],$$

where $C\colon [0, 1]^3 \mapsto [0, 1]$ is a tri-variate copula. Let $U = F_L(L)$, $V = F_X(X)$ and $W = F_R(R)$ be the probability integral transforms. Then, the conditional distribution function is

$$h(u, v, w) \equiv \Pr(U \le u, W \le w | V = v) = \frac{\partial C_\alpha(u, v, w)}{\partial v}.$$

**Theorem 3** *The inclusion probability is written as the univariate integral*

$$c = \int_0^1 h\left[F_L\{F_X^{-1}(v), v, 1\right] dv - \int_0^1 h\left[F_L\{F_X^{-1}(v), v, F_R\{F_X^{-1}(v)\right] dv.$$

*Proof* By straightforward calculations, we have

$$\begin{aligned} c &= \Pr(L \le X, X \le R) = \Pr\left[U \le F_L\left\{F_X^{-1}(V)\right\}, F_R\left\{F_X^{-1}(V)\right\} \le W\right] \\ &= E\left(\Pr\left[U \le F_L\left\{F_X^{-1}(V)\right\}, F_R\left\{F_X^{-1}(V)\right\} \le W | V\right]\right) \\ &= E\left(\Pr\left[U \le F_L\left\{F_X^{-1}(V)\right\} | V\right]\right) \\ &\quad - E\left(\Pr\left[U \le F_L\left\{F_X^{-1}(V)\right\}, W < F_R\left\{F_X^{-1}(V)\right\} | V\right]\right). \end{aligned}$$

Finally, the last expression is re-written as the integral form that appears in Theorem 3. □

Theorem 3 implies that all the procedures developed for left-truncated data may be extended to doubly-truncated data. However, the details of this topic is beyond the scope of the paper.

## 8 Conclusion

This paper considers a copula model to account for dependence between a pair of lifetime variables $(L, X)$, where $X$ is left-truncated by $L$. To resolve the methodological difficulty of maximum likelihood estimation, we propose a novel expression of the inclusion probability (Theorem 1) and its derivatives. Also, in order to derive a goodness-of-fit procedure, we propose a similar expression on the truncated distribution function (Theorem 2). Due to space limitations, we have only demonstrated these theorems under the Clayton copula with Weibull margins (Sect. 4). However, we emphasize that the new expressions are fairly general, which can be applied to almost all types of continuous bivariate lifetime models. Indeed, we have developed the R package "depend.truncation" that can fit three different models.

Unlike the semiparametric approaches of Chaieb et al. (2006), we adopt a parametric approach that specified the form of the lifetime distribution. For purpose of reliability assessments, parametric models are more easily analyzed and interpreted by engineers than semiparametric models (see Sect. 6). A practical advantage of the parametric approach is that one can easily deduce the mean lifetime and the hazard function. For instance, our real data analysis (Sect. 6) demonstrated that the shape of the hazard function expresses the natural aging process of brake pads after long drive, giving some practical advice for engineers and consumers. In addition, our data analysis gives the informative conclusion that the brake pad needs to be replaced after a long use. Note that these conclusions are not straightforwardly deduced from the semiparametric approaches.

The left-truncated data considered in Hong et al. (2009) provide a unique challenge for applying the proposed approach. In their data, the left-truncation time is defined as the elapsed time from the installation date of a machine to the study initiation date. For those installed after the initiation date, the truncation time may be set as 0 (Hong et al. 2009), or simply undefined (see Appendix II of Emura and Shiu 2016). In the former case, the left-truncation distribution has a probability mass concentrated on 0. Since the proposed method in this paper is developed for continuous left-truncation times, further consideration is needed for handling this type of left-truncation.

One may consider the extension of our parametric models in the presence of covariates. In the context of dependent censoring models, regression analyses have been developed well (e.g., Escarela and Carriere 2003; Braekers and Veraverbeke 2005; Chen 2010; Emura et al. 2015; Staplin et al. 2015; Emura and Chen 2016; Emura et al. 2017b). However, to the best of our knowledge, only Ding (2012) shortly discussed the covariate models under a copula-based dependent truncation model.

Another direction of future research may be a reliability aspect of the lifetime model under dependent truncation. Our real data example has demonstrated the successful application of the Weibull model to express the aging process of the brake pad after long drive. While our presentation of the aging process employs the marginal hazard plots and marginal survival curves (Fig. 5), there are several other alternatives. A recent work by Noughabi and Kayid (2017) suggests a bivariate quantile residual plots. While such a plot is more difficult to interpret than the marginal plots, it contains rich information. For instance, as explained by Noughabi and Kayid (2017), the quantile residual plots may be used as an individualized prediction measure of the residual life (Emura et al. 2017a). Prediction of remaining life in a highly reliable products is an important issue in the presence of left-truncation (Hong et al. 2009), but is less studied in the literature.

**Acknowledgements** The authors thank two anonymous reviewers for their helpful comments that improved the paper. This work is supported by the Research Grants funded by the Government of Taiwan (MOST 103-2118-M-008-MY2; MOST 105-2118-M-008-003-MY2).

## References

Bakoyannis G, Touloumi G (2012) Practical methods for competing risks data: a review. Stat Method Med Res 21:257–272

Bakoyannis G, Touloumi G (2017) Impact of dependent left truncation in semiparametric competing risks methods: a simulation study. Commun Stat Simul Comput 46(3):2025–2042
Braekers R, Veraverbeke N (2005) A copula-graphic estimator for the conditional survival function under dependent censoring. Can J Stat 33:429–447
Chaieb LL, Rivest LP, Abdous B (2006) Estimating survival under a dependent truncation. Biometrika 93:665–669
Chen YH (2010) Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. J R Stat Soc B 72:235–251
Cortese G, Ventura L (2013) Accurate higher-order likelihood inference on P(Y < X). Comput Stat 28(3):1035–1059
De Uña-Álvarez J (2012) On the Markov three-state progressive model. In: Recent advances in system reliability. Springer, New York
Ding AA (2012) Copula identifiability conditions for dependent truncated data model. Lifetime Data Anal 18(4):397–407
Domma F, Giordano S (2013) A copula-based approach to account for dependence in stress-strength models. Stat Pap 54(3):807–826
Emura T (2017) R depend.truncation: statistical inference for parametric and semiparametric models based on dependently truncated data, CRAN: version 2.8. https://CRAN.R-project.org/package=depend.truncation
Emura T, Chen YH (2016) Gene selection for survival data under dependent censoring: a copula-based approach. Stat Method Med Res 25(6):2840–2857
Emura T, Konno Y (2012a) Multivariate normal distribution approaches for dependently truncated data. Stat Pap 53:133–149
Emura T, Konno Y (2012b) A goodness-of-fit tests for parametric models based on dependently truncated data. Comput Stat Data Anal 56:2237–2250
Emura T, Michimae H (2017) A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae. Environ Ecol Stat 24(1):151–173
Emura T, Murotani K (2015) An algorithm for estimating survival under a copula-based dependent truncation model. TEST 24(4):734–751
Emura T, Shiu SK (2016) Estimation and model selection for left-truncated and right-censored lifetime data with application to electric power transformers analysis. Commun Stat Simul Comput 45(9):3171–3189
Emura T, Wang W (2010) Testing quasi-independence for truncation data. J Multivar Anal 101:223–239
Emura T, Wang W (2012) Nonparametric maximum likelihood estimation for dependent truncation data based on copulas. J Multivar Anal 110:171–188
Emura T, Nakatochi M, Murotani K, Rondeau V (2015) A joint frailty-copula model between tumour progression and death for meta-analysis. Stat Methods Med Res. doi:10.1177/0962280215604510
Emura T, Nakatochi M, Matsui S, Michimae H, Rondeau V (2017a) Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: meta-analysis with a joint model. Stat Methods Med Res. doi:10.1177/0962280216688032
Emura T, Chen YH, Matsui S, Rondeau V (2017b) Survival analysis with dependent censoring and correlated endpoints—copula-based approaches. JSS research series in statistics. Springer, Berlin
Emura T, Wang W (2016) Semiparametric inference for an accelerated failure time model with dependent truncation. Ann Inst Stat Math 68(5):1073–1094
Escarela G, Carriere JF (2003) Fitting competing risks with an assumed copula. Stat Methods Med Res 12(4):333–349
Gardes L, Stupfler G (2014) Estimating extreme quantiles under random truncation. TEST 24(2):207–227
Greco L, Ventura L (2011) Robust inference for the stress-strength reliability. Stat Pap 52(4):773–788
Hong Y, Meeker WQ, Escobar LA (2008) Avoiding problems with normal approximation confidence intervals for probabilities. Technometrics 50(1):64–68
Hong Y, Meeker WQ, McCalley JD (2009) Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. Ann Appl Stat. doi:10.1214/00-AOAS231
Hsu TM, Emura T, Fan TH (2016) Reliability inference for a copula-based series system life test under multiple type-I censoring. IEEE Trans Reliab 65(2):1069–1080
Hu YH, Emura T (2015) Maximum likelihood estimation for a special exponential family under random double-truncation. Comput Stat 30(4):1199–1229

Joe H (1993) Parametric families of multivariate distributions with given margins. J Multivar Anal 46:262–282

Kalbfleisch JD, Lawless JF (1992) Some useful statistical methods for truncated data. J Qual Technol 24:145–152

Khuri AI (2003) Advanced calculus with applications in statistics (vol 486). Wiley. doi:10.1002/0471394882

Klein JP, Moeschberger ML (2003) Survival analysis techniques for censored and truncated data, 2nd edn. Springer, New York

Knight K (2000) Mathematical statistics. Chapman and Hall, Boca Raton

Lawless JF (2003) Statistical models and methods for lifetime data, 2nd edn. Wiley, Hoboken

Lynden-Bell D (1971) A method of allowing for known observational selection in small samples applied to 3RC quasars. Mon Not R Astron Soc 155:95–118

Martin EC, Betensky RA (2005) Testing quasi-independence of failure and truncation via conditional Kendall's tau. J Am Stat Assoc 100:484–492

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York

Noughabi MS, Kayid M (2017) Bivariate quantile residual life: a characterization theorem and statistical properties. Stat Pap. doi:10.1007/s00362-017-0905-9

Schepsmeier U, Stöber J (2014) Derivatives and Fisher information of bivariate copulas. Stat Pap 55:525–542

Schepsmeier U, Stöber J, Brechmann et al (2015) R VineCopula: statistical inference of vine copulas, version 1.5, CRAN

Staplin ND, Kimber AC, Collett D, Roderick PJ (2015) Dependent censoring in piecewise exponential survival models. Stat Methods Med Res 24(3):325–341

Strzalkowska-Kominiak E, Stute W (2013) Empirical copulas for consecutive survival data. TEST 22:688–714

Tsai WY (1990) Testing the assumption of independence of truncation time and failure time. Biometrika 77:169–177

Wang MC (1991) Nonparametric estimation from cross-sectional survival data. J Am Stat Assoc 86(413):130–143

Woodroofe M (1985) Estimating a distribution function with truncated data. Ann Stat 13:163–177