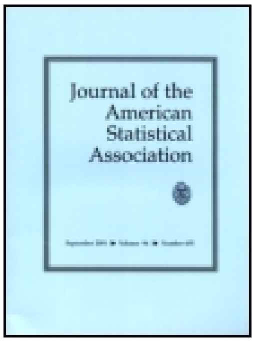

Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Copula Based Cox Proportional Hazards Models for Dependent Censoring

Negera Wakgari Deresa & Ingrid Van Keilegom

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2022.2161387

Copula Based Cox Proportional Hazards Models for Dependent Censoring

Negera Wakgari Deresa and Ingrid Van Keilegom

ORSTAT, KU Leuven, Leuven, Belgium

ABSTRACT
Most existing copula models for dependent censoring in the literature assume that the parameter defining the copula is known. However, prior knowledge on this dependence parameter is often unavailable. In this article we propose a novel model under which the copula parameter does not need to be known. The model is based on a parametric copula model for the relation between the survival time (T) and the censoring time (C), whereas the marginal distributions of T and C follow a semiparametric Cox proportional hazards model and a parametric model, respectively. We show that this model is identified, and propose estimators of the nonparametric cumulative hazard and the finite-dimensional parameters. It is shown that the estimators of the model parameters and the cumulative hazard function are consistent and asymptotically normal. We also investigate the performance of the proposed method using finite-sample simulations. Finally, we apply our model and estimation procedure to a follicular cell lymphoma dataset. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2022
Accepted December 2022

KEYWORDS
Association; Nonparametric function; Semiparametric regression; Survival analysis

1. Introduction

When studying the effect of some covariates X on a survival time T, a popular model in the literature is a semiparametric proportional hazards model (or Cox model), which is defined by the following conditional survival function:

$$S_{T|X}(t|x) = \exp\{-\Lambda(t)e^{x^\top \beta}\}, \tag{1.1}$$

where β is a vector of regression parameters and Λ belongs to a space $\mathcal{G}$ of cumulative hazard functions; we refer to the seminal paper by Cox (1972) for details. In the presence of random right censoring, the latter paper proposed an elegant estimation procedure provided that the survival time T and the (right) censoring time C are stochastically independent given covariates. This assumption holds in many situations, mainly when censoring occurs at the end of the study (administrative censoring). However, there are also numerous contexts where the independent censoring assumption does not hold. For example, in medical studies, patients may withdraw from treatment when their prognosis is poor, and hence the reason for withdrawal may be linked to the expected event of interest, which induces dependent censoring; see Deresa and Van Keilegom (2021) for more examples. In the case where the independence assumption is doubtful, the event time analysis based on the Cox model may lead to serious estimation bias (Huang and Zhang 2008).

A convenient way to deal with dependent censoring is to use a copula model for the joint distribution of T and C. However, since the copula between T and C is not identifiable in a fully nonparametric setting (Tsiatis 1975), some authors work with a known copula model, which completely specifies the association between the two variables. The first paper in that line was Zheng and Klein (1995), who proposed a consistent nonparametric estimator for the marginal distributions of T and C. This estimator generalizes the Kaplan and Meier (1958) estimator to the case of dependent censoring. Rivest and Wells (2001) further studied the proposal of Zheng and Klein (1995) for the particular case of an Archimedean copula. For the case with covariates, Huang and Zhang (2008) modeled the marginal distributions of both T and C using the Cox model under a known copula and demonstrated the nice performance of their model through simulations. Later on, Chen (2010) extended this copula approach to semiparametric transformation models, which include the Cox model as a special case.

A known copula approach has the advantage of not making strict model assumptions about the marginal distributions, but it has the disadvantage of assuming that the dependence structure is known. Supposing the copula to be known is often an unrealistic assumption, but as shown in Crowder (1991), identification is impossible if the copula is entirely unknown. It is therefore essential to check identifiability within certain parametric copulas, where the copula structure is specified up to a finite-dimensional parameter.

Other approaches not based on copulas have also been proposed to handle dependent censoring. In this regard, the inverse probability of censoring weighted (IPCW) method (see, Robins and Finkelstein 2000; Scharfstein and Robins 2002; Collett 2015, among others), where the weight is obtained from a censoring time model using the auxiliary variables as covariates, and the multiple imputation method of Jackson et al. (2014), where the censored times are imputed under user-specified deviations from independent censoring, are also helpful to adjust for dependent censoring in the Cox model. These methods either

CONTACT Negera Wakgari Deresa negerawakgari.deresa@kuleuven.be ORSTAT, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

assume the availability of auxiliary variables to reduce bias due to dependent censoring or they impose nonidentifiable parameters concerning the dependence between survival and censoring times for which it is required to conduct a sensitivity analysis.

In contrast, our approach does not require a sensitivity analysis on the dependence parameter because we will be able to identify and estimate this parameter along with all other parameters from the observed survival data. In addition, we do not need to assume the existence of auxiliary variables in order to deal with dependent censoring.

In this article we propose a semiparametric copula approach for survival data subject to dependent censoring. The proposed method is based on the Cox model for the marginal distribution of T, as in model (1.1), whereas the marginal distribution of C can be defined within any parametric model provided that certain regularity conditions are satisfied. Then, their dependence structure is modeled by a parametric copula. We will show that this model is identifiable based on the distribution of the observed survival data. In particular, we will identify the association between T and C even though we observe either one of them, but never both of them. The identifiability is shown only based on the distribution of the observed data without imposing further restrictive assumptions on the dependence parameter. Our approach is more flexible than a parametric copula approach recently proposed in Czado and Van Keilegom (2023) and Deresa et al. (2022) in the sense that we do not need to impose a parametric restriction on the marginal distribution of T. We also propose a method to estimate this nonparametric function using martingale ideas when T and C are dependent given a set of covariates. Moreover, under certain regularity conditions, we prove the consistency and asymptotic normality of the parameter estimators and the estimator of the nonparametric function. The performance of the proposed method is also assessed through finite-sample simulations. Interestingly, our method is not sensitive to the type of parametric copula family we assume between survival and dependent censoring times.

The remainder of the article is organized as follows. In the next section we state the model and derive the distribution of the observed quantities. The identifiability of our model is studied in Section 3. Section 4 describes how to estimate the model parameters and the cumulative hazard function. The asymptotic properties of the parameter and nonparametric estimators are established in Section 5, whereas their proofs are deferred to the supplementary material. Section 6 presents a simulation study to illustrate the finite-sample performance of the proposed method. An application to a real data example is given in Section 7, and Section 8 contains some conclusions and discussion.

2. Model Specification

In this article we will assume that in addition to the survival time T and the dependent censoring time C, there is an independent censoring time A. The variable A can for instance represent the time until the end of the study, which is usually independent of T. Due to the presence of right censoring, we observe $Z = \min(T, C, A)$ and the censoring indicators (Δ_1, Δ_2) given by $\Delta_1 = \mathbf{I}(Z = T)$ and $\Delta_2 = \mathbf{I}(Z = C)$, where $\mathbf{I}(\cdot)$ is the indicator function. Let X (of dimension p) be the covariates related to T, and let W (of dimension q) be the covariates that influence C. The vectors X and W can have some or all of the variables in common, but they can also be completely distinct. We allow for dependence between T and C, even after conditioning out the effect of covariates, and will model this dependence using a copula model. The variable A is assumed to be independent of (T, C), given (X, W).

The marginal model for T is given by the following Cox model:

$$F_{T|X}(t|x) = \Pr(T \le t|X = x) = 1 - \exp\{-\Lambda(t)e^{x^\top \beta}\}, \quad (2.1)$$

where β is a vector of regression coefficients and Λ is an unknown increasing and differentiable cumulative baseline hazard function of unspecified form with $\Lambda(0) = 0$. We assume that T is absolutely continuous. The corresponding baseline hazard function is denoted by λ, which is defined as $\lambda(t) = d\Lambda(t)/dt$. Moreover, we assume that $X^\top\beta$ does not contain an intercept for identification reasons. Under the independent censoring assumption, Cox (1972) proposed a procedure to estimate the regression parameter β based on the concept of partial likelihood, in which the baseline hazard function is canceled out from the likelihood, and he obtained likelihood-type asymptotic theory for the parameter estimator. Our inference procedure will not follow the partial likelihood approach because we allow for the case where T and C are dependent given the covariates. We therefore rely on a pseudo-likelihood approach as detailed in Section 4.

We also need a model for the marginal distribution of C. It is assumed that the margin of C follows a parametric model:

$$F_{C|W} \in \{F_{C|W,\eta} : \eta \in H\}, \quad (2.2)$$

for some parameter space H, where $F_{C|W}(c|w) = \Pr(C \le c|W = w)$. The corresponding density is assumed to exist and is denoted by $f_{C|W}(c|w)$.

We now propose a joint model through a bivariate copula in order to take the association between T and C into account for given X and W. A bivariate copula is a distribution function $\mathcal{C}$ defined over the unit square $[0, 1]^2$ with uniform margins; see Nelsen (2006) for details on copulas. Hence, the conditional distribution function of T and C can be modeled, for all $t, c > 0$, by

$$\Pr(T \le t, C \le c|X = x, W = w) = \mathcal{C}\{F_{T|X}(t|x), F_{C|W}(c|w)\}, \quad (2.3)$$

for some copula $\mathcal{C}$, whereas the distribution of A is left entirely unspecified. The conditional distributions are assumed to be continuous, so that the copula is unique (Sklar 1959). Throughout the article we assume that the copula $\mathcal{C}$ is parametric, that is,

$$\mathcal{C} \in \{\mathcal{C}_\gamma : \gamma \in \Gamma\}$$

for some parameter space Γ.

For what follows we require the following notations for $0 \le u, v \le 1$:

$$h_{T|C}(u|v) = \frac{\partial}{\partial v}\mathcal{C}(u, v), \quad \text{and} \quad h_{C|T}(v|u) = \frac{\partial}{\partial u}\mathcal{C}(u, v).$$

It can be shown that

$$\Pr(T \le t | C = c, X = x, W = w) = h_{T|C}(F_{T|X}(t|x)|F_{C|W}(c|w))$$
$$\Pr(C \le c | T = t, X = x, W = w) = h_{C|T}(F_{C|W}(c|w)|F_{T|X}(t|x)).$$

The proof is given in Deresa et al. (2022). We shall assume that

(C1) (T, C) and A are independent given (X, W).
(C2) A and (X, W) are independent. In addition, we assume that the matrices $\text{var}(X)$ and $\text{var}(W)$ have full rank, and the vectors X and W contain at least one continuous variable.
(C3) The probabilities $\Pr(Z = T|X, W)$ and $\Pr(Z = C|X, W)$ are strictly positive for almost all (X, W).
(C4) The censoring by A is noninformative for (T, C), that is, the distribution of A does not depend on any of the model parameters.

Based on the distribution of (Z, Δ_1, Δ_2) given (X, W), we will study identifiability of models (2.1)–(2.3) in the next section. Let us now derive the formulas needed in this identifiability proof and for the estimation of our model. Let $G(z, \delta_1, \delta_2|x, w)$ denote the sub-distribution function of the triplet (Z, Δ_1, Δ_2) given X and W, that is,

$$G(z, \delta_1, \delta_2|x, w) = \Pr(Z \le z, \Delta_1 = \delta_1, \Delta_2 = \delta_2|X = x, W = w) \tag{2.4}$$

and let $g(z, \delta_1, \delta_2|x, w)$ be the corresponding sub-density. Then, we can show that

$$\begin{aligned} g(z,1,0|x,w) &= \lambda(z)e^{x^\top\beta}\exp\{-\Lambda(z)e^{x^\top\beta}\} \\ &\quad \times \left\{1-h_{C|T}(F_{C|W}(z|w)|F_{T|X}(z|x))\right\}\{1-F_A(z)\} \\ g(z,0,1|x,w) &= f_{C|W}(z|w)\left\{1-h_{T|C}(F_{T|X}(z|x)|F_{C|W}(z|w))\right\} \\ &\quad \times \{1-F_A(z)\} \\ g(z,0,0|x,w) &= S(z|x,w)f_A(z), \end{aligned} \tag{2.5}$$

where

$$S(z|x,w) = \tilde{\mathcal{C}}\big(F_{T|X}(z|x), F_{C|W}(z|w)\big) \quad \text{with}$$
$$\tilde{\mathcal{C}}(u,v) = 1 - u - v + \mathcal{C}(u,v),$$

for $0 \le u, v \le 1$, $F_A(z) = \Pr(A \le z)$ and $f_A(z) = (d/dz)F_A(z)$. We can add parameters to these functions to emphasize their dependency on the finite- and/or infinite-dimensional parameters, for instance $S_\alpha(z|x, w)$ with $\alpha = (\gamma, \beta, \Lambda, \eta) \in \Gamma \times \mathbb{R}^p \times \mathcal{G} \times H$. Finally,

$$\begin{aligned} G(z|x,w) :&= \sum\nolimits_{\delta_1+\delta_2\le 1} G(z,\delta_1,\delta_2|x,w) \\ &= 1 - S(z|x,w)\{1-F_A(z)\}. \end{aligned} \tag{2.6}$$

The proofs of (2.5) and (2.6) are provided in the supplementary material.

3. Model Identification

In this section we study in detail the identifiability of our model based on the distribution of the observed vector $(Z, \Delta_1, \Delta_2, X, W)$. Under general regularity conditions on the marginal distributions and on the copula, we will show that the model parameters (γ, β, η) and the function Λ are identifiable, in the sense that any two different sets of parameters give different joint distributions of $(Z, \Delta_1, \Delta_2, X, W)$. Studying identifiability under dependent censoring is far from trivial, especially for the dependence parameter γ. This is because the right censoring mechanism hinders us from observing the pair (T, C) simultaneously, and hence the relation between T and C cannot be directly seen from the observed data. Compared to the identifiability proof under a fully parametric model in Czado and Van Keilegom (2023) and Deresa et al. (2022), the nonparametric component in the marginal model for T poses an additional challenge to show identifiability.

We need to make some assumptions on the joint model in order to show identifiability. The first assumption is about the marginal density of C (given the covariates):

(C5) For all $\eta_1, \eta_2 \in H$, we have:

$$\lim_{t\to 0} \frac{f_{C|W,\eta_1}(t|w)}{f_{C|W,\eta_2}(t|w)} = 1 \quad \text{for all } w \iff \eta_1 = \eta_2.$$

Condition (C5) is satisfied for many parametric families of densities, like for example, the families of Weibull, log-normal and log-logistic densities; see Czado and Van Keilegom (2023) and Deresa et al. (2022) for more details. For the marginal distribution of T, note that we do not need to impose any further assumptions other than those mentioned in Section 2.

In addition to (C5), we need the following conditions on the dependence structure, which is also required to identify the joint model:

(C6) (*i*) For all $(\gamma, \beta, \Lambda, \eta) \in \Gamma \times \mathbb{R}^p \times \mathcal{G} \times H$,

$$\lim_{t\to 0} h_{T|C,\gamma}(F_{T|X,\zeta}(t|x)|F_{C|W,\eta}(t|w)) = 0 \ \text{ for all } (x, w),$$

where $\zeta = (\beta, \Lambda)$. The same holds true for $h_{C|T,\gamma}(F_{C|W,\eta}(t|w)|F_{T|X,\zeta}(t|x))$.
(*ii*) For all $\gamma_k, \zeta_k = (\beta, \Lambda_k), \eta$ $(k = 1, 2)$ that are such that $\lim_{t\to 0} \lambda_1(t)/\lambda_2(t) = 1$, we have

$$\lim_{t\to 0} \frac{c_{\gamma_1}(F_{T|X,\zeta_1}(t|x), F_{C|W,\eta}(t|w))}{c_{\gamma_2}(F_{T|X,\zeta_2}(t|x), F_{C|W,\eta}(t|w))} = 1$$
$$\text{for all } (x,w) \iff \gamma_1 = \gamma_2,$$

where c_γ denotes the copula density.

We will verify condition (C6) for two popular classes of copulas, specifically Archimedean copulas and Gaussian copulas. Here are some examples of commonly used Archimedean copulas, including the Frank copula (Frank 1979; Genest 1987), given by

$$\mathcal{C}_\gamma(u,v) = -\frac{1}{\gamma}\log\left\{1 + \frac{(e^{-\gamma u}-1)(e^{-\gamma v}-1)}{e^{-\gamma}-1}\right\}, \quad \gamma \ne 0,$$

the Gumbel copula (Gumbel 1960), given by

$$\mathcal{C}_\gamma(u,v) = \exp\left\{-\left[(-\log u)^\gamma + (-\log v)^\gamma\right]^{1/\gamma}\right\}, \quad \gamma \ge 1,$$

and the Clayton copula (Clayton 1978), given by

$$\mathcal{C}_\gamma(u,v) = (u^{-\gamma} + v^{-\gamma} - 1)^{-1/\gamma}, \quad \gamma > 0.$$

The most commonly used member of the family of elliptical copulas is the Gaussian copula, given by

$$\mathcal{C}_\gamma(u,v) = \Phi_\gamma\big(\Phi^{-1}(u), \Phi^{-1}(v)\big),$$

where $-1 < \gamma < 1$, Φ_γ is the cumulative distribution function of a standard bivariate normal distribution with correlation parameter γ, and Φ^{-1} is the quantile function of a standard normal variable. Note that the dependence parameter γ has a one-to-one relation with Kendall's τ parameter for these common copula families, which is easier to interpret than γ because it determines the strength of the association that is universal for all copula families; see Nelsen (2006) for details.

In the next lemma we show how condition (C6) can be validated for the above families of common copulas. The proof of Lemma 3.1 (*a*) is similar to that of Theorem 3.3 in Czado and Van Keilegom (2023), and thus it is not shown here, whereas the proof of Lemma 3.1 (*b*) is given in the supplementary material. Note that the Clayton copula is not included in this lemma, as the identifiability cannot be proved the same way as for the other common copulas. We refer to the latter paper for more details, and focus on the other copulas in what follows.

Lemma 3.1.

(a) Assumption (C6) (*i*) is satisfied by

1. the Frank copula, independently of the marginal distributions;
2. the Gumbel copula if $\lim_{t\to 0} \log F_{T|X,\zeta}(t|x)/\log F_{C|W,\eta}(t|w) \in (0,\infty)$ for all (x,w) and all (ζ,η);
3. the Gaussian copula if $\lim_{t\to 0} J_\alpha(t|x,w) = -\infty$ and $\lim_{t\to 0} L_\alpha(t|x,w) = -\infty$ for all (x,w,α), where $J_\alpha(t|x,w) = \Phi^{-1}(F_{T|X,\zeta}(t|x)) - \gamma\Phi^{-1}(F_{C|W,\eta}(t|w))$ and $L_\alpha(t|x,w) = \Phi^{-1}(F_{C|W,\eta}(t|w)) - \gamma\Phi^{-1}(F_{T|X,\zeta}(t|x))$.

(b) Assumption (C6) (*ii*) is satisfied for the Frank, Gumbel and Gaussian copulas.

We refer to Czado and Van Keilegom (2023) and Deresa et al. (2022) for more details regarding the required conditions for the Gumbel and Gaussian copula. We are now ready to give our identifiability result in the following theorem. The proof is given in the supplementary material.

Theorem 3.1. Assume that conditions (C1)–(C6) hold true. Then, model (2.1)–(2.3) is identifiable. This means that if for $k = 1, 2$, the pair (T_k, C_k, A_k) satisfies model (2.1)–(2.3) with parameter vector $\alpha_k = (\gamma_k, \beta_k, \Lambda_k, \eta_k) \in \Gamma \times \mathbb{R}^p \times \mathcal{G} \times H$, and if $(Z_1, \Delta_{11}, \Delta_{21})$ given (X, W) and $(Z_2, \Delta_{12}, \Delta_{22})$ given (X, W) have the same distribution, then

$$\gamma_1 = \gamma_2, \beta_1 = \beta_2, \Lambda_1 = \Lambda_2, \eta_1 = \eta_2,$$

where $Z_k = \min(T_k, C_k, A_k)$, $\Delta_{1k} = \mathbf{I}(Z_k = T_k)$ and $\Delta_{2k} = \mathbf{I}(Z_k = C_k)$.

4. Estimation

We will now discuss an estimation procedure for models (2.1)–(2.3). Assume that the data consist of n iid replications $(Z_i, \Delta_{1i}, \Delta_{2i}, X_i, W_i)$ $(i = 1, \ldots, n)$ of $(Z, \Delta_1, \Delta_2, X, W)$, and let $\theta = (\gamma, \beta, \eta)$. Then, the joint likelihood function is derived from (2.5) by

$$\begin{aligned} L(\theta,\Lambda) &= \prod_{i=1}^n g_{\theta,\Lambda}(Z_i,\Delta_{1i},\Delta_{2i}|X_i,W_i) \qquad (4.1)\\ &= \prod_{i=1}^n \Big[\lambda(Z_i)e^{X_i^\top\beta}\exp\{-\Lambda(Z_i)e^{X_i^\top\beta}\}\\ &\quad\times\big\{1-h_{C|T,\gamma}(F_{C|W,\eta}(Z_i|W_i)|F_{T|X,\zeta}(Z_i|X_i))\big\}\Big]^{\Delta_{1i}}\\ &\quad\times\Big[f_{C|W,\eta}(Z_i|W_i)\big\{1-h_{T|C,\gamma}(F_{T|X,\zeta}(Z_i|X_i)\\ &\qquad|F_{C|W,\eta}(Z_i|W_i))\big\}\Big]^{\Delta_{2i}}\\ &\quad\times\Big[\tilde{\mathcal{C}}_\gamma\Big\{F_{T|X,\zeta}(Z_i|X_i),F_{C|W,\eta}(Z_i|W_i)\Big\}\Big]^{(1-\Delta_{1i})(1-\Delta_{2i})} \end{aligned}$$

since assumption (C4) implies that the density and distribution of A can be omitted from the likelihood function.

The direct maximization of this likelihood can be challenging since it involves the unknown function Λ. The idea is now to estimate θ by replacing the unknown function in the above likelihood with a nonparametric estimator for a fixed value of θ. The parameter θ is then estimated by solving the score equation derived from the pseudolikelihood function, $L(\theta,\Lambda)$ with Λ replaced by its estimated value.

Let us first construct our nonparametric estimator for Λ by forming an estimating equation when T and C are dependent given the set of covariates. This will be done based on martingale ideas. Let $N_i(z) = \mathbf{I}(Z_i \le z, \Delta_{1i} = 1)$ and $Y_i(z) = \mathbf{I}(Z_i \ge z)$ for all $i \in \{1,\ldots,n\}$. Using Theorem 1.3.1 in Fleming and Harrington (1991), it follows that

$$M_i(z) = N_i(z) - \int_0^z Y_i(s)\lambda^\#(s|X_i,W_i)ds$$

is a martingale with respect to the filtration $\mathcal{F}_z^i = \sigma\{Y_i(s), N_i(s), X_i, W_i; 0 \le s \le z \le \tau\}$, where $\lambda^\#(z|X,W)$, the conditional crude hazard rate, is given by

$$\lambda^\#(z|X,W) = \frac{-\frac{\partial}{\partial u}\Pr(T\ge u, C\ge z|X,W)|_{u=z}}{\Pr(T\ge z, C\ge z|X,W)},$$

and τ_0 is a finite maximum follow-up time. We refer to Rivest and Wells (2001) for a similar martingale construction under dependent censoring. Next, define the following functions:

$$\begin{aligned} A_\alpha(t|x,w) &= h_{C|T,\gamma}(F_{C|W,\eta}(t|w)|F_{T|X,\zeta}(t|x)),\\ B_\alpha(t|x,w) &= h_{T|C,\gamma}(F_{T|X,\zeta}(t|x)|F_{C|W,\eta}(t|w)). \end{aligned}$$

Under the general parametric copula model given in (2.3), we have that

$$M_i(z) = N_i(z) - \int_0^z Y_i(s)\exp(\psi_i(s,\theta_0,\Lambda_0))d\Lambda_0(s), \quad (4.2)$$

where

$$\begin{aligned} \psi_i(z,\theta_0,\Lambda_0) &= X_i^\top\beta_0 - \Lambda_0(z)\exp(X_i^\top\beta_0)\\ &\quad+\log(1-A_{\alpha_0}(z|X_i,W_i)) - \log(S_{\alpha_0}(z|X_i,W_i)), \end{aligned}$$

where $\theta_0 = (\gamma_0, \beta_0, \eta_0)$ and Λ_0 are the true values of θ and Λ, respectively, and $\alpha_0 = (\theta_0, \Lambda_0)$. Therefore, motivated by the fact that $M_i(z)$ is a martingale process, we estimate Λ for a given θ by solving the following estimating equation:

$$\sum_{i=1}^{n}\{\mathrm{d}N_i(z) - Y_i(z)\exp(\psi_i(z,\theta,\Lambda))d\Lambda(z)\} = 0 \quad (0 \le z \le \tau_0), \tag{4.3}$$

with Λ satisfying $\Lambda(0) = 0$. It is easily seen that the estimator of Λ is a nondecreasing step function with jumps only at the observed survival times, denoted by $z_1 < \cdots < z_K < \infty$.

However, we will not use (4.3) directly to estimate the cumulative hazard since it involves a complex iterative optimization process. Instead, we propose an alternative version of (4.3) that is simple for computational purposes. For a given value of θ, we estimate Λ as a step function with jumps at the ordered observed survival times $z_1, \ldots, z_K$, using the following equation:

$$\Delta\hat{\Lambda}(z_k,\theta) = \frac{\sum_{i=1}^{n} dN_i(z_k)}{\sum_{i=1}^{n} Y_i(z_k)\exp\{\psi_i(z_{k-1},\theta,\hat{\Lambda})\}}, \tag{4.4}$$

with $\hat{\Lambda}(z_1-,\theta) = 0$ and $\Delta\hat{\Lambda}(z_k,\theta) = \hat{\Lambda}(z_k,\theta) - \hat{\Lambda}(z_{k-1},\theta)$. Note that the estimator of the kth jump depends on $\hat{\Lambda}$ up to and including time z_{k-1}. This technique avoids the iterative optimization scheme for estimating the cumulative hazard.

We now estimate θ by substituting $\hat{\Lambda}(z,\theta)$ for $\Lambda(z)$ in (4.1) and setting the derivative of the resulting pseudo-likelihood function with respect to θ to zero. This gives the following estimating equation:

$$U_n(\theta,\hat{\Lambda}) = n^{-1}\sum_{i=1}^{n} U(Z_i,\Delta_{1i},\Delta_{2i},\theta,\hat{\Lambda}(\cdot,\theta)) = 0, \tag{4.5}$$

where

$$U(Z_i,\Delta_{1i},\Delta_{2i},\theta,\Lambda) = \frac{\partial}{\partial\theta}\log g_{\theta,\Lambda}(Z_i,\Delta_{1i},\Delta_{2i}|X_i,W_i).$$

Finally, $\hat{\theta}$ is defined as a solution of this score equation. Note that Λ can be estimated without performing an iterative optimization, but we need to estimate θ using an optimization algorithm. The details of our estimation algorithm can be found in the supplementary material.

Remark 1. In the particular case where the independence copula $\mathcal{C}_\gamma(u,v) = uv$ is specified, the estimator of Λ cancels out from the formula of $\psi_i(z_{k-1},\theta,\hat{\Lambda})$. Consequently, our nonparametric estimator of Λ in (4.4) reduces to the standard Breslow estimator for the cumulative hazard function in the Cox model (Breslow 1974). However, when $\mathcal{C}_\gamma$ is different from the independence copula, the quantity $\psi_i(z_{k-1},\theta,\hat{\Lambda})$ in (4.4) involves $\hat{\Lambda}(z_{k-1})$, contrary to the denominator of the Breslow estimator. In this respect, our estimator is similar to that of Zucker (2005). Hence, we will show the consistency of $\hat{\Lambda}$ in the next section using similar arguments as in Zucker's consistency proof for the estimated cumulative hazard function.

5. Asymptotic Properties

This section gives the consistency and the asymptotic normality of $\hat{\theta}$ and $\hat{\Lambda}$ and the set of assumptions needed to prove these results. For the Cox model (and for some variants thereof), the partial likelihood function does not involve infinite-dimensional parameters, and hence it is relatively easy to establish an asymptotic theory. However, for the proposed approach, the estimating equations unavoidably involve both finite and infinite-dimensional parameters. Hence, the derivation of the asymptotic theory is much more demanding. We will show this in several steps. Let us first introduce the following quantities:

$$\mu_1(v|u) = \frac{h^{10}_{C|T,\gamma}(v|u)}{1-h_{C|T,\gamma}(v|u)} \qquad \mu_2(v|u) = \frac{h^{01}_{C|T,\gamma}(v|u)}{1-h_{C|T,\gamma}(v|u)}$$

$$\mu_3(v|u) = \frac{\dot{h}_{C|T,\gamma}(v|u)}{1-h_{C|T,\gamma}(v|u)} \qquad \nu_1(t|x,w) = \frac{1-A_\alpha(t|x,w)}{S_\alpha(t|x,w)}$$

$$\nu_2(t|x,w) = \frac{1-B_\alpha(t|x,w)}{S_\alpha(t|x,w)} \qquad \nu_3(t|x,w) = \frac{\partial S_\alpha(t|x,w)/\partial\gamma}{S_\alpha(t|x,w)}$$

$$B(z) = E[Y_i(z)\exp(\psi_i(z,\theta_0,\Lambda_0))],$$

$$B_1(z) = E[Y_i(z)\dot{\psi}_i(z,\theta_0,\Lambda_0)\exp(\psi_i(z,\theta_0,\Lambda_0))],$$

$$A(z) = \exp\Big(\int_0^z \frac{B_1(s)}{B(s)}d\Lambda_0(s)\Big),$$

where $h^{10}_{C|T,\gamma}(v|u) = \partial h_{C|T,\gamma}(v|u)/\partial u$, $h^{01}_{C|T,\gamma}(v|u) = \partial h_{C|T,\gamma}(v|u)/\partial v$, $\dot{h}_{C|T,\gamma}(v|u) = \partial h_{C|T,\gamma}(v|u)/\partial\gamma$, $\dot{\psi}_i(z,\theta_0,\Lambda_0) = \frac{\partial\psi_i(z,\theta_0,\xi)}{\partial\xi}\Big|_{\xi=\Lambda_0(z)}$. We further define

$$\Sigma_1 = E\Big[\frac{\partial U(Z,\Delta_1,\Delta_2,\theta_0,\Lambda_0)}{\partial\theta^\top}\Big] + E\Big[\frac{\partial U(Z,\Delta_1,\Delta_2,\theta_0,\xi)}{\partial\xi}\Big|_{\xi=\Lambda_0(Z)}\frac{1}{A_1(Z)}\int_0^Z \frac{A_1(s)}{B(s)}dD(s)\Big],$$

and

$$\Sigma_2 = E\Big[\Big\{U(Z_i,\Delta_{1i},\Delta_{2i},\theta_0,\Lambda_0) - \int_0^{\tau_0} C(s)dM_i(s)\Big\}^{\otimes 2}\Big],$$

where $e^{\otimes 2} = ee^\top$ for any vector e,

$$C(z) = E\Big[\frac{\partial U(Z,\Delta_1,\Delta_2,\theta_0,\xi)}{\partial\xi}\Big|_{\xi=\Lambda_0(Z)}\frac{\mathbf{I}(z\le Z)}{A(Z)}\Big]\frac{A(z)}{B(z)},$$

$$\begin{aligned}D(z) = E\Big(\int_0^z \Big[&Y_i(s)\Big\{(1-\Lambda_0(s)\exp(X_i^\top\beta_0))\frac{\partial X_i^\top\beta_0}{\partial\theta^\top} \\ &- \mu_2(F_{C|W,\eta_0}(s|W_i)|F_{T|X,\zeta_0}(s|X_i)) \\ &\times\frac{\partial F_{C|W,\eta_0}(s|W_i)}{\partial\theta^\top} - \mu_3(F_{C|W,\eta_0}(s|W_i)|F_{T|X,\zeta}(s|X_i))\frac{\partial\gamma_0}{\partial\theta^\top} \\ &+ \nu_2(s|X_i,W_i)\frac{\partial F_{C|W,\eta_0}(s|W_i)}{\partial\theta^\top} \\ &- \nu_3(s|X_i,W_i)\frac{\partial\gamma_0}{\partial\theta^\top} - \Big(\mu_1(F_{C|W,\eta_0}(s|W_i)|F_{T|X,\zeta_0}(s|X_i)) \\ &- \nu_1(s|X_i,W_i)\Big)\Lambda_0(s)\end{aligned}$$

$$\times \exp(X_i^\top \beta_0 - \Lambda_0(s) e^{X_i^\top \beta_0}) \frac{\partial X_i^\top \beta_0}{\partial \theta^\top} \Big\}$$
$$\times \exp(\psi_i(s, \theta_0, \Lambda_0(s))) d\Lambda_0(s) \Big] \Big),$$

and

$$A_1(z) = \exp\Big\{ \int_0^z \frac{B_2(s)}{B(s)} d\Lambda_0(s) \Big\},$$

where

$$\begin{aligned} B_2(z) &= E\Big[Y_i(z) \big\{ \mu_1(F_{C|W,\eta_0}(z|W_i) | F_{T|X,\zeta_0}(z|X_i)) \\ &\quad - \nu_1(z|X_i, W_i) \big\} \exp(X_i^\top \beta_0 - \Lambda_0(z) e^{X_i^\top \beta_0} \\ &\quad + \psi_i(z, \theta_0, \Lambda_0(z)) \Big]. \end{aligned}$$

In addition to assumptions (C1)–(C6), we need the following regularity conditions to establish asymptotic properties of our estimators ($\|\cdot\|$ denotes the Euclidean norm of a vector). Let $\Lambda_0(z, \theta)$ be the solution of the expected value of (4.3).

(D1) The parameter vector θ lies in a compact set, say Θ, that contains an open neighborhood of the true parameter vector θ_0.
(D2) The covariate vectors X and W have bounded support.
(D3) The function $\Lambda_0(t)$ is monotone increasing and differentiable with derivative $\lambda_0(t)$. In addition, $\lambda_0(t)$ is bounded by some constant $\lambda_{\max}$ for $t \in [0, \tau_0]$.
(D4) There is a finite maximum follow-up time $\tau_0 > 0$, with $y^* = \Pr(Y_i(\tau_0) = 1) > 0$.
(D5) We have that $\Pr(T > \tau_0, C > \tau_0 | X = x, W = w) > 0$ and $\Pr(A > \tau_0) > 0$ for all x and w.
(D6) $A_{\theta,\Lambda}(t|x, w)$, $B_{\theta,\Lambda}(t|x, w)$ and $\tilde{\mathcal{C}}_\gamma \big\{ F_{T|X,\zeta}(t|x), F_{C|W,\eta}(t|w) \big\}$ exist and are twice continuously differentiable with respect to the components of θ. In addition, all derivatives of order two are bounded, uniformly in Λ, t, x and w.
(D7) Σ_1 is a finite and nonsingular matrix.
(D8) For all $\delta > 0$, there exists $\epsilon > 0$ such that $\inf_{\|\theta - \theta_0\| > \delta} \|E[U(Z, \Delta_1, \Delta_2, \theta, \Lambda_0(\cdot, \theta))]\| > \epsilon$.

Assumptions (D1)–(D4) are similar to those in Zucker (2005) for showing the asymptotic properties of a pseudo-partial likelihood estimator. Assumption (D5) is a technical assumption required to ensure that the denominator of our nonparametric estimator and that denominator in various expansions used in the proofs are nonzero. Finally, assumptions (D6)–(D8) are needed for the application of Theorems 1 and 2 in Chen et al. (2003).

Before we state the asymptotic properties of the estimators, we need some preliminary results stated in the lemma below, which will be needed in the proof of the main theorem. All proofs are given in the supplementary material.

Lemma 5.1. Assume that conditions (D1)–(D8) hold true, and that assumptions (C1)–(C6) are satisfied. Then,

(i) Boundedness of $\hat{\Lambda}(\cdot, \theta)$: There exists some n' such that for all $z \in [0, \tau_0]$ and $\theta \in \Theta$,

$$\hat{\Lambda}(z, \theta) \leq 1.01/(y^* \phi_{\min}) \text{ for all } n \geq n',$$

where $\phi_{\min}$ is a lower bound on $\exp(\psi_i(z, \theta, \Lambda))$.

(ii) Consistency and rate of convergence of $\hat{\Lambda}(\cdot, \theta)$:

$$\sup_{\theta \in \Theta, 0 \leq z \leq \tau_0} |\hat{\Lambda}(z, \theta) - \Lambda_0(z, \theta)| = O_p(n^{-1/2}).$$

(iii) Iid representation of $\hat{\Lambda}(\cdot, \theta_0) - \Lambda_0(\cdot)$:

$$\begin{aligned} &\hat{\Lambda}(z, \theta_0) - \Lambda_0(z) \\ &\quad = \frac{1}{A(z)} \frac{1}{n} \sum_{i=1}^n \int_0^z \frac{A(s)}{B(s)} dM_i(s) + R_n(z), \quad (5.1) \end{aligned}$$

where $\sup_{0 \leq z \leq \tau_0} |R_n(z)| = o_p(n^{-1/2})$.

(iv) Consistency of $(\partial/\partial\theta)\hat{\Lambda}(\cdot, \theta_0)$: For every $z \in [0, \tau_0]$,

$$\left. \frac{\partial \hat{\Lambda}(z, \theta)}{\partial \theta} \right|_{\theta = \theta_0} = \frac{1}{A_1(z)} \int_0^z \frac{A_1(s)}{B(s)} dD(s) + o_p(1). \quad (5.2)$$

For any Λ, let

$$\|\Lambda\|_{\mathcal{G}} = \sup_{0 \leq z \leq \tau_0, \theta \in \Theta} |\Lambda(z, \theta)|.$$

Also, note that $E[U(Z, \Delta_1, \Delta_2, \theta_0, \Lambda_0)] = 0$ and $U_n(\hat{\theta}, \hat{\Lambda}(\cdot, \hat{\theta})) = 0$. We will show the asymptotic properties of $\hat{\theta}$ using the results in Chen et al. (2003). The latter paper gives sufficient conditions for the consistency and asymptotic normality of a class of semiparametric Z-estimators. We present in the following theorem the consistency and the asymptotic distribution of $\hat{\theta}$. The proof is deferred to the supplementary material.

Theorem 5.1. Assume that conditions (D1)–(D8) hold true, and that assumptions (C1)–(C6) are satisfied. Then,

(i) Consistency of $\hat{\theta}$: $\hat{\theta} \xrightarrow{P} \theta_0$.
(ii) Asymptotic normality of $\hat{\theta}$:

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow \mathcal{N}\{0, \Sigma_1^{-1} \Sigma_2 (\Sigma_1^{-1})^\top\},$$

in distribution.

The proof of Theorem 5.1 (i) is done by verifying the conditions of Theorem 1 in Chen et al. (2003). The important requirement to apply the results in the latter paper is a uniform consistency of the estimator of the infinite-dimensional parameter obtained in Lemma 5.1 (ii). Moreover, Theorem 5.1 (ii) reports the limiting distribution of $\hat{\theta}$, for which the crucial requirement is now a linear representation of the infinite-dimensional parameter, which is shown in Lemma 5.1 (iii).

Given that the asymptotic variance derived in Theorem 5.1 has a complicated expression involving integrals, we rely on a nonparametric bootstrap approach for drawing inference. This approach consists in drawing resamples $(Z_i^*, \Delta_{1i}^*, \Delta_{2i}^*, X_i^*, W_i^*)$ $(i = 1, \ldots, n)$ randomly with replacement from the original sample $(Z_i, \Delta_{1i}, \Delta_{2i}, X_i, W_i)$ $(i = 1, \ldots, n)$. For the bootstrap sample, the parameter estimates $\hat{\theta}^*$ and $\hat{\Lambda}^*$ are obtained following the same procedure as in Section 4. Then the asymptotic variance of $\hat{\theta}$ can be approximated by the empirical variance of $\hat{\theta}^*$.

Table 1. Bias, empirical standard deviation (ESD), and root mean squared error (RMSE) for samples of size $n = 500$ using the Frank and independence copulas.

	$\tau = 0.2$			$\tau = 0.4$			$\tau = 0.8$		
	Bias	ESD	RMSE	Bias	ESD	RMSE	Bias	ESD	RMSE
				Frank copula					
β_1	−0.010	0.134	0.134	−0.014	0.131	0.131	−0.024	0.126	0.129
β_2	−0.003	0.098	0.098	−0.004	0.099	0.099	−0.013	0.102	0.103
η_0	−0.005	0.139	0.139	−0.006	0.123	0.123	−0.022	0.113	0.115
η_1	0.001	0.136	0.136	0.004	0.125	0.125	0.013	0.110	0.111
η_2	−0.002	0.118	0.118	−0.002	0.109	0.109	−0.012	0.099	0.100
σ	−0.002	0.052	0.052	−0.001	0.052	0.052	0.003	0.051	0.051
τ	0.012	0.112	0.112	0.010	0.090	0.090	0.008	0.037	0.038
				Independence copula					
β_1	0.024	0.135	0.137	0.058	0.137	0.149	0.124	0.141	0.188
β_2	0.081	0.085	0.118	0.167	0.087	0.188	0.327	0.092	0.340
η_0	0.165	0.111	0.199	0.314	0.109	0.333	0.520	0.110	0.532
η_1	0.052	0.140	0.150	0.100	0.137	0.170	0.169	0.133	0.215
η_2	0.129	0.095	0.160	0.248	0.095	0.265	0.415	0.101	0.427
σ	0.001	0.055	0.055	−0.013	0.055	0.057	−0.082	0.053	0.098

6. Simulation Study

In this section we study the finite-sample performance of the proposed estimators. First, we investigate the effect of not considering the dependence between T and C in models (2.1)–(2.3). We will do this by comparing our method with the method that assumes independent censoring. Second, we assess the performance of the proposed method when the copula structure is misspecified. Finally, we evaluate the convergence of the parameter estimators to a normal limit for finite-sample sizes. The simulation study is summarized using the following three scenarios. We refer to the supplementary material for additional simulation examples.

Scenario 1. Under this scenario, we will compare our model with the model that assumes that T and C are independent. We consider the following data generating model:

$$\Pr(T \le z, C \le z | X = x, W = x) = \mathcal{C}_\gamma\{F_{T|X}(z|x), F_{C|W}(z|x)\},$$

where the copula $\mathcal{C}_\gamma$ is a Frank copula with association parameter γ specified to give a Kendall's τ of 0.2, 0.4 or 0.8. We specified the margin of T by the Cox proportional hazards model:

$$F_{T|X}(z|x) = 1 - \exp\Big(- \Lambda(z) e^{\beta_1 x_1 + \beta_2 x_2}\Big)$$

and the margin of C by the Weibull model:

$$F_{C|W}(z|x) = 1 - \exp\Big(- \exp\Big(\frac{\log(z) - \mu(x)}{\sigma}\Big)\Big),$$

where $x = (x_1, x_2)^\top$, $\mu(x) = \eta_0 + \eta_1 x_1 + \eta_2 x_2$, $\Lambda(z) = 0.25z^{3/4}$, $W = X = (X_1, X_2)^\top$, $X_1 \sim \text{Bern}(0.5)$, $X_2 \sim \mathcal{N}(0, 1)$, and X_1 and X_2 are independent. The regression parameters are set as follows: $\beta_1 = 0.45, \beta_2 = 1, \eta_0 = 1.35, \eta_1 = 0.3, \eta_2 = 1, \sigma = 1$. In addition, the administrative censoring variable A is generated from a uniform distribution on $[0, 15]$ independent of all other variables. Under this scenario, the average proportion of observations in the simulated data is approximately 45% T, 40% C, and 15% A.

We created a total of 1000 datasets with a sample size of $n = 500$. For each of them, the model parameters are estimated under the Frank copula and the independence copula model corresponding to $\tau = 0$. Note that for the marginal distribution of C, we work with the Weibull model. In Table 1 we present the bias, the Empirical Standard Deviation (ESD), and the root mean squared error (RMSE) of the parameter estimates based on the 500 replications. The table shows that the Frank copula model performs better than the model assuming independence, even when there is a weak dependence between T and C (see the case $\tau = 0.2$ in the table). While the estimates for the model parameters are unbiased under the Frank copula, they are biased when the censoring mechanism is wrongly assumed to be independent, especially when $\tau = 0.4$ and 0.8. Also, as expected, the biases and RMSE for the independence copula increase as we increase τ.

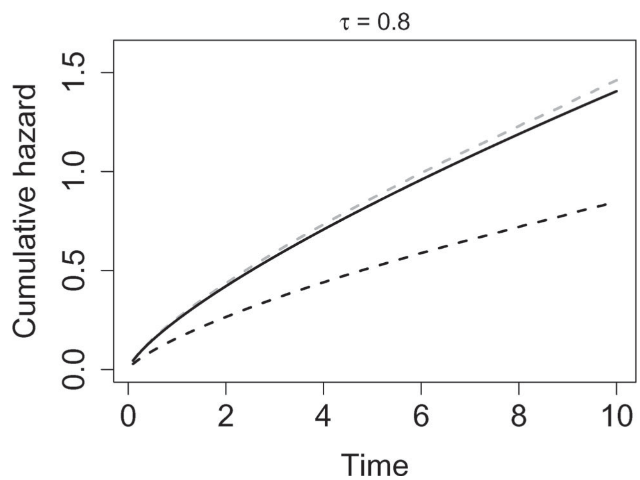

Figure 1. The average of the estimated cumulative hazard functions based on the Frank copula (dashed gray line) and independence copula (dashed black line) overlaid with the true cumulative hazard function $\Lambda(z) = 0.25z^{3/4}$ (solid line).

Moreover, we investigate the performance of the proposed method by estimating the cumulative hazard function $\Lambda(z) = 0.25z^{3/4}$ of the survival time T. The average of the 500 estimated cumulative hazard functions based on the Frank and independence copulas for the case $\tau = 0.8$ is shown in Figure 1. It can be observed that the average of the estimated functions $\hat{\Lambda}(z, \hat{\theta})$ based on the Frank copula model is very close to the

Table 2. Results for bias, empirical standard deviation (ESD), and root mean squared error (RMSE) based on Gumbel and Gaussian copulas when the data are generated from a Frank copula.

	$\tau = 0.2$			$\tau = 0.4$			$\tau = 0.8$		
	Bias	ESD	RMSE	Bias	ESD	RMSE	Bias	ESD	RMSE
					Gumbel copula				
β_1	−0.009	0.139	0.140	−0.021	0.138	0.139	−0.019	0.126	0.128
β_2	0.006	0.104	0.104	−0.003	0.108	0.108	−0.012	0.099	0.100
η_0	−0.005	0.158	0.158	−0.022	0.141	0.142	−0.026	0.112	0.115
η_1	0.002	0.139	0.139	0.000	0.129	0.129	0.006	0.112	0.112
η_2	−0.004	0.135	0.135	−0.018	0.124	0.125	−0.019	0.100	0.102
σ	−0.014	0.052	0.054	−0.014	0.052	0.053	0.006	0.051	0.052
τ	−0.001	0.130	0.130	0.013	0.109	0.110	0.000	0.036	0.036
					Gaussian copula				
β_1	−0.013	0.135	0.135	−0.018	0.132	0.133	−0.008	0.124	0.125
β_2	−0.011	0.107	0.107	−0.016	0.105	0.106	−0.000	0.098	0.098
η_0	−0.018	0.159	0.160	−0.021	0.132	0.133	−0.001	0.107	0.107
η_1	−0.002	0.139	0.139	−0.000	0.127	0.127	0.012	0.112	0.113
η_2	−0.010	0.133	0.133	−0.012	0.117	0.117	0.007	0.097	0.097
σ	0.004	0.053	0.054	0.012	0.053	0.054	0.020	0.053	0.056
τ	0.022	0.140	0.142	0.014	0.096	0.097	−0.047	0.043	0.064

true cumulative hazard function $\Lambda(z)$, which suggests that the proposed estimation method gives acceptable estimates of the true cumulative hazard function. However, under the false independence assumption, the average of the estimated cumulative hazard functions is far from the truth.

Scenario 2. Here, we would like to know whether the estimated model parameters are sensitive to the misspecification of the copula structure. In order to study this, we analyze the data simulated above under Scenario 1 using the correct marginal distributions but with a misspecified copula. Table 2 displays the bias, ESD and RMSE of the parameter estimates based on Gumbel and Gaussian copulas when the data are generated from a Frank copula model. It is seen that the misspecification in the copula structure has little to no influence on the proposed method under the considered setting. We also show the average of the estimated cumulative hazard functions for the case $\tau = 0.8$ in Figure 3 of the supplementary material. The figure shows that the average of the estimated cumulative hazard functions based on the Gumbel and Gaussian copulas are very close to the true cumulative hazard function. This result is somewhat similar to the results obtained in Huang and Zhang (2008) and Chen (2010), who showed that the bias due to misspecification of the copula structure is usually moderate.

Scenario 3. In this scenario, we examine how close is the sampling distribution of $\hat{\theta}$ to a normal distribution for samples of size $n = 500$. We simulate data based on a Gaussian copula with $\gamma = 0.75$, corresponding to a Kendall's τ of 0.54. The marginal distribution of T follows the same Cox model as under Scenario 1, except that $\Lambda(z) = 0.25z^{1/2}$, and the margin of C follows a lognormal distribution, specified by

$$F_{C|W}(z|x) = \Phi\Big(\frac{\log(z) - \mu(x)}{\sigma}\Big),$$

where $\mu(x)$, X_1 and X_2 are defined in Scenario 1. The other model parameters are set as $\beta_1 = 0.45, \beta_2 = 1, \eta_0 = 1, \eta_1 = 0.5, \eta_2 = 0.75$ and $\sigma = 1.5$. The censoring variable A is generated from a $\mathcal{U}[0, 25]$ distribution independently of everything else. We estimate the model parameters under a Gaussian copula model using the proposed method. Figure 2 provides the quantile-quantile plots for the estimated parameters based on 500 generated samples, and it can be observed that the distribution of the parameter estimators is close to a normal limit. However, for $\hat{\tau}$, there is a slight deviation from the straight line in the left tail. After applying Fisher's Z transformation, the distribution appears to be more symmetric (see the plot for $\hat{\omega}$ in Figure 2), where $\hat{\omega}$ is Fisher's Z transformation of $\hat{\tau}$.

Table 3 presents the bias, ESD, bootstrap standard error (BSE), RMSE, and 95% coverage rate (CR) using the above model. The bootstrap standard errors are computed from 150 bootstrap resamples for each of 500 samples, and the coverage rate is calculated using the normal approximation with bootstrap standard error. To get the confidence interval for τ, we first apply Fisher's Z transformation and the Delta method to construct the confidence interval on the transformed scale and then transform the confidence interval limits to the original scale. It is clear from the table that the bootstrap standard errors are very close to the corresponding empirical standard deviations, which suggests that the bootstrap method works well in practice. The coverage rates are close to the 95% nominal level, which also indicates that the asymptotic normality of our estimators is approximately satisfied. Moreover, the average of the estimated cumulative hazard functions based on the Gaussian copula is close to the true hazard function; see Figure 3 in the supplementary material.

We also conducted a simulation study to examine the sensitivity of the proposed method to the misspecification of the censoring model or to the misspecification of the copula structure and censoring model. From Table 6 of the supplementary material, we see that the parameter estimators for the survival model still behave well when the censoring model or when both the censoring model and copula structure are misspecified. However, compared to the results in Table 3 (when there is no misspecification), we notice a larger bias for model parameters and a decrease in coverage rate for Kendall's τ. The details can be

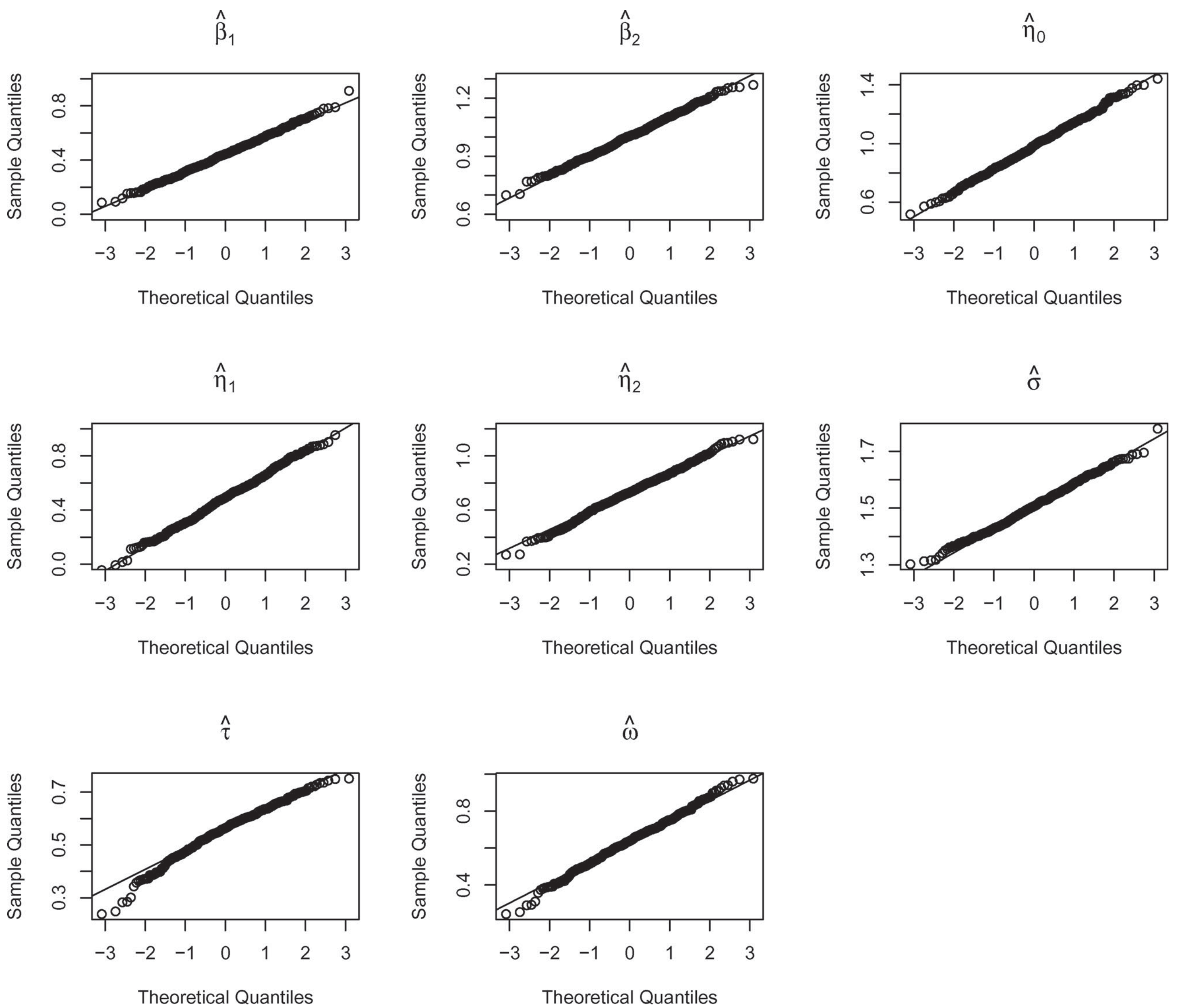

Figure 2. Normal quantile-quantile plot for the estimates ($\hat{\beta}_1, \hat{\beta}_2, \hat{\eta}_0, \hat{\eta}_1, \hat{\eta}_2, \hat{\sigma}, \hat{\tau}, \hat{\omega}$) based on 500 simulated data.

found in Scenario 4 of the supplementary material. To assist the user of our model in evaluating the quality of the fitted model, we proposed a formal goodness-of-fit test in the supplementary material. The test is based on the L_2 distance between a model-based estimator and a model-free estimator of the distribution function of the minimum of T and C. In the simulations, this test shows reasonable control of the Type I error rate and exhibits a good power in rejecting a misspecified censoring model. We refer to Scenario 5 of the supplementary material for details.

Table 3. Bias, empirical standard deviation (ESD), bootstrap standard error (BSE), root mean squared error (RMSE) and coverage rate (CR) based on the Gaussian copula model.

Par.	Bias	ESD	BSE	RMSE	CR
β_1	−0.006	0.130	0.136	0.130	0.959
β_2	0.002	0.100	0.104	0.100	0.958
η_0	−0.021	0.160	0.172	0.161	0.968
η_1	−0.011	0.178	0.173	0.179	0.950
η_2	−0.024	0.148	0.156	0.150	0.948
σ	0.008	0.076	0.078	0.077	0.958
τ	0.019	0.084	0.085	0.086	0.928

7. Data Application

We now apply the proposed model and estimation method to a follicular cell lymphoma dataset, given in the book of Pintilie (2006) and collected at the Princess Margaret Hospital, Toronto, where patients entered the database as they registered for treatment. The data given in Pintilie (2006) consist of 541 patients with early disease stage (I or II) and treated with Radiation Alone (RT) or with Radiation and Chemotherapy (CMT). In this study the endpoints of interest are what comes first: relapse of the disease or death in remission. We are mainly interested in the marginal distribution of the time to relapse, so death in remission is considered as a (dependent) censoring event, whereas censoring at the end of the study is considered as independent censoring.

Among the 541 patients, 272 (50.3%) of them experienced disease relapse, another 76 (14.0%) patients died in remission, and the remaining 193 (35.7%) patients were censored at the end of the study. We use the following four covariates in our analysis: Treatment is the binary treatment covariate (0 for RT, 1 for

Table 4. Results for the follicular cell lymphoma dataset: parameter estimates (Est.), bootstrap standard errors (BSE) and *p*-values, using the Gumbel copula with Cox–Weibull and Cox-Lognormal margins.

	Cox–Weibull			Cox-Lognormal		
	Est.	BSE	*p*-value	Est.	BSE	*p*-value
Disease relapse						
Treatment	−0.350	0.165	0.034	−0.336	0.172	0.050
Age	0.398	0.081	0.000	0.431	0.075	0.000
Hgb	0.038	0.061	0.535	0.063	0.067	0.345
Clinstg	−0.660	0.150	0.000	−0.617	0.141	0.000
Death in remission						
Intercept	2.497	0.279	0.000	1.774	0.325	0.000
Treatment	0.185	0.179	0.304	0.304	0.205	0.138
Age	−0.654	0.071	0.000	−0.681	0.076	0.000
Hgb	−0.026	0.068	0.701	−0.039	0.084	0.639
Clinstg	0.532	0.178	0.003	0.682	0.180	0.000
σ	0.697	0.135	0.000	1.101	0.078	0.000
τ	0.579	0.165	0.000	0.767	0.159	0.000

CMT); Age is the age of the patient at diagnosis (in years); Hgb is the hemoglobin level in g/l; and Clinstg is the clinical stage of the disease at the time of diagnosis (1 for stage I, 0 for stage II). The data are available in the R package `randomForestSRC`. Our objective is to study the effect of treatment on relapse of the disease after adjusting for other covariate information. The knowledge of a specific group of patients with a higher risk of disease relapse would inform policymakers on the treatment strategies and type of follow-up foreseen.

We model the time of interest (time to disease relapse) jointly with the dependent censoring time (time to death in remission). We apply the proposed estimation procedure to fit models (2.1)–(2.3) to this dataset. The copula in (2.3) is assumed to be the Gumbel, Frank or Gaussian copula, where the two marginal regressions for the disease relapse and death in remission are specified by a Cox proportional hazards and a Weibull or log-normal model, respectively.

In Table 4 we display the parameter estimates, bootstrap standard errors and *p*-values based on the Gumbel copula. The results for the Frank and Gaussian copulas are similar, and hence are not reported. This is supported by our simulations in Section 6 (see Scenario 2). The *p*-values are calculated based on a Wald statistic using bootstrap standard errors, which are obtained from 200 resamples of the observed data. It can be seen that three covariates, namely treatment, age and clinical stage of the disease, have significant effects on the time to the disease relapse based on Cox and Weibull (Cox–Weibull) margins. The same set of covariates are found to be significant for the time to disease relapse when we model the margins using the Cox and lognormal (Cox-Lognormal) models, except that the treatment effect is now borderline significant at 5% level. It can be concluded that the CMT group has a lower risk of disease relapse compared to the RT group. In contrast, two covariates are significantly related to the time to death in remission based on the two models, namely age and clinical stage of the disease. Concerning the treatment effect, there is no difference between those who are treated with RT and those who are treated with CMT in reducing the risk of death in remission. As measured by Kendall's $\hat{\tau}$, the two endpoints, time to disease relapse and death, are significantly correlated even after adjusting for covariates.

Table 5. Parameter estimates (Est.), bootstrap standard errors (BSE) and *p*-values, using the independence copula with Cox–Weibull and Cox-Lognormal margins.

	Cox–Weibull			Cox-Lognormal		
	Est.	BSE	*p*-value	Est.	BSE	*p*-value
Disease relapse						
Treatment	−0.370	0.162	0.026	−0.370	0.162	0.026
Age	0.329	0.069	0.000	0.329	0.069	0.000
Hgb	0.041	0.064	0.518	0.041	0.064	0.518
Clinstg	−0.652	0.136	0.000	−0.652	0.136	0.000
Death in remission						
Intercept	3.208	0.145	0.000	3.155	0.193	0.000
Treatment	0.127	0.209	0.545	0.154	0.259	0.553
Age	−0.703	0.088	0.000	−0.762	0.097	0.000
Hgb	−0.029	0.091	0.749	−0.036	0.105	0.732
Clinstg	0.333	0.161	0.039	0.351	0.210	0.094
σ	0.608	0.051	0.000	1.125	0.091	0.000

We therefore present in Table 5 the parameter estimates under the independence copula model for purposes of comparison. The parameter estimates for the time to disease relapse are the same under the Cox–Weibull and Cox-Lognormal models, which is not surprising given that the Cox model is used in both models and that the two endpoints are independent under the independence copula model. Comparing results in Table 5 with those in Table 4, we see that the parameter estimates are somewhat different for the Gumbel copula and the independence copula models, which shows that the association between the two endpoints affects the parameter estimates. We observe that the same variables are significant and insignificant under the Cox–Weibull model. However, under the Cox-Lognormal model, the clinical stage of the disease is relevant for the time to death in remission when the copula is the Gumbel copula, but it ceases to be significant at the 5% level in the independence copula model. Moreover, the estimated cumulative hazard function under the independence copula deviates much from the one under the Gumbel copula, as shown in Figure 5 of the supplementary material.

Finally, the goodness-of-fit *p*-value proposed in the supplementary material is computed based on 200 bootstrap samples for the fitted models. From Table 8 of the supplementary material, we see that the *p*-values for the Gumbel copula with Cox–Weibull and Cox-Lognormal margins are larger those for the corresponding independence copula. Hence, there is a preference for the models taking the dependence between time to disease relapse and death into account. From the fitted semiparametric copula models, the model combination of the Gumbel copula with Cox-Lognormal margins has the largest *p*-value of 0.56 among all fitted models.

8. Discussion

This article proposed a copula-based semiparametric model and a corresponding estimation strategy for survival data with dependent censoring. We gave sufficient conditions under which the proposed model is identifiable and illustrated that these conditions are satisfied for many models. Most importantly, our method allows the estimation of the association parameter based on the observed survival data instead of assuming this dependence parameter to be known. We also developed a nonparametric estimator for the cumulative hazard function under dependent censoring.

The proposed method has been assessed both in an asymptotic way and via finite-sample simulations. We proved that the parameter estimators are consistent and asymptotically normal as the sample size tends to infinity. In addition, we showed that the estimator of the cumulative hazard function converges uniformly to the true hazard function. The numerical study also demonstrated the good performance of the proposed method, even when the copula structure or the margin for the censoring model are misspecified. Furthermore, the coverage rate for the finite-dimensional parameters is quite close to the 95% nominal level, suggesting that the asymptotic normality is approximately satisfied even in finite samples.

As a topic of future research, one could extend the current model to semiparametric or nonparametric regression functions, using splines, orthogonal series or kernel methods. Another potential topic for future research is the use of Cox models for the marginal distribution of both the survival time and the censoring time, whereas their dependence structure is modeled using a parametric copula. A crucial issue is then showing the identifiability of this flexible model.

Supplementary Materials

The supplementary material contains all the proofs in this article, estimation algorithm, goodness-of-fit test, additional simulation examples, and additional analysis results for the follicular cell lymphoma dataset.

Funding

Financial support from the European Research Council (2016–2022, Horizon 2020 ERC grant agreement no. 694409) is gratefully acknowledged.

ORCID

Negera Wakgari Deresa https://orcid.org/0000-0002-1302-3725
Ingrid Van Keilegom https://orcid.org/0000-0001-8827-7642

References

Breslow, N. (1974), "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, 89–99. [5]

Chen, X., Linton, O., and Van Keilegom, I. (2003), "Estimation of Semiparametric Models when the Criterion Function is not Smooth," *Econometrica*, 71, 1591–1608. [6]

Chen, Y.-H. (2010), "Semiparametric Marginal Regression Analysis for Dependent Competing Risks under an Assumed Copula," *Journal of the Royal Statistical Society*, Series B, 72, 235–251. [1,8]

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151. [3]

Collett, D. (2015), *Modelling Survival Data in Medical Research* (3rd ed.), New York: Taylor and Francis. [1]

Cox, D. R. (1972), Regression Models and Life Tables," *Journal of the Royal Statistical Society*, Series B, 34, 187–220. [1,2]

Crowder, M. (1991), "On the Identifiability Crisis in Competing Risks Analysis," *Scandinavian Journal of Statistics*, 18, 223–233. [1]

Czado, C., and Van Keilegom, I. (2023), "Dependent Censoring based on Copulas," *Biometrika* (in press, arXiv preprint, *https://arxiv.org/abs/2104.06872*). [2,3,4]

Deresa, N. W., and Van Keilegom, I. (2021), "On Semiparametric Modelling, Estimation and Inference for Survival Data Subject to Dependent Censoring," *Biometrika*, 108, 965–979. [1]

Deresa, N. W., Van Keilegom, I., and Antonio, K. (2022), "Copula-based Inference for Bivariate Survival Data with Left Truncation and Dependent Censoring," *Insurance: Mathematics and Economics*, 107, 1–21. [2,3,4]

Fleming, T. R., and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, Hoboken, NJ: Wiley. [4]

Frank, M. J. (1979), "On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$," *Aequationes Mathematicae*, 19, 194–226. [3]

Genest, C. (1987), "Frank's Family of Bivariate Distributions," *Biometrika*, 74, 549–555. [3]

Gumbel, E. J. (1960), "Distributions des valeurs extrêmes en plusieurs dimensions," *Publications de l'Institut de Statistique de l'Université de Paris*, 9, 171–173. [3]

Huang, X., and Zhang, N. (2008), "Regression Survival Analysis with an Assumed Copula for Dependent Censoring: A Sensitivity Analysis Approach," *Biometrics*, 64, 1090–1099. [1,8]

Jackson, D., White, I. R., Seaman, S., Evans, H., Baisley, K., and Carpenter, J. (2014), "Relaxing the Independent Censoring Assumption in the Cox Proportional Hazards Model Using Multiple Imputation," *Statistics in Medicine*, 33, 4681–4694. [1]

Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481. [1]

Nelsen, R. B. (2006), *An Introduction to Copulas* (2nd ed.), New York: Springer. [2,4]

Pintilie, M. (2006), *Competing Risks: A Practical Perspective*, Chichester: Wiley. [9]

Rivest, L.-P., and Wells, M. T. (2001), "A Martingale Approach to the Copula-Graphic Estimator for the Survival Function under Dependent Censoring," *Journal of Multivariate Analysis*, 79, 138–155. [1,4]

Robins, J. M., and Finkelstein, D. H. (2000), "Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests," *Biometrics*, 56, 779–788. [1]

Scharfstein, D. O., and Robins, J. M. (2002), "Estimation of the Failure Time Distribution in the Presence of Informative Censoring," *Biometrika*, 89, 617–634. [1]

Sklar, M. (1959), "Fonctions de répartition à *n* dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231. [2]

Tsiatis, A. (1975), "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20–22. [1]

Zheng, M., and Klein, J. P. (1995), "Estimates of Marginal Survival for Dependent Competing Risks based on an Assumed Copula," *Biometrika*, 82, 127–138. [1]

Zucker, D. M. (2005), "A Pseudo-Partial Likelihood Method for Semiparametric Survival Regression with Covariate Errors," *Journal of the American Statistical Association*, 100, 1264–1277. [5,6]