

# Two-stage pseudo maximum likelihood estimation of semiparametric copula-based regression models for semi-competing risks data

Sakie J. Arachchige

Department of Mathematics and Statistics  
Mississippi State University, MS, USA

Xinyuan Chen

Department of Mathematics and Statistics  
Mississippi State University, MS, USA

Qian M. Zhou

Department of Mathematics and Statistics  
Mississippi State University, MS, USA, qz70@msstate.edu

## Abstract

We propose a two-stage estimation procedure for a copula-based model with semi-competing risks data, where the non-terminal event is subject to dependent censoring by the terminal event, and both events are subject to independent censoring. With a copula-based model, the marginal survival functions of individual event times are specified by semiparametric transformation models, and the dependence between the bivariate event times is specified by a parametric copula function. For the estimation procedure, in the first stage, the parameters associated with the marginal of the terminal event are estimated using only the corresponding observed outcomes, and in the second stage, the marginal parameters for the non-terminal event time and the copula parameter are estimated together via maximizing a pseudo-likelihood function based on the joint distribution of the bivariate event times. We derived the asymptotic properties of the proposed estimator and provided an analytic variance estimator for inference. Through simulation studies, we showed that our approach leads to consistent estimates with less computational cost and more robustness than the one-stage procedure developed in Chen (2012), where all parameters were estimated simultaneously. In addition, our approach demonstrates more desirable finite-sample performances over another existing two-stage estimation method proposed in Zhu et al. (2021). An **R** package **PMLE4SCR** is developed to implement our proposed method.

**Key words:** Copulas Pseudo maximum likelihood estimator Semi-competing risks Semiparametric regression Stage-wise estimation

## 1 Introduction

Medical studies oftentimes involve situations where multiple time-to-event outcomes exist simultaneously. These events can be divided into two categories: *non-terminal*, e.g., cancer relapse, and *terminal*, e.g., death. The occurrence of the terminal event precludes that of the non-terminal, but not vice versa; the resulting data are commonly referred to as the *semi-competing risks* (Fine et al., 2001). Let  $T$  and  $D$  denote times to a non-terminal and terminal event, respectively, and they are assumed to be subject to independent administrative censoring. The distribution of  $T$  and the dependence between  $T$  and  $D$  are often of interest.

Copula-based models have been used to model the between-event dependency for semi-competing risks data that arise from studies of aging (Varadhan et al., 2014; Sun et al., 2023, 2024) and disease

progression among cancer patients (Zhou et al., 2016). In such models, the joint survival function of  $(T, D)$ ,  $S(t, d) \equiv \mathbb{P}(T > t, D > d)$ , is directly specified as  $S(t, d) = \mathbb{C}(S_T(t), S_D(d))$ . Here,  $\mathbb{C}(\cdot, \cdot) : [0, 1]^2 \rightarrow [0, 1]$  is a bivariate copula function (Sklar, 1959), and  $S_T(t) \equiv \mathbb{P}(T > t)$  and  $S_D(d) \equiv \mathbb{P}(D > d)$  are marginal survival functions of  $T$  and  $D$ , respectively. This copula-based approach enjoys the flexibility from the separate modeling of (i) the marginals and (ii) the dependence structure between  $T$  and  $D$ . Shih and Louis (1995) proposed a two-stage estimation procedure for multivariate survival outcome under independent censoring, where, in the first stage, the marginals were estimated separately for each event without accounting for the between-event dependencies, and, in the second stage, the parameter associated with the copula was estimated via maximizing the log-likelihood with the plug-in estimates of the marginals. The resulting estimator for the copula parameter is referred to as the pseudo maximum likelihood estimator (PMLE).

Accurately estimating the copula parameter relies on the consistent estimation of the marginals. The validity of the above stage-wise estimation procedure requires that individual event times are only subject to independent censoring. Thus, such an approach has been mostly used on clustered survival data. However, this assumption does not hold for semi-competing risks data because  $T$  is dependently censored by  $D$ . In other words, standard tools such as the Kaplan-Meier (KM) estimator are not consistent for  $S_T(\cdot)$ , which poses challenges in estimating the copula parameter via the above two-stage procedure. As a result, earlier estimators were obtained from estimating equations based on a concordance indicator (Fine et al., 2001), the Doob–Meyer (DM) decomposition of a counting process (Wang, 2003), or conditional Kendall’s  $\tau$  (Lakhal et al., 2008). Later, these procedures were extended to incorporate a discrete covariate for the marginal distribution via a regression model and for the between-event dependence parameter (Ghosh, 2006; Hsieh et al., 2008). For estimating the marginal of  $T$ , they predominantly considered a *plug-in* estimator, first proposed in Fine et al. (2001), through the relationship of  $S_T(\cdot)$  with  $S_D(\cdot)$  and the survival function  $S_{T^*}(\cdot)$  of  $T^* \equiv T \wedge D$ , the time to the first event (non-terminal or terminal). However, this method is limited to the Archimedean copula family, which provides an explicit expression for such a relationship. Different from the above methods, Peng and Fine (2007) considered a class of time-varying coefficient semiparametric regression models for the marginals and copulas with time-varying parameters. They simultaneously estimated the copula parameter and the marginal of  $T$  using non-linear estimating equations.

Chen (2012) considered the maximum likelihood estimation (MLE) approach, i.e., simultaneous estimation of all parameters associated with the marginals and the copula via maximizing the log-likelihood function. Zhou et al. (2016) proposed a PMLE for the copula parameter based on the log-likelihood function expressed in terms of  $S_D(\cdot)$  and  $S_{T^*}(\cdot)$  without covariates. Zhu et al. (2021) extended the procedure in Zhou et al. (2016) to incorporate covariates into the marginals specified by time-varying coefficient semiparametric regression models. They proposed a two-stage estimation procedure: in the first stage, the PMLE of the copula parameter was obtained following the idea in Zhou et al. (2016), and, in the second stage, the regression coefficients associated with  $S_T(\cdot)$  were estimated from non-linear estimating equations developed in Peng and Fine (2007).

In this article, we propose a new estimation procedure for copula-based models to analyze semi-competing risks data, with the primary goals of investigating (i) the association between  $T$  and  $D$  and (ii) the direct covariate effects on their marginals. Along the line of Chen (2012), we consider a class of models where a parametric copula is employed to model the dependence between  $T$  and  $D$ , and the marginals are specified by semi-parametric transformation models. We estimate the marginals and copula parameters via a two-stage PMLE approach. Specifically, in the first stage, we obtain a consistent estimate of the marginal of  $D$  since it is only subject to independent censoring, and in the second stage, we estimate the marginal of  $T$  and the copula parameter simultaneously by maximizing the pseudo-likelihood function with the plug-in estimates of  $S_D(\cdot)$ .

Our two-stage estimation approach adapts from the one-stage MLE method of Chen (2012), which is expected to be asymptotically more efficient but likely computationally challenging in practice. In contrast, our PMLE approach reduces the dimension of the parameter space in each stage while yielding consistent estimates. Despite not utilizing the association between  $T$  and  $D$  in the first stage, which might result in efficiency loss, we observed through simulation studies that, compared to the MLE, the PMLE approach gives more desirable finite-sample performances, including smaller biases for copula parameters. In addition, our approach enjoys robustness against copula misspecification when estimating

the marginals. When the copula function is misspecified, our estimation of  $S_D(\cdot)$  in the first stage is guaranteed to be consistent. Furthermore, although the PMLE of  $S_T(\cdot)$  is not consistent under copula misspecification, with a consistent estimate of  $S_D(\cdot)$ , it leads to a smaller bias compared to the one-stage MLE, which is additionally affected by the biased estimates of  $S_D(\cdot)$ .

Our method exhibits three advantages compared to the stage-wise estimation procedure developed by Zhu et al. (2021). First, our method does not involve the first-event time  $T^*$ , and consequently, besides Archimedean copulas, it can also be applied to the elliptical copula family, e.g., Gaussian or  $t$  copulas. Second, given the same estimator of the marginal of  $D$  from both methods, our estimators for the marginal of  $T$  and the copula parameter can be regarded as the MLE, but the estimator in Zhu et al. (2021) is still the PMLE based on plug-in estimates. Consequently, our estimators are inherently more efficient. Third, we provide variance estimators based on analytic expressions for inference. Compared to the non-parametric bootstrap employed in Zhu et al. (2021) and Zhou et al. (2016), our inference procedure is computationally more efficient. To implement our method, an **R** package **PMLE4SCR** is available in the GitHub repository <https://github.com/michellezhou2009/PMLE4SCR>.

The remainder of the article is organized as follows. In Section 2, we propose the two-stage PMLE estimation procedure and discuss its asymptotic properties. Section 3 presents simulation studies evaluating the finite-sample performance of the PMLE and the analysis of Bone Marrow Transplant (BMT) data as an illustration. We provide concluding remarks in Section 4. Technical details and additional numerical results are given in the Online Supplement of the Supporting Information.

## 2 Method

### 2.1 Copula-based model specification

Let  $A$  denote the time to independent administrative censoring. The observed event times are denoted by  $C \equiv D \wedge A$  and  $X \equiv T \wedge C$ , with respective censoring status defined as  $\delta_D = \mathbb{1}(D \leq A)$  and  $\delta_T = \mathbb{1}(T \leq C)$ . In addition, let  $\mathbf{Z} \in \mathbb{R}^p$  denote a vector of baseline covariates, and given  $\mathbf{Z}$ ,  $A$  is independent of both  $T$  and  $D$ , but  $T$  and  $D$  are associated. The observed data from  $n$  independent subjects is denoted by  $\mathfrak{S} = \{(X_i, C_i, \delta_{T,i}, \delta_{D,i}, \mathbf{Z}_i), i = 1, \dots, n\}$ . We have the following commonly adopted assumptions regarding the study setting and covariates:

- (A1) There exists a maximum follow-up time  $\xi < \infty$  in the study such that  $\mathbb{P}(X > \xi | \mathbf{Z}) > 0$ ,  $\mathbb{P}(\delta_T = 0, \delta_D = 0, X = \xi | \mathbf{Z}) > 0$ ,  $\mathbb{P}(C > \xi | \mathbf{Z}) > 0$ , and  $\mathbb{P}(\delta_T = 1, \delta_D = 0, C = \xi | \mathbf{Z}) > 0$ , with probability 1;
- (A2) The covariates  $\mathbf{Z}$  is bounded with probability 1.

Assumptions (A1) - (A2) are regulatory conditions which have seen frequent adoption in the semi-competing risks literature, see Fine et al. (2001, §1.2), Zeng and Lin (2006, Conditions 2 - 3), Chen (2010, Assumptions (b) - (c)), and Chen (2012, Assumptions (II) - (III)). These two conditions are required to obtain the uniform convergence of the estimators for  $S_T(t | \mathbf{Z})$  and  $S_D(d | \mathbf{Z})$  on  $[0, \xi]$ , which will be given in the later sections. Assumption (A1) is plausible in practice because, in general, most studies terminate after a finite observation time due to various constraints, i.e., time and/or budget, and subjects who have not experienced any events throughout the entire follow-up are regarded as administratively censored. For example, in the critical care trial of acute lung injury patients (Acute Respiratory Distress Syndrome Network, 2000) studied in Chen et al. (2024), the administrative censoring (maximum follow-up) time is 180 days from when a patient is admitted to the intensive care unit.

Given covariates  $\mathbf{Z}$ , the copula-based approach models the joint survival function of  $(T, D)$ ,  $\mathbb{P}(T > t, D > d | \mathbf{Z})$ , as a function of their respective marginals  $S_T(t | \mathbf{Z}) = \mathbb{P}(T > t | \mathbf{Z})$  and  $S_D(d | \mathbf{Z}) = \mathbb{P}(D > d | \mathbf{Z})$ , i.e.,

$$\mathbb{P}(T > t, D > d | \mathbf{Z}) = \mathbb{C}(S_T(t | \mathbf{Z}), S_D(d | \mathbf{Z}); \alpha), \quad (1)$$

where  $\mathbb{C}(\cdot, \cdot; \alpha) : [0, 1]^2 \rightarrow [0, 1]$  is a copula function parametrically specified with parameter  $\alpha$ . In the remainder of this article, we define  $\dot{\mathbb{C}}_{u_j} \equiv \partial \mathbb{C}(u_T, u_D; \alpha) / \partial u_j$ ,  $j = T, D$ , as the first-order partial derivatives of the copula function, and  $\ddot{\mathbb{C}}_{u_T u_D} \equiv \partial^2 \mathbb{C}(u_T, u_D; \alpha) / \partial u_T \partial u_D$  as the second-order partial derivative.

**Remark 1** The Archimedean and Elliptical copula families are often selected for specifying  $\mathbb{C}(\cdot, \cdot; \alpha)$ . Archimedean copulas are defined via a generator function  $\zeta(u; \alpha)$  for  $u \in [0, 1]$ , which is continuous, convex, and strictly decreasing with  $\zeta(1; \alpha) = 0$ . A bivariate Archimedean copula can be constructed as  $\mathbb{C}(u_1, u_2; \alpha) = \zeta^{-1}\{\zeta(u_1; \alpha) + \zeta(u_2; \alpha); \alpha\}$ . Below, we present some commonly used generator functions and corresponding bivariate copulas and refer readers to Nelsen (2006) for a more detailed introduction.

(i) Clayton copula,  $\zeta(u; \alpha) = (u^{-\alpha} - 1)/\alpha$  for  $\alpha \in (-1, \infty) \setminus \{0\}$ , and

$$\mathbb{C}(u_1, u_2; \alpha) = \{\max(u_1^{-\alpha} + u_2^{-\alpha} - 1, 0)\}^{-1/\alpha}; \quad (2)$$

(ii) Frank copula,  $\zeta(u; \alpha) = -\log\{(e^{-\alpha u} - 1)/(e^{-\alpha} - 1)\}$  for  $\alpha \in \mathbb{R} \setminus \{0\}$ , and

$$\mathbb{C}(u_1, u_2; \alpha) = -\alpha^{-1} \log\left\{1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1}\right\}; \quad (3)$$

(iii) Gumbel copula,  $\zeta(u; \alpha) = \{-\log(u)\}^\alpha$  for  $\alpha \in [1, \infty)$ , and

$$\mathbb{C}(u_1, u_2; \alpha) = \exp\left(-[\{-\log(u_1)\}^\alpha + \{-\log(u_2)\}^\alpha]^{1/\alpha}\right). \quad (4)$$

Elliptical copulas join marginals via an elliptical distribution, e.g., a Gaussian distribution. A bivariate Gaussian copula with parameter  $\alpha$  is defined as

$$\mathbb{C}(u_1, u_2; \alpha) = \Phi_2\{\Phi^{-1}(u_1), \Phi^{-1}(u_2); \alpha\}, \quad (5)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the standard univariate Gaussian distribution, and  $\Phi_2(\cdot, \cdot; \alpha)$  is the standard bivariate Gaussian cumulative distribution function with  $\alpha$  being the correlation parameter.

We specify the marginal survival functions of  $T$  and  $D$  via semi-parametric transformation models. Specifically, for  $j = T, D$ ,

$$S_j(t|\mathbf{Z}) = \exp\left[-G_j\left\{R_j(t)e^{\beta_j^T \mathbf{Z}}\right\}\right], \quad (6)$$

where  $\beta_j$  is a  $p$ -dimensional vector of regression coefficients, the unspecified baseline function  $R_j(\cdot)$  is assumed to be increasing, and  $G_j(\cdot)$  is a specified non-negative, strictly increasing, and continuously differentiable function. For example, if  $G_j(t) = t$ , (6) becomes a Cox proportional hazards (PH) model; if  $G_j(t) = \log(1+t)$ , (6) becomes a proportional odds model. The marginal survival function in (6) can be easily extended to accommodate time-varying covariates  $\mathbf{Z}(t)$ , where  $S_j(t|\mathbf{Z}) = \exp[-G_j\{R_j(t)e^{\beta_j^T \mathbf{Z}(t)}\}]$ .

## 2.2 The two-stage estimation procedure

Let  $\boldsymbol{\theta}_D = \{\beta_D, R_D(\cdot)\}$  and  $\boldsymbol{\theta}_T = \{\beta_T, R_T(\cdot)\}$  denote the parameters associated with the marginals of  $D$  and  $T$ , respectively. We propose a two-stage estimation procedure for  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_D, \boldsymbol{\theta}_T, \alpha\}$ , where, in the first stage,  $\boldsymbol{\theta}_D$  is estimated by itself via an MLE approach with observed data  $\mathfrak{S}_D = \{(C_i, \delta_{D,i}, \mathbf{Z}_i), i = 1, \dots, n\}$ , and, in the second stage,  $\boldsymbol{\theta}_1 \equiv \{\boldsymbol{\theta}_T, \alpha\}$  is estimated via maximizing a pseudo-log-likelihood depending on the estimate  $\hat{\boldsymbol{\theta}}_D$  obtained from the first stage. We hereafter refer to the proposed estimation procedure as the two-stage PMLE for brevity.

### 2.2.1 The first stage estimation of $\boldsymbol{\theta}_D$

In the semi-competing risks setting,  $\boldsymbol{\theta}_D$  can be consistently estimated by an MLE with observed data  $\mathfrak{S}_D$ , since  $D$  is only subject to independent censoring by  $A$  (Zeng and Lin, 2006). The baseline function  $R_D(\cdot)$  is regarded as a non-decreasing step function that jumps only at the observed event times. Let  $0 < d_1 < d_2 < \dots < d_{\kappa_D} \leq \xi$  denote the ordered observed death times, and let  $dR_{D,l} > 0$  denote the jump size of  $R_D(\cdot)$  at time  $d_l$ , for  $l = 1, \dots, \kappa_D$ . Thus,  $\boldsymbol{\theta}_D$  can be re-expressed as  $\boldsymbol{\theta}_D = (\beta_D^T, dR_{D,1}, \dots, dR_{D,\kappa_D})^T$ .

The log-likelihood function of  $\boldsymbol{\theta}_D$  given  $\mathfrak{S}_D$  is  $\ell_D(\boldsymbol{\theta}_D) = n^{-1} \sum_{i=1}^n \ell_{D,i}(\boldsymbol{\theta}_D)$ , where

$$\begin{aligned}\ell_{D,i}(\boldsymbol{\theta}_D) &= \int_0^{\xi} \left[ \log dR_D(t) + \boldsymbol{\beta}_D^\top \mathbf{Z}_i + \log \dot{G}_D \left\{ R_D(t) e^{\boldsymbol{\beta}_D^\top \mathbf{Z}_i} \right\} \right] dN_{D,i}(t) \\ &\quad - G_D \left\{ R_D(C_i) e^{\boldsymbol{\beta}_D^\top \mathbf{Z}_i} \right\},\end{aligned}\tag{7}$$

with  $N_{D,i}(t) = \mathbb{1}(C_i \leq t) \delta_{D,i}$ . The MLE  $\hat{\boldsymbol{\theta}}_D = (\hat{\boldsymbol{\beta}}_D^\top, \hat{dR}_{D,1}, \dots, \hat{dR}_{D,\kappa_D})^\top$  is the solution to score equations  $\Psi_D(\boldsymbol{\theta}_D) = n^{-1} \sum_{i=1}^n \Psi_{D,i}(\boldsymbol{\theta}_D) = 0$  with  $\Psi_{D,i}(\boldsymbol{\theta}_D) = \partial \ell_{D,i}(\boldsymbol{\theta}_D) / \partial \boldsymbol{\theta}_D$ .

### 2.2.2 The second stage estimation of $\boldsymbol{\theta}_1$

Different from  $\boldsymbol{\theta}_D$ ,  $\boldsymbol{\theta}_T$  cannot be consistently estimated using only the observed data  $\{(X_i, \delta_{T,i}, \mathbf{Z}_i), i = 1, \dots, n\}$ , since  $T$  is subject to both dependent and independent censoring by  $D$  and  $A$ , respectively. A consistent estimator must be obtained from a full likelihood function based on the joint survival function of  $(T, D)$  with  $\mathfrak{S}$  (Fine et al., 2001). Similar to  $R_D(\cdot)$ ,  $R_T(\cdot)$  is also regarded as a non-decreasing step function that jumps only at the observed non-terminal event times  $0 < t_1 < t_2 < \dots < t_{\kappa_T} \leq \xi$ , where  $dR_{T,l} > 0$  denotes the jump size of  $R_T(\cdot)$  at time  $t_l$ , for  $l = 1, \dots, \kappa_T$ . The parameter  $\boldsymbol{\theta}_T$  can be re-expressed as  $\boldsymbol{\theta}_T = (\boldsymbol{\beta}_T^\top, dR_{T,1}, \dots, dR_{T,\kappa_T})^\top$ . The full log-likelihood function given  $\mathfrak{S}$  is  $\ell(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$ , where

$$\begin{aligned}\ell_i(\boldsymbol{\theta}) &= \delta_{T,i} \delta_{D,i} \log \ddot{\mathbb{C}}_{u_T u_D} + \delta_{T,i} (1 - \delta_{D,i}) \log \dot{\mathbb{C}}_{u_T} \\ &\quad + (1 - \delta_{T,i}) \delta_{D,i} \log \dot{\mathbb{C}}_{u_D} + (1 - \delta_{T,i})(1 - \delta_{D,i}) \log \mathbb{C} \\ &\quad + \delta_{T,i} \left[ -G_T \left\{ R_T(X_i) e^{\boldsymbol{\beta}_T^\top \mathbf{Z}_i} \right\} + \log \dot{G}_T \left\{ R_T(X_i) e^{\boldsymbol{\beta}_T^\top \mathbf{Z}_i} \right\} \right. \\ &\quad \left. + \log dR_T(X_i) + \boldsymbol{\beta}_T^\top \mathbf{Z}_i \right] \\ &\quad + \delta_{D,i} \left[ -G_D \left\{ R_D(C_i) e^{\boldsymbol{\beta}_D^\top \mathbf{Z}_i} \right\} + \log \dot{G}_D \left\{ R_D(C_i) e^{\boldsymbol{\beta}_D^\top \mathbf{Z}_i} \right\} \right. \\ &\quad \left. + \log dR_D(C_i) + \boldsymbol{\beta}_D^\top \mathbf{Z}_i \right].\end{aligned}\tag{8}$$

The PMLE of  $\boldsymbol{\theta}_1$  is obtained by maximizing the following pseudo-log-likelihood function:

$$\hat{\ell}(\boldsymbol{\theta}_1) \equiv \ell(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_D) = n^{-1} \sum_{i=1}^n \ell_i(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_D),$$

where  $\ell_i$  is given in (8). The PMLE  $\hat{\boldsymbol{\theta}}_1 = (\hat{\alpha}, \hat{\boldsymbol{\beta}}_T^\top, \hat{dR}_{T,1}, \dots, \hat{dR}_{T,\kappa_T})^\top$  is the solution to pseudo score equations  $\hat{\Psi}_1(\boldsymbol{\theta}_1) = n^{-1} \sum_{i=1}^n \hat{\Psi}_{1,i}(\boldsymbol{\theta}_1) = 0$  with  $\hat{\Psi}_{1,i}(\boldsymbol{\theta}_1) = \hat{\ell}_{\boldsymbol{\theta}_1,i}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_D) = \partial \ell_i(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_D) / \partial \boldsymbol{\theta}_1$ . The score function  $\ell_{\boldsymbol{\theta}_1,i}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_D) = (\Psi_{\alpha,i}, \Psi_{\beta_T,i}^\top, \Psi_{dR_{T,1},i}, \dots, \Psi_{dR_{T,\kappa_T},i})^\top$ , with the expressions given in Section S1 of the Online Supplement, is a function of  $\boldsymbol{\theta}_D$  only through  $U_{D,i}$ .

**Remark 2** The copula in (1) can be extended by allowing the copula parameter to depend on some baseline covariates. In the data analysis of Chen (2012), the copula parameter varies across different levels of a categorical variable. Nikoloulopoulos and Karlis (2008) regressed the copula parameter on covariates for nonsurvival outcomes. Following these works, we can assume  $\alpha = \phi(\boldsymbol{\gamma}^\top \mathbf{W})$ , where  $\mathbf{W}$  is a subset of the covariates  $\mathbf{Z}$ , and  $\phi(\cdot)$  is a specified link function. The copula parameter thus becomes  $\boldsymbol{\gamma}$ , and its corresponding score function is modified as  $\Psi_{\boldsymbol{\gamma},i} = \Psi_{\alpha,i} \phi'(\boldsymbol{\gamma}^\top \mathbf{W}_i) \mathbf{W}_i$ , where  $f'(\cdot)$  denotes the first-order derivative of a univariate function  $f(\cdot)$ .

**Remark 3** The log-likelihood function in (8) is equivalent to the log-likelihood in Chen (2012) derived based on three observed counting processes: the first event being non-terminal, the first event being terminal, and the second event occurring subsequently to the uncensored first event. Therefore, the one-stage MLE obtained via maximizing (8) shares the same asymptotic properties given in Theorems 1 and 2 of Chen (2012).

### 2.3 Asymptotic properties of the two-stage PMLE

We present the asymptotic properties of the two-stage PMLE. Define the following quantities:

$$\Upsilon_1(g_T, g_D) = -\log \mathbb{C}\left(e^{-G_T(g_T)}, e^{-G_D(g_D)}; \alpha\right) \quad \text{and} \quad \eta_{1j}(g_T, g_D) = \frac{\partial}{\partial g_j} \Upsilon_1(g_T, g_D),$$

for  $j = T, D$ ; also

$$\Upsilon_2(g_T, g_D) = -\log \dot{\mathbb{C}}_{u_T}\left(e^{-G_T(g_T)}, e^{-G_D(g_D)}; \alpha\right) \quad \text{and} \quad \eta_2(g_T, g_D) = \frac{\partial}{\partial g_D} \Upsilon_2(g_T, g_D).$$

Let  $\boldsymbol{\theta}^0 = \{\alpha^0, \boldsymbol{\beta}_T^0, R_T^0(\cdot), \boldsymbol{\beta}_D^0, R_D^0(\cdot)\}$  be the true values of  $\boldsymbol{\theta}$ , and  $\Theta$  denote the total parameter space. The following conditions are required to establish the asymptotic properties of  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_D^\top, \hat{\boldsymbol{\theta}}_1^\top)^\top$ .

- (C1) The true baseline functions  $R_T^0(\cdot)$  and  $R_D^0(\cdot)$  are strictly increasing and continuously differentiable, and the true values  $\alpha^0, \boldsymbol{\beta}_T^0$  and  $\boldsymbol{\beta}_D^0$ , fall in the interior of a compact parameter space.
- (C2) For any sequence  $0 < h_1 < \dots < h_r \leq g$ ,

$$\prod_{s=1}^r \{(1+h_s)\dot{G}_D(h_s)\} \exp\{-G_D(g)\} \leq \mu_0^r (1+g)^{-\nu_0},$$

where  $\mu_0$  and  $\nu_0$  are positive constants.

- (C3) The function  $G_j(\cdot)$  satisfies the conditions:  $G_j(0) = 0$ ,  $G_j(\infty) = \infty$ , and  $\dot{G}_j(0) > 0$  for  $j = T, D$ . The copula function  $\mathbb{C}(\cdot, \cdot; \alpha)$  satisfies the conditions:  $\mathbb{C}(0, 0; \alpha) = 0$  and  $\eta_{1j}(0, 0) > 0$  for  $j = T, D$ . Also, for positive constants  $c_1, c_2$ , and  $c_3$ ,

$$\begin{aligned} \lim_{g_1, g_2 \rightarrow \infty} \frac{\sum_{j=1}^2 \log\{g_j \times \sup_{s_1 \leq g_1, s_2 \leq g_2} \eta_{1j}(s_1, s_2)\}}{\Upsilon_1(c_1 g_1, c_2 g_2)} &= 0, \\ \lim_{g_1, g_2 \rightarrow \infty} \frac{\log\{g_2 \times \sup_{s_1 \leq g_1, s_2 \leq g_2} \eta_2(s_1, s_2)\}}{\inf_{w_1 \geq 0, w_2 \geq 0} \{\Upsilon_2(c_1 w_1, c_3 g_2) - \Upsilon_2(c_1 w_1, c_2 w_2)\}} &= 0. \end{aligned}$$

- (C4)  $\sum_{i=1}^n \ell_i(\boldsymbol{\theta})$ ,  $\sum_{i=1}^n \dot{\ell}_{\boldsymbol{\theta}_1, i}(\boldsymbol{\theta})$ ,  $\sum_{i=1}^n \ddot{\ell}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1, i}(\boldsymbol{\theta})$ , and  $\sum_{i=1}^n \ddot{\ell}_{\boldsymbol{\theta}_1 u_D, i}(\boldsymbol{\theta})$  are continuous and bounded on  $[0, \xi] \times \Theta$ , where  $\dot{\ell}_{\boldsymbol{\theta}_1, i}(\boldsymbol{\theta}) = \partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1$ ,  $\ddot{\ell}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1, i}(\boldsymbol{\theta}) = \partial^2 \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top$ , and  $\ddot{\ell}_{\boldsymbol{\theta}_1 u_D, i}(\boldsymbol{\theta}) = \partial^2 \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1 \partial u_D$ .
- (C5) Information matrices  $\mathcal{I}_D(\boldsymbol{\theta}_D) = \mathbb{E}\{\mathcal{I}_{D,i}(\boldsymbol{\theta}_D)\}$  with  $\mathcal{I}_{D,i}(\boldsymbol{\theta}_D) = -\partial^2 \ell_{D,i}(\boldsymbol{\theta}_D) / \partial \boldsymbol{\theta}_D \partial \boldsymbol{\theta}_D^\top$  and  $\mathcal{I}_1(\boldsymbol{\theta}) = \mathbb{E}\{\mathcal{I}_{1,i}(\boldsymbol{\theta})\}$  with  $\mathcal{I}_{1,i}(\boldsymbol{\theta}) = -\partial^2 \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top$  evaluated at the true value of  $\boldsymbol{\theta}_D^0$  and  $\boldsymbol{\theta}^0$  are positive definite, and their eigenvalues are bounded below and above by some finite positive constants. The explicit expressions of  $\mathcal{I}_D(\boldsymbol{\theta}_D)$  and  $\mathcal{I}_1(\boldsymbol{\theta})$  are given in Section S2 of the Online Supplement.
- (C6) The variance-covariance matrix  $\text{Var}\{\dot{\ell}_{\boldsymbol{\theta}_1, i}(\boldsymbol{\theta}) + n^{-1} \sum_{k=1}^n \ddot{\ell}_{\boldsymbol{\theta}_1 u_D, k}(\boldsymbol{\theta}) \psi_{u_D, k, i}(\boldsymbol{\theta}_D)\}$  evaluated at  $\boldsymbol{\theta}^0$  is positive definite, and its eigenvalues are bounded below and above by some finite positive constants, where  $\psi_{u_D, k, i}(\boldsymbol{\theta}_D)$  is given in Section S3 of the Online Supplement.

Assumptions (A1) - (A2) and Conditions (C1) - (C2) are from the regulatory conditions listed in Zeng and Lin (2006) for establishing the consistency and asymptotic distribution of  $\hat{\boldsymbol{\theta}}_D$ . Specifically, with probability 1,  $\hat{\boldsymbol{\beta}}_D$  converges to  $\boldsymbol{\beta}_D^0$  and  $\hat{R}_D(\cdot) \equiv \{\hat{d}\hat{R}_{D,l}, l = 1, \dots, \kappa_D\}$  converges to  $R_D^0(\cdot)$  uniformly in the interval  $[0, \xi]$ . In addition,  $\sqrt{n}(\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D)$  converges weakly to a zero-mean Gaussian process. Let  $\mathcal{I}_D^0 = \mathbb{E}\{\mathcal{I}_{D,i}(\boldsymbol{\theta}_D^0)\}$ . For a Hadamard differentiable functional  $F_D(\boldsymbol{\theta}_D)$  of  $\boldsymbol{\theta}_D$  (Chen, 2012), the functional delta method suggests that  $\sqrt{n}\{F_D(\hat{\boldsymbol{\theta}}_D) - F_D(\boldsymbol{\theta}_D)\}$  converges weakly to a zero-mean Gaussian process with the variance-covariance matrix  $\dot{F}_D(\boldsymbol{\theta}_D)^\top (\mathcal{I}_D^0)^{-1} \dot{F}_D(\boldsymbol{\theta}_D)$ , where  $\dot{F}_D(\boldsymbol{\theta}_D)$  is the derivative of  $F_D(\boldsymbol{\theta}_D)$  with respect to  $\boldsymbol{\theta}_D$ . The matrix  $(\mathcal{I}_D^0)^{-1}$  is estimated by  $\bar{\mathcal{I}}_{D,n}^{-1}(\hat{\boldsymbol{\theta}}_D)$  where  $\bar{\mathcal{I}}_{D,n}(\hat{\boldsymbol{\theta}}_D) = n^{-1} \sum_{i=1}^n \mathcal{I}_{D,i}(\hat{\boldsymbol{\theta}}_D)$ . Alternatively, one can estimate  $(\mathcal{I}_D^0)^{-1}$  via a robust estimator, given by  $\hat{\Sigma}_{D,n} \equiv n^{-1} \sum_{i=1}^n \psi_{\boldsymbol{\theta}_D, i}(\hat{\boldsymbol{\theta}}_D) \psi_{\boldsymbol{\theta}_D, i}(\hat{\boldsymbol{\theta}}_D)^\top$  with  $\psi_{\boldsymbol{\theta}_D, i}(\hat{\boldsymbol{\theta}}_D) = \bar{\mathcal{I}}_{D,n}^{-1}(\hat{\boldsymbol{\theta}}_D) \Psi_{D,i}(\hat{\boldsymbol{\theta}}_D)$ , where  $\Psi_{D,i}(\boldsymbol{\theta}_D)$  is the score function for the log-likelihood function  $\ell_{D,i}$  discussed in Section 2.2.1.

Condition (C3) was adopted in Chen (2012), which regulates the tail behavior of the estimated baseline functions, and together with Assumption (A1), the boundedness of  $\hat{R}_D(\cdot)$  and  $\hat{R}_T(\cdot)$  is guaranteed. Conditions (C4) - (C5) are standard conditions guaranteeing the existence of the score functions, information matrices, and certain components for the variance-covariance matrix of  $\hat{\theta}_1$ . Condition (C6) assumes the variance-covariance matrix of  $\hat{\theta}_1$  is well-defined. Additionally, since, for the Gumbel copula,  $\eta_{1j}(0,0) > 0$  (in Condition (C3)) does not hold, we adopt the same modification as Chen (2012) by adding a small positive value, e.g.,  $n^{-1}$ , to the cumulative hazard function,  $G_j\{R_j(t)e^{\beta_j^\top \mathbf{z}}\}$ , at time 0 to make  $\eta_{1j}(0,0)$  positive for  $j = T, D$ .

The consistency and asymptotic distribution of  $\hat{\theta}_1$  are established in the following Theorems 1 and 2, respectively.

**Theorem 1** *Assume Assumptions (A1) - (A2) and Conditions (C1) - (C6) hold. With probability 1,  $\hat{\alpha}$  converges to  $\alpha^0$ ,  $\hat{\beta}_T$  converges to  $\beta_T^0$ , and  $\hat{R}_T(\cdot) \equiv \{\hat{dR}_{T,l}, l = 1, \dots, \kappa_T\}$  converges to  $R_T^0(\cdot)$  uniformly in the interval  $[0, \xi]$ .*

**Theorem 2** *Assume Assumptions (A1) - (A2) and Conditions (C1) - (C6) hold, and then  $\sqrt{n}(\hat{\theta}_1 - \theta_1^0)$  converges weakly to a zero-mean Gaussian process. Let  $\mathcal{I}_1^0 = \mathbb{E}\{\mathcal{I}_{1,i}(\theta^0)\}$ . For a Hadamard differentiable functional  $F_1(\theta_1)$  of  $\theta_1$ ,  $\sqrt{n}\{F_1(\hat{\theta}_1) - F_1(\theta_1^0)\}$  converges weakly to a zero-mean Gaussian process with the variance-covariance matrix  $\dot{F}_1(\theta_1^0)^\top \Sigma_1^0 \dot{F}_1(\theta_1^0)$ , where  $\dot{F}_1(\theta_1)$  is the derivative of  $F_1(\theta_1)$  with respect to  $\theta_1$ , and  $\Sigma_1^0 \equiv \text{Var}\{\psi_{\theta_1,i}(\theta^0)\}$  with*

$$\psi_{\theta_1,i}(\theta^0) = (\mathcal{I}_1^0)^{-1} \left\{ \dot{\ell}_{\theta_1,i}(\theta^0) + n^{-1} \sum_{k=1}^n \ddot{\ell}_{\theta_1 u_D, k}(\theta^0) \psi_{u_D, k, i}(\theta_D^0) \right\}.$$

The expression of  $\psi_{u_D, k, i}(\theta_D^0)$  is given in Section S3 of the Online Supplement.

An application of Theorem 2 is to consider linear functionals in the form of  $\varphi = \mathbf{b}_{\beta_T}^\top \beta_T + \int_0^\xi b_{R_T}(t) dR_T(t) + b_\alpha \hat{\alpha}$  for  $F_1(\theta_1)$ , where  $b_\alpha$  and  $\mathbf{b}_{\beta_T}$  are a known scalar and vector, respectively, and  $\mathbf{b}_{R_T} \equiv (b_{R_T,1}, \dots, b_{R_T,\kappa_T})^\top$  is the known vector of the values of  $b_{R_T}(t)$  evaluated at  $0 < t_1 < t_2 < \dots < t_{\kappa_T} \leq \xi$ . The asymptotic variance of  $\varphi$  is given by  $\mathbf{b}_1^\top \Sigma_1^0 \mathbf{b}_1$ , where  $\mathbf{b}_1 = (\mathbf{b}_{\beta_T}^\top, \mathbf{b}_{R_T}^\top, b_\alpha)^\top$ . The variance-covariance matrix  $\Sigma_1^0$  can be consistently estimated by

$$\widehat{\Sigma}_{1,n} = n^{-1} \sum_{i=1}^n \psi_{\theta_1,i}(\widehat{\theta}) \psi_{\theta_1,i}(\widehat{\theta})^\top, \quad (9)$$

where

$$\psi_{\theta_1,i}(\widehat{\theta}) = \bar{\mathcal{I}}_{1,n}^{-1}(\widehat{\theta}) \left\{ \dot{\ell}_{\theta_1,i}(\widehat{\theta}) + n^{-1} \sum_{k=1}^n \ddot{\ell}_{\theta_1 u_D, k}(\widehat{\theta}) \psi_{u_D, k, i}(\widehat{\theta}_D) \right\}$$

with  $\bar{\mathcal{I}}_{1,n}(\widehat{\theta}) = n^{-1} \sum_{i=1}^n \mathcal{I}_{1,i}(\widehat{\theta})$ . This estimator can be re-written in the form of a robust sandwich variance estimator:  $\widehat{\Sigma}_{1,n} = \bar{\mathcal{I}}_{1,n}^{-1}(\widehat{\theta}) \bar{\mathcal{V}}_{1,n}(\widehat{\theta}) \bar{\mathcal{I}}_{1,n}^{-1}(\widehat{\theta})$  with  $\bar{\mathcal{V}}_{1,n}(\widehat{\theta}) = n^{-1} \sum_{i=1}^n \{\dot{\ell}_{\theta_1,i}(\widehat{\theta}) + n^{-1} \sum_{k=1}^n \ddot{\ell}_{\theta_1 u_D, k}(\widehat{\theta}) \psi_{u_D, k, i}(\widehat{\theta}_D)\} \{\dot{\ell}_{\theta_1,i}(\widehat{\theta}) + n^{-1} \sum_{k=1}^n \ddot{\ell}_{\theta_1 u_D, k}(\widehat{\theta}) \psi_{u_D, k, i}(\widehat{\theta}_D)\}^\top$ .

With such an analytic variance estimator, we can construct a 95% confidence interval (CI) for  $\alpha$ :  $\hat{\alpha} \pm 1.96 \times s_{\hat{\alpha}}$ , where the standard error (SE)  $s_{\hat{\alpha}}$  is the square root of the last diagonal element of  $\widehat{\Sigma}_{1,n}$  (i.e.,  $\mathbf{b}_{\beta_T} = \mathbf{0}$ ,  $\mathbf{b}_{R_T} = \mathbf{0}$ , and  $b_\alpha = 1$ ); similarly, a 95% CI for  $\beta_{T,k}$ , the  $k$ -th element of  $\beta_T$ :  $\hat{\beta}_{T,k} \pm 1.96 \times s_{\hat{\beta}_{T,k}}$ , where  $\hat{\beta}_{T,k}$  is the  $k$ -th element of  $\widehat{\beta}_T$ , and the SE  $s_{\hat{\beta}_{T,k}}$  is the square root of the  $k$ -th diagonal element of  $\widehat{\Sigma}_{1,n}$  (i.e.,  $\mathbf{b}_{\beta_T}$  is a vector with the  $k$ -th element being one and other elements being zero,  $\mathbf{b}_{R_T} = \mathbf{0}$ , and  $b_\alpha = 0$ ). For a given  $t_0$ , let  $k_{t_0} = \sum_{k=1}^{\kappa_T} \mathbb{1}(t_k \leq t_0)$ . A point estimate of the baseline survival probabilities  $S_{T,0}(t_0) = \exp[-G_T\{R_T(t_0)\}]$  is  $\widehat{S}_T(t_0) = \exp[-G_T\{\widehat{R}_T(t_0)\}]$  with  $\widehat{R}_T(t_0) = \mathbf{b}_{R_T(t_0)}^\top \widehat{dR}_T$ , where  $\widehat{dR}_T = (\widehat{dR}_{T,1}, \dots, \widehat{dR}_{T,\kappa_T})^\top$ , and  $\mathbf{b}_{R_T(t_0)}$  is a  $\kappa_T$ -dimensional vector with the first  $k_{t_0}$  elements being one, and the rest being zero. In addition, a 95% CI for  $S_{T,0}(t_0)$  is

$[e^{-G_T(R_L)}, e^{-G_T(R_U)}]$  with  $[R_L, R_U] = \widehat{R}_T(t_0) \pm 1.96 \times s_{\widehat{R}_T(t_0)}$ , where the SE  $s_{\widehat{R}_T(t_0)}$  is the square root of  $\mathbf{b}_1^\top \widehat{\Sigma}_{1,n} \mathbf{b}_1$  with  $\mathbf{b}_1 = (\mathbf{0}^\top, \mathbf{b}_{R_T(t_0)}^\top, 0)^\top$  (i.e.,  $\mathbf{b}_{\beta_T} = \mathbf{0}$  and  $b_\alpha = 0$ ).

### 3 Numerical studies

#### 3.1 Simulation studies

##### 3.1.1 Data generation

We generated baseline covariates  $Z_{1,i} \sim \mathcal{N}(1, 0.5) \times \mathbb{1}[0, 2]$ , a truncated normal distribution, and  $Z_{2,i} \sim \mathcal{B}(0.8)$ , a Bernoulli distribution. Given  $\mathbf{Z}_i = (Z_{1,i}, Z_{2,i})^\top$ , the outcomes  $(T, D)$  were generated from  $\log(T_i/3) = -\beta_{T,1}Z_{1,i} - \beta_{T,2}Z_{2,i} + e_{T,i}$  and  $\log(D_i/3) = -\beta_{D,1}Z_{1,i} - \beta_{D,2}Z_{2,i} + e_{D,i}$ , where  $(e_{T,i}, e_{D,i})$  followed a bivariate distribution that corresponds to a certain parametric copula function. Here, we considered two commonly used copula families: Gumbel and Clayton, with different strengths of association specified by Kendall's  $\tau$ , defined as  $\tau = 4 \int_{[0,1]^2} \mathbb{C}(u_1, u_2) d\mathbb{C}(u_1, u_2) - 1$ . Kendall's  $\tau$  has a one-to-one correspondence with  $\alpha$ : for example,  $\tau \equiv \tau(\alpha) = (\alpha - 1)/\alpha$  for Gumbel and  $\tau(\alpha) = \alpha/(\alpha + 2)$  for Clayton. We considered three dependence levels  $\tau = 0.4, 0.6, 0.8$ . As an example, to generate  $(e_{T,i}, e_{D,i})$  from the Gumbel copula with  $\tau = 0.6$ , we first generated  $(u_{T,i}, u_{D,i})$  with parameter  $\alpha = 2.5$  corresponding to  $\tau = 0.6$ , and then obtained  $e_{j,i} = \log\{-\log(u_{j,i})\}$  for  $j = T, D$ . We set  $\beta_D = (0.2, 0)^\top$ , which leads to about 18% of death times censored, and considered two sets of  $\beta_T = (1, 1)^\top$  and  $(1, 0.5)^\top$ , which leads to different censoring rates, 3%-12%, and 5%-18%, respectively, for the non-terminal event.

##### 3.1.2 Simulation study I

The first simulation study focused on evaluating the estimation accuracy and efficiency of the proposed PMLE approach and comparing its finite-sample performances with the one-stage MLE method from Chen (2012) under each combination of  $(\tau, \beta_T)$ . The sample size  $n$  was set to 200 and 400, with 1,000 replications of simulated datasets. We specified  $G_T(t) = G_D(t) = t$  (i.e., PH model) for the marginal distributions and the copula function as the true copula family. Both the PMLE and one-stage MLE were obtained based on the same log-likelihood function in (8) using the **R** package **trust** (Geyer., 2020), which implements the trust region optimization algorithm (Fletcher, 2000). The **trust** package requires users to provide the starting values of the parameters. To achieve faster convergence, we used the following estimates instead of random values. For  $\theta_D$  and  $\theta_T$ , we fit marginal models with the data  $(C_i, \delta_{D,i}, \mathbf{Z}_i)$  and  $(X_i, \delta_{T,i}, \mathbf{Z}_i)$ , correspondingly. These two estimates were plugged into (8), and the resulting pseudo-log-likelihood function was maximized to obtain the starting value of the copula parameter  $\alpha$ . For most combinations of the copula family and  $(n, \tau, \beta_T)$ , the optimization converged for all 1,000 replications. The exceptions are under Clayton with Kendall's  $\tau = 0.8$ , there were seven and four non-converged replications for  $\beta_T = (1, 1)^\top$  and  $(1, 0.5)^\top$  respectively with  $n = 200$ , and three and one non-converged replications for  $\beta_T = (1, 1)^\top$  and  $(1, 0.5)^\top$  respectively with  $n = 400$ .

Tables 1 and 2 present the summary statistics of the MLE and PMLE of  $(\beta_T, \alpha)$  for sample size  $n = 200$  under the Gumbel and Clayton copula, respectively. In addition, we examined the estimation of the baseline survival function of  $T$ , where, specifically, three time points,  $t_0 = 0.863, 2.079$ , and  $4.159$ , were selected such that the corresponding baseline survival probability  $S_{T,0}(t_0) = 0.75, 0.5$ , and  $0.25$ . Tables 3 and 4 present the summary statistics for the MLE and PMLE of these three baseline survival probabilities. The summary statistics include the relative bias (BIAS), empirical standard deviation (ESD), average standard error (ASE), and the root mean square error (rMSE), as well as the empirical coverage percentage of the 95% CI using the analytic standard error estimator. The PMLE's SEs were calculated via the procedure described in Section 2.3. The MLE's SEs were calculated by a robust variance-covariance matrix estimator:  $\bar{\mathcal{I}}_n^{-1}(\hat{\theta}_{\text{MLE}}) \bar{\mathcal{V}}_n(\hat{\theta}_{\text{MLE}}) \bar{\mathcal{I}}_n^{-1}(\hat{\theta}_{\text{MLE}})$  with  $\bar{\mathcal{I}}_n(\theta) = -n^{-1} \sum_{i=1}^n \ddot{\ell}_{\theta\theta,i}(\theta)$  and  $\bar{\mathcal{V}}_n(\theta) = n^{-1} \sum_{i=1}^n \dot{\ell}_{\theta,i}(\theta) \dot{\ell}_{\theta,i}(\theta)^\top$ , where  $\dot{\ell}_{\theta,i}(\theta) = \partial \ell_i(\theta) / \partial \theta$  and  $\ddot{\ell}_{\theta\theta,i}(\theta) = \partial^2 \ell_i(\theta) / \partial \theta \partial \theta^\top$  are the first-order and second-order derivatives of the log-likelihood function  $\ell_i(\theta)$  in (8) with respect to  $\theta$ . The results for sample size  $n = 400$  are reported in Web Tables 1 - 4 of the Online Supplement.

For both copula families, the PMLE performs well with a small bias relative to the estimation variability. As the sample size  $n$  increases, the bias decreases, demonstrating its consistency, and the ESD and, consequently, rMSE also decrease. The censoring rate also affects PMLE's performance. Compared to  $\beta_T = (1, 1)^\top$ ,  $\beta_T = (1, 0.5)^\top$ , resulting in a higher censoring rate, leads to a larger BIAS, ESD, and rMSE. The same pattern is also observed for the MLE. In addition, for most scenarios, the PMLE's ASE is close to its ESD, and the empirical coverage percentages of the CI are close to the nominal level, 95%. These results confirm the validity of our analytic SE estimator. The only exception is for the baseline survival probability  $S_{T,0}(t_0) = 0.25$ , where the empirical coverage percentages for both MLE and PMLE are significantly lower than 95%. This is because fewer subjects have observed events at a larger time point due to censoring.

Compared with the MLE, the PMLE's ESD for the marginal distribution parameters is generally larger, which is expected since the MLE is asymptotically more efficient. The PMLE of the copula parameter  $\alpha$  has a smaller ESD in some scenarios, especially for sample size  $n = 200$ . When the sample size increases to 400, the lead by the PMLE diminishes, and for some scenarios, the MLE exhibits better efficiency. To confirm the relationship of asymptotic efficiency between the PMLE and the MLE, we conducted additional simulations for sample size  $n = 1,000$  under the scenario of Kendall's  $\tau = 0.8$  and  $\beta_T = (1, 1)^\top$ ; the results are reported in Web Table 5 of the Online Supplement. When the sample size increases to 1,000, the MLE has a smaller ESD than the PMLE, confirming that the MLE is asymptotically more efficient.

We also compared the computational time between the PMLE and MLE with sample sizes  $n = 200$ , 400, and 1,000 for Kendall's  $\tau = 0.8$  and  $\beta_T = (1, 1)^\top$ . When the copula family is Gumbel, using a Mac Air with Apple M2 chip and 24 GB memory, the average running times of 10 replications for the PMLE are 0.53 secs, 2.14 secs, and 20.5 secs, for sample sizes  $n = 200$ , 400, and 1,000, respectively, and the average running times for MLE are 1.85 secs, 10.5 secs, and 200 secs. When the copula family is Clayton, the average running times for the three sample sizes are 0.33 secs, 1.16 secs, 13.8 secs for the PMLE, 1.02 secs, 5.6 secs, and 171 secs for MLE. This, together with the estimation results, shows that the PMLE is computationally more efficient than the MLE while maintaining comparable performance. Furthermore, we compared the MLE of the marginal distribution parameter for  $D$  with our first-stage estimator  $\hat{\theta}_D$  using its own survival outcome without accounting for the dependence between  $T$  and  $D$ . The results are reported in Web Tables 6 - 13 of the Online Supplement. Note that the three time points  $t_0 = 0.863, 2.079, 4.159$ , selected for the baseline probabilities of  $T$ , also leads to the baseline probabilities of  $D$  being  $S_{D,0}(t_0) = 0.75, 0.5, 0.25$ , respectively. Both estimators have a small bias, and their rMSE is dominated by the variance. The MLE has a smaller ESD because it utilizes extra information on the joint distribution of  $(T, D)$ .

### 3.1.3 Simulation study II

We conducted a second simulation study to assess the robustness of the PMLE and MLE under the misspecification of the copula function. Specifically, we considered two settings: (i) we generated  $(T, D)$  from the Gumbel copula but fit the Clayton copula, and (ii) vice versa. The sample size was  $n = 400$ , Kendall's  $\tau = 0.6$ , and  $\beta_T = (1, 0.5)^\top$ . When the copula function is misspecified, the MLE or PMLE estimator  $\hat{\alpha}$  will not converge to the true value of the copula parameter under the true copula family, and the plug-in estimate of Kendall's  $\tau$ ,  $\hat{\tau} = \tau(\hat{\alpha})$ , will not converge to the true Kendall's  $\tau$  either. Thus, it is more meaningful to focus on estimating the marginals in this setting. Table 5 presents the BIAS, ESD, and rMSE for  $\beta_D$  and  $\beta_T$  with or without the copula misspecification. When the copula is correctly specified, the bias of the PMLE and MLE for  $\beta_D$  are similar; when the copula is misspecified, the PMLE's bias remains unchanged. This is because the PMLE estimate of  $\beta_D$  is obtained without the information of the copula function and, consequently, is still a consistent estimator under copula misspecification. In contrast, the MLE's bias increases substantially under the copula misspecification. For the estimation of  $\beta_T$ , both the MLE and PMLE perform less ideally due to the misspecification. Still, when the true copula is Gumbel, but the Clayton copula is fitted, the PMLE gives a much smaller bias, demonstrating better finite-sample robustness. When the true copula is Clayton but the Gumbel copula is fitted, the biases for the PMLE and MLE are similar.

$\tau$	$\beta_T$	Method	BIAS	ESD	ASE	rMSE	CP
$\widehat{\beta}_{T,1}$							
0.4	(1,1)	MLE	0.005	0.160	0.160	0.160	94.0
		PMLE	0.006	0.162	0.163	0.162	94.7
	(1,0.5)	MLE	0.006	0.163	0.163	0.163	94.6
		PMLE	0.007	0.165	0.166	0.166	94.7
0.6	(1,1)	MLE	0.001	0.156	0.157	0.156	94.3
		PMLE	0.002	0.163	0.165	0.163	94.7
	(1,0.5)	MLE	0.000	0.158	0.158	0.158	94.6
		PMLE	0.003	0.164	0.166	0.164	94.4
0.8	(1,1)	MLE	0.007	0.156	0.156	0.156	94.4
		PMLE	0.005	0.168	0.171	0.168	95.3
	(1,0.5)	MLE	0.005	0.157	0.157	0.157	94.4
		PMLE	0.006	0.169	0.171	0.169	94.8
$\widehat{\beta}_{T,2}$							
0.4	(1,1)	MLE	0.012	0.203	0.197	0.203	94.2
		PMLE	0.009	0.204	0.201	0.205	95.4
	(1,0.5)	MLE	0.024	0.393	0.378	0.394	93.8
		PMLE	0.020	0.396	0.385	0.397	93.8
0.6	(1,1)	MLE	0.016	0.198	0.191	0.198	93.6
		PMLE	0.009	0.204	0.200	0.204	95.2
	(1,0.5)	MLE	0.025	0.383	0.366	0.383	93.3
		PMLE	0.015	0.394	0.384	0.394	93.9
0.8	(1,1)	MLE	0.024	0.196	0.189	0.197	93.4
		PMLE	0.006	0.211	0.205	0.211	94.6
	(1,0.5)	MLE	0.033	0.382	0.363	0.383	93.2
		PMLE	0.011	0.409	0.394	0.409	94.5
$\widehat{\alpha}$							
0.4	(1,1)	MLE	0.036	0.075	0.078	0.083	96.8
		PMLE	0.023	0.073	0.075	0.077	96.4
	(1,0.5)	MLE	0.037	0.078	0.080	0.086	96.1
		PMLE	0.024	0.075	0.076	0.079	96.3
0.6	(1,1)	MLE	0.054	0.087	0.088	0.103	93.6
		PMLE	0.026	0.083	0.084	0.087	95.6
	(1,0.5)	MLE	0.055	0.089	0.090	0.105	94.1
		PMLE	0.027	0.084	0.085	0.088	95.7
0.8	(1,1)	MLE	0.096	0.105	0.104	0.142	89.7
		PMLE	0.020	0.093	0.095	0.095	95.3
	(1,0.5)	MLE	0.094	0.106	0.105	0.142	89.5
		PMLE	0.020	0.093	0.095	0.096	95.3

Table 1: Simulation results for the estimates of  $(\beta_T, \alpha)$  under the Gumbel copula with sample size  $n = 200$ . BIAS: relative bias; ESD: relative empirical standard deviation; ASE: relative average standard error; rMSE: relative root mean square error; CP: empirical coverage percentage of 95% confidence interval.

$\tau$	$\beta_T$	Method	BIAS	ESD	ASE	rMSE	CP
$\widehat{\beta}_{T,1}$							
0.4	(1,1)	MLE	0.014	0.154	0.152	0.155	94.6
		PMLE	0.014	0.160	0.160	0.161	94.7
	(1,0.5)	MLE	0.016	0.160	0.157	0.161	93.5
		PMLE	0.016	0.166	0.165	0.167	93.8
0.6	(1,1)	MLE	0.018	0.143	0.139	0.144	93.5
		PMLE	0.018	0.161	0.163	0.162	95.6
	(1,0.5)	MLE	0.022	0.147	0.142	0.149	93.6
		PMLE	0.021	0.165	0.166	0.167	95.3
0.8	(1,1)	MLE	0.034	0.133	0.133	0.137	94.4
		PMLE	0.027	0.173	0.174	0.175	96.0
	(1,0.5)	MLE	0.035	0.135	0.135	0.140	94.9
		PMLE	0.029	0.174	0.175	0.177	95.7
$\widehat{\beta}_{T,2}$							
0.4	(1,1)	MLE	0.009	0.197	0.191	0.197	93.6
		PMLE	0.003	0.207	0.201	0.207	93.8
	(1,0.5)	MLE	0.026	0.372	0.359	0.373	93.4
		PMLE	0.008	0.392	0.381	0.393	93.2
0.6	(1,1)	MLE	0.008	0.177	0.171	0.177	93.4
		PMLE	0.012	0.210	0.201	0.210	93.9
	(1,0.5)	MLE	0.023	0.332	0.317	0.333	93.6
		PMLE	0.005	0.392	0.377	0.392	92.8
0.8	(1,1)	MLE	0.001	0.161	0.158	0.161	94.2
		PMLE	0.025	0.223	0.210	0.225	93.6
	(1,0.5)	MLE	0.011	0.303	0.295	0.304	94.1
		PMLE	0.019	0.413	0.393	0.414	93.5
$\widehat{\alpha}$							
0.4	(1,1)	MLE	0.063	0.186	0.186	0.196	94.8
		PMLE	0.026	0.184	0.183	0.186	94.4
	(1,0.5)	MLE	0.061	0.191	0.190	0.201	94.7
		PMLE	0.025	0.189	0.186	0.190	94.6
0.6	(1,1)	MLE	0.067	0.148	0.146	0.162	94.8
		PMLE	0.011	0.146	0.145	0.146	94.5
	(1,0.5)	MLE	0.065	0.151	0.148	0.165	94.1
		PMLE	0.010	0.149	0.146	0.150	93.6
0.8	(1,1)	MLE	0.101	0.138	0.136	0.171	92.1
		PMLE	0.011	0.125	0.128	0.126	95.1
	(1,0.5)	MLE	0.098	0.139	0.136	0.170	92.7
		PMLE	0.013	0.127	0.128	0.127	93.7

Table 2: Simulation results for the estimates of  $(\beta_T, \alpha)$  under the Clayton copula with  $n = 200$ . BIAS: relative bias; ESD: relative empirical standard deviation; ASE: relative average standard error; rMSE: relative root mean square error; CP: empirical coverage percentage of 95% confidence interval.

$\tau$	$\beta_T$	Time	Prob.	Method	BIAS	ESD	ASE	rMSE	CP
0.4	(1,1)	0.863	0.750	MLE	0.007	0.077	0.076	0.078	94.3
				PMLE	0.005	0.078	0.077	0.078	94.2
	(1,0.5)	2.079	0.500	MLE	0.009	0.172	0.170	0.173	93.4
				PMLE	0.006	0.174	0.172	0.174	93.5
	(1,1)	4.159	0.250	MLE	0.162	0.397	0.334	0.428	82.4
				PMLE	0.168	0.397	0.338	0.432	81.7
	(1,0.5)	0.863	0.750	MLE	0.007	0.077	0.075	0.077	94.0
				PMLE	0.005	0.077	0.076	0.078	93.6
	(1,1)	2.079	0.500	MLE	0.008	0.169	0.169	0.170	93.3
				PMLE	0.005	0.170	0.171	0.171	93.8
	(1,0.5)	4.159	0.250	MLE	0.066	0.373	0.339	0.379	89.0
				PMLE	0.073	0.375	0.343	0.382	89.0
0.6	(1,1)	0.863	0.750	MLE	0.009	0.075	0.074	0.075	95.2
				PMLE	0.006	0.077	0.076	0.077	94.8
	(1,0.5)	2.079	0.500	MLE	0.009	0.166	0.165	0.166	94.2
				PMLE	0.004	0.169	0.170	0.169	94.0
	(1,1)	4.159	0.250	MLE	0.133	0.379	0.327	0.401	84.1
				PMLE	0.142	0.382	0.337	0.408	84.7
	(1,0.5)	0.863	0.750	MLE	0.008	0.074	0.073	0.075	94.5
				PMLE	0.005	0.076	0.075	0.076	94.7
	(1,1)	2.079	0.500	MLE	0.008	0.162	0.163	0.162	93.8
				PMLE	0.003	0.166	0.168	0.166	93.9
	(1,0.5)	4.159	0.250	MLE	0.052	0.357	0.328	0.360	89.2
				PMLE	0.060	0.362	0.339	0.367	89.6
0.8	(1,1)	0.863	0.750	MLE	0.011	0.073	0.072	0.074	95.7
				PMLE	0.005	0.077	0.076	0.077	94.4
	(1,0.5)	2.079	0.500	MLE	0.008	0.163	0.162	0.163	93.9
				PMLE	0.001	0.171	0.171	0.171	93.5
	(1,1)	4.159	0.250	MLE	0.116	0.359	0.321	0.378	86.3
				PMLE	0.130	0.369	0.340	0.391	86.8
	(1,0.5)	0.863	0.750	MLE	0.011	0.073	0.072	0.074	95.2
				PMLE	0.004	0.077	0.075	0.077	94.2
	(1,1)	2.079	0.500	MLE	0.007	0.160	0.160	0.160	93.4
				PMLE	0.000	0.169	0.169	0.169	93.5
	(1,0.5)	4.159	0.250	MLE	0.050	0.338	0.323	0.341	90.7
				PMLE	0.062	0.349	0.342	0.355	91.2

Table 3: Simulation results for the baseline survival function of  $T$  under the Gumbel copula with sample size  $n = 200$ . BIAS: relative bias; ESD: relative empirical standard deviation; ASE: relative average standard error; rMSE: relative root mean square error; CP: empirical coverage percentage of 95% confidence interval.

$\tau$	$\beta_T$	Time	Prob.	Method	BIAS	ESD	ASE	rMSE	CP
0.4	(1,1)	0.863	0.750	MLE	0.003	0.077	0.075	0.077	93.2
				PMLE	0.001	0.078	0.078	0.078	93.5
	(1,0.5)	2.079	0.500	MLE	0.002	0.168	0.163	0.168	93.2
				PMLE	0.005	0.172	0.170	0.172	93.1
	(1,1)	4.159	0.250	MLE	0.134	0.380	0.313	0.403	81.6
				PMLE	0.151	0.381	0.326	0.410	82.5
	(1,0.5)	0.863	0.750	MLE	0.002	0.076	0.075	0.076	92.9
				PMLE	0.000	0.077	0.077	0.077	93.5
	(1,1)	2.079	0.500	MLE	0.001	0.164	0.164	0.164	92.9
				PMLE	0.005	0.168	0.171	0.169	93.3
0.6	(1,0.5)	4.159	0.250	MLE	0.056	0.350	0.319	0.355	89.1
				PMLE	0.073	0.358	0.333	0.365	89.2
	(1,1)	0.863	0.750	MLE	0.000	0.069	0.068	0.069	92.8
				PMLE	0.001	0.077	0.076	0.077	92.6
	(1,1)	2.079	0.500	MLE	0.000	0.151	0.146	0.151	92.9
				PMLE	0.009	0.167	0.165	0.167	93.1
	(1,0.5)	4.159	0.250	MLE	0.120	0.341	0.286	0.362	83.1
				PMLE	0.145	0.358	0.321	0.386	85.0
	(1,0.5)	0.863	0.750	MLE	0.002	0.068	0.067	0.068	92.8
				PMLE	0.002	0.076	0.075	0.076	93.1
0.8	(1,1)	2.079	0.500	MLE	0.003	0.148	0.146	0.148	93.2
				PMLE	0.007	0.165	0.165	0.165	92.9
	(1,1)	4.159	0.250	MLE	0.048	0.309	0.290	0.312	90.5
				PMLE	0.070	0.334	0.325	0.341	91.0
	(1,1)	0.863	0.750	MLE	0.006	0.063	0.062	0.063	93.2
				PMLE	0.005	0.078	0.076	0.079	92.6
	(1,0.5)	2.079	0.500	MLE	0.008	0.139	0.137	0.139	93.5
				PMLE	0.014	0.169	0.168	0.170	92.0
	(1,0.5)	4.159	0.250	MLE	0.141	0.320	0.272	0.350	83.7
				PMLE	0.171	0.362	0.330	0.400	85.5
	(1,0.5)	0.863	0.750	MLE	0.009	0.062	0.062	0.063	92.5
				PMLE	0.005	0.078	0.077	0.078	92.9
	(1,0.5)	2.079	0.500	MLE	0.009	0.135	0.136	0.135	93.4
				PMLE	0.010	0.167	0.168	0.167	92.7
	(1,0.5)	4.159	0.250	MLE	0.058	0.288	0.275	0.293	90.9
				PMLE	0.079	0.338	0.335	0.347	90.7

Table 4: Simulation results for the baseline survival function of  $T$  under the Clayton copula with sample size  $n = 200$ . BIAS: relative bias; ESD: relative empirical standard deviation; ASE: relative average standard error; rMSE: relative root mean square error; CP: empirical coverage percentage of 95% confidence interval.

Parameter	Method	BIAS		ESD		MSE	
		GC	GG	GC	GG	GC	GG
True: Gumbel							
$\beta_{D,1}$	MLE	0.070	0.003	0.108	0.111	0.129	0.111
	PMLE	0.002	0.002	0.118	0.118	0.118	0.118
$\beta_{D,2}$	MLE	0.019	0.002	0.122	0.134	0.124	0.134
	PMLE	0.001	0.001	0.144	0.144	0.144	0.144
$\beta_{T,1}$	MLE	0.195	0.001	0.116	0.113	0.226	0.113
	PMLE	0.078	0.002	0.115	0.118	0.139	0.118
$\beta_{T,2}$	MLE	0.104	0.008	0.130	0.137	0.166	0.137
	PMLE	0.045	0.006	0.137	0.143	0.144	0.143
True: Clayton							
		CG	CC	CG	CC	CG	CC
$\beta_{D,1}$	MLE	0.053	0.005	0.123	0.095	0.133	0.095
	PMLE	0.003	0.003	0.117	0.117	0.117	0.117
$\beta_{D,2}$	MLE	0.021	0.013	0.149	0.112	0.151	0.113
	PMLE	0.001	0.001	0.145	0.145	0.145	0.145
$\beta_{T,1}$	MLE	0.041	0.011	0.123	0.103	0.129	0.103
	PMLE	0.041	0.013	0.125	0.116	0.131	0.117
$\beta_{T,2}$	MLE	0.013	0.006	0.152	0.118	0.152	0.118
	PMLE	0.018	0.003	0.154	0.140	0.155	0.140

Table 5: Simulation results for copula misspecification. GC: true Gumbel - fit Clayton; GG: Gumbel - Gumbel; GC: Gumbel - Clayton; CG: Clayton - Gumbel; CC: Clayton - Clayton; CG: Clayton - Gumbel. BIAS: bias; ESD: empirical standard deviation; rMSE: root mean square error.

### 3.1.4 Simulation study III

We also compared our PMLE with the two-stage estimator developed in Zhu et al. (2021) under the Gumbel copula. Specifically, Zhu et al. (2021), in the first stage, estimated the regression coefficient  $\beta_D$  first, which produced  $\widehat{U}_{D,i} = S_D(X_{D,i}|\mathbf{Z}_i; \widehat{\beta}_D)$ , and with the estimated marginal of  $D$  and a non-parametric estimator of the marginal of  $T^*$  (time to the first event, which is only subject to independent censoring), a PMLE of the copula parameter was obtained. In the second stage, the time-varying regression coefficient,  $\beta_T(t)$ , was estimated via a separate set of non-linear estimating equations. For comparison purposes, we followed the same simulation setting of Zhu et al. (2021). Since the marginal specifications are different, we focused on the comparison of the copula parameter estimation in terms of the BIAS, ESD, and rMSE, presented in Table 6, where the results of Zhu et al. (2021)'s method were extracted from Table 1 of their paper directly. It needs to be pointed out that the definition of the Gumbel copula parameter in Zhu et al. (2021) is different from ours, and their Gumbel copula function is defined as

$$C(u_1, u_2; \alpha^*) = \exp \left( - \left[ \{-\log(u_1)\}^{1/\alpha^*} + \{-\log(u_2)\}^{1/\alpha^*} \right]^{\alpha^*} \right),$$

which indicates that our Gumbel copula parameter in (4)  $\alpha = 1/\alpha^*$ . Based on this relationship, we obtained the PMLE of  $\alpha^*$  by  $\widehat{\alpha}^* = 1/\widehat{\alpha}$ , and with this estimator, we calculated the summary statistics presented in Table 6. Across all the combinations  $(n, \tau, \beta_T)$ , our PMLE has given more desirable performances in all three metrics.

## 3.2 Analysis of the BMT data

We applied the proposed method to the BMT data, available in the **R** package **SemiCompRisks**, which consists of 137 patients with acute myelocytic leukemia (AML) or acute lymphoblastic leukemia (ALL) aged 7 to 52 (Klein and Moeschberger, 2003). This data set contains time to death (terminal event,

$\tau$	$\alpha^*$	$\beta_T$	Method	$n = 200$			$n = 400$		
				BIAS	ESD	rMSE	BIAS	ESD	rMSE
0.4	0.6	(1,1)	PMLE	0.011	0.042	0.043	0.005	0.030	0.030
			Zhu et al. (2021)	0.034	0.059	0.068	0.028	0.058	0.064
	0.6	(1,0.5)	PMLE	0.011	0.043	0.044	0.006	0.031	0.031
			Zhu et al. (2021)	0.032	0.061	0.069	0.029	0.059	0.066
0.6	0.4	(1,1)	PMLE	0.008	0.031	0.032	0.004	0.022	0.023
			Zhu et al. (2021)	0.058	0.064	0.086	0.041	0.057	0.070
	0.4	(1,0.5)	PMLE	0.008	0.032	0.033	0.004	0.023	0.023
			Zhu et al. (2021)	0.051	0.060	0.079	0.039	0.056	0.068
0.8	0.2	(1,1)	PMLE	0.002	0.018	0.018	0.001	0.012	0.012
			Zhu et al. (2021)	0.078	0.056	0.096	0.057	0.042	0.071
	0.2	(1,0.5)	PMLE	0.002	0.018	0.018	0.001	0.012	0.012
			Zhu et al. (2021)	0.068	0.048	0.083	0.055	0.042	0.069

Table 6: Simulation results comparing the proposed PMLE and the method in Zhu et al. (2021). BIAS: bias; ESD: empirical standard deviation; rMSE: root mean square error.  $\alpha^*$  is the copula parameter of Gumbel copula given by  $\mathbb{C}(u_1, u_2; \alpha^*) = \exp(-\{-\log(u_1)\}^{1/\alpha^*} + \{-\log(u_2)\}^{1/\alpha^*} \alpha^*)$ .

$D$ ), subject to the administrative censoring of study time, and time to relapse (non-terminal event,  $T$ ), which is subject to censoring by both death and study time. We were interested in investigating the dependence between times to relapse and death and the effects of disease groups, AML-low, AML-high, and ALL, on the time to relapse and death. Here, AML-low was set as the baseline group.

The PH models were assumed for both marginal distributions, and three copula families, Clayton, Gumbel, and Frank, were considered. In addition, we assumed that under each copula family, the copula parameter may vary with the disease group. Both the MLE and PMLE approaches were implemented to estimate the parameters. For inference, the standard errors of PMLE and MLE were calculated by the same procedure used for the simulation study I.

The parameter estimates with their standard errors (in the parentheses) are shown in Table 7. For the dependence between the bivariate event times, we calculated the plug-in estimate of Kendall's  $\tau$ :  $\hat{\tau} = \tau(\hat{\alpha})$  under each copula family, where  $\hat{\alpha}$  is the MLE or PMLE of  $\alpha$ . As described earlier,  $\tau(\alpha) = (\alpha - 1)/\alpha$  for Gumbel, and  $\tau(\alpha) = \alpha/(\alpha + 2)$  for Clayton. In addition, for Frank,  $\tau(\alpha) = 1 - \frac{4}{\alpha} + \frac{4}{\alpha} \int_0^{\alpha} \frac{t}{e^{t-1}} dt$ . The standard error of  $\hat{\tau}$  was obtained via the delta method:  $\hat{\sigma}_{\tau} = |\dot{\tau}(\hat{\alpha})| \hat{\sigma}_{\alpha}$ , where  $\hat{\sigma}_{\tau}$  and  $\hat{\sigma}_{\alpha}$  denote the standard error for  $\hat{\tau}$  and  $\hat{\alpha}$ , respectively.

Comparing the MLE and PMLE under each copula family, the two approaches produce similar point estimates and standard errors. For marginal distributions, compared to the AML-low group, both the AML-high and ALL groups have significantly higher risks of death and relapse of leukemia under all copula families. Figure 1 plots the estimated survival functions for the time to relapse for each disease group with 95% confidence intervals from the PMLE. The estimates from the MLE are given in Web Figure 1 of the Online Supplement. The AML-low group has the highest survival rates, and the AML-high has the lowest rates.

Regarding the dependence between times to relapse and death, under each copula family, a strong and similar association is observed for the three groups. A difference across the three copula families is that under Clayton, the association is strongest for the AML-low group, but the AML-high group has the strongest association under Frank and Gumbel. We compared the log-likelihood function evaluated at the PMLE and MLE, and the Clayton copula family has the highest value using both methods. Thus, Clayton is the most suitable of these three families, using the log-likelihood as the selection criterion (equivalent to using the Akaike information criterion or Bayesian information criterion because

Covariates	Gumbel		Clayton		Frank	
	PMLE	MLE	PMLE	MLE	PMLE	MLE
Regression coefficients for leukemia relapse						
AML high	1.239 (0.317)	1.147 (0.313)	1.168 (0.311)	1.116 (0.306)	1.137 (0.302)	1.032 (0.305)
ALL	0.854 (0.345)	0.764 (0.325)	0.710 (0.324)	0.669 (0.320)	0.716 (0.318)	0.645 (0.318)
Regression coefficients for death						
AML high	1.022 (0.276)	0.953 (0.281)	1.022 (0.276)	0.977 (0.271)	1.022 (0.276)	0.905 (0.272)
ALL	0.611 (0.285)	0.553 (0.284)	0.611 (0.285)	0.577 (0.280)	0.611 (0.285)	0.535 (0.279)
Kendall's $\tau$ between event times						
AML low	0.679 (0.140)	0.709 (0.122)	0.814 (0.084)	0.821 (0.079)	0.737 (0.105)	0.738 (0.106)
AML high	0.726 (0.090)	0.769 (0.089)	0.769 (0.108)	0.777 (0.102)	0.760 (0.094)	0.770 (0.095)
ALL	0.686 (0.090)	0.726 (0.079)	0.767 (0.100)	0.773 (0.093)	0.721 (0.094)	0.730 (0.095)
Log-Likelihood	-4.474	-4.456	-4.436	-4.432	-4.447	-4.443

Table 7: Estimation results on the BMT data using the PMLE and MLE methods. Standard errors are given in parentheses.

all alternatives have the same number of parameters). Furthermore, under Clayton, the difference between the PMLE and MLE for  $\beta_D$  is the smallest. As shown in simulation study II, the PMLE and MLE for the marginal distribution parameter of  $D$  are similar when the copula family is correctly specified, but under copula misspecification, they diverge. Thus, the smallest gap between these two estimators under Clayton is another piece of evidence suggesting that Clayton is the best family among the three candidates.

## 4 Conclusions

This article proposed a two-stage likelihood-based estimation procedure for analyzing semi-competing risks data. The dependence between the non-terminal and terminal event times is modeled by a parametric copula function, and the marginals are specified via semi-parametric transformation models. One motivation for such a stage-wise estimation is to alleviate the computational challenge arising from the one-stage MLE approach. Through simulation studies, our proposed PMLE method has given a more desirable finite-sample performance than the MLE for the copula parameter. On the other hand, despite being less efficient than the MLE for the regression coefficients of the marginal for  $D$ , the proposed PMLE method is robust against copula misspecification. This is because the estimation of  $S_D(\cdot)$  in the first stage is guaranteed to be consistent. As a result, the estimation of  $S_T(\cdot)$  might lead to a smaller bias than the one-stage MLE. Despite this robustness, the PMLE of  $S_T(\cdot)$  is inconsistent if the copula function is misspecified. Thus, it is ideal to pair the method with a goodness-of-fit test for the copula specification, where the biggest challenge is to distinguish the misspecification of the copula function from the misspecification of  $S_T(\cdot)$ , which is a future research problem.

Although our log-likelihood function in (8) is equivalent to the counting-process-based log-likelihood in Chen (2012) (Remark 3), the respective one-stage MLEs would be different for finite samples due to the discretization of  $R_j(\cdot)$  for  $j = T, D$ . Thus, we observe close but not exactly the same results between Table 7 of this article and Table 5 of Chen (2012). However, we expect this difference will decrease with a larger sample size.

The popular strategy for addressing the dependent censoring in estimating the survival function of  $T$  was using the first-event time  $T^*$ . However, its implementation is limited to the Archimedean copula family, and when  $T$  is heavily censored by  $D$ , the estimation for the tail of  $S_T(\cdot)$  is challenging (Fine et al., 2001; Lakhhal et al., 2008). In addition,  $S_{T^*}(\cdot)$  is often estimated nonparametrically even with the covariates, contributing to the less ideal estimation efficiency. Though this could be improved by using

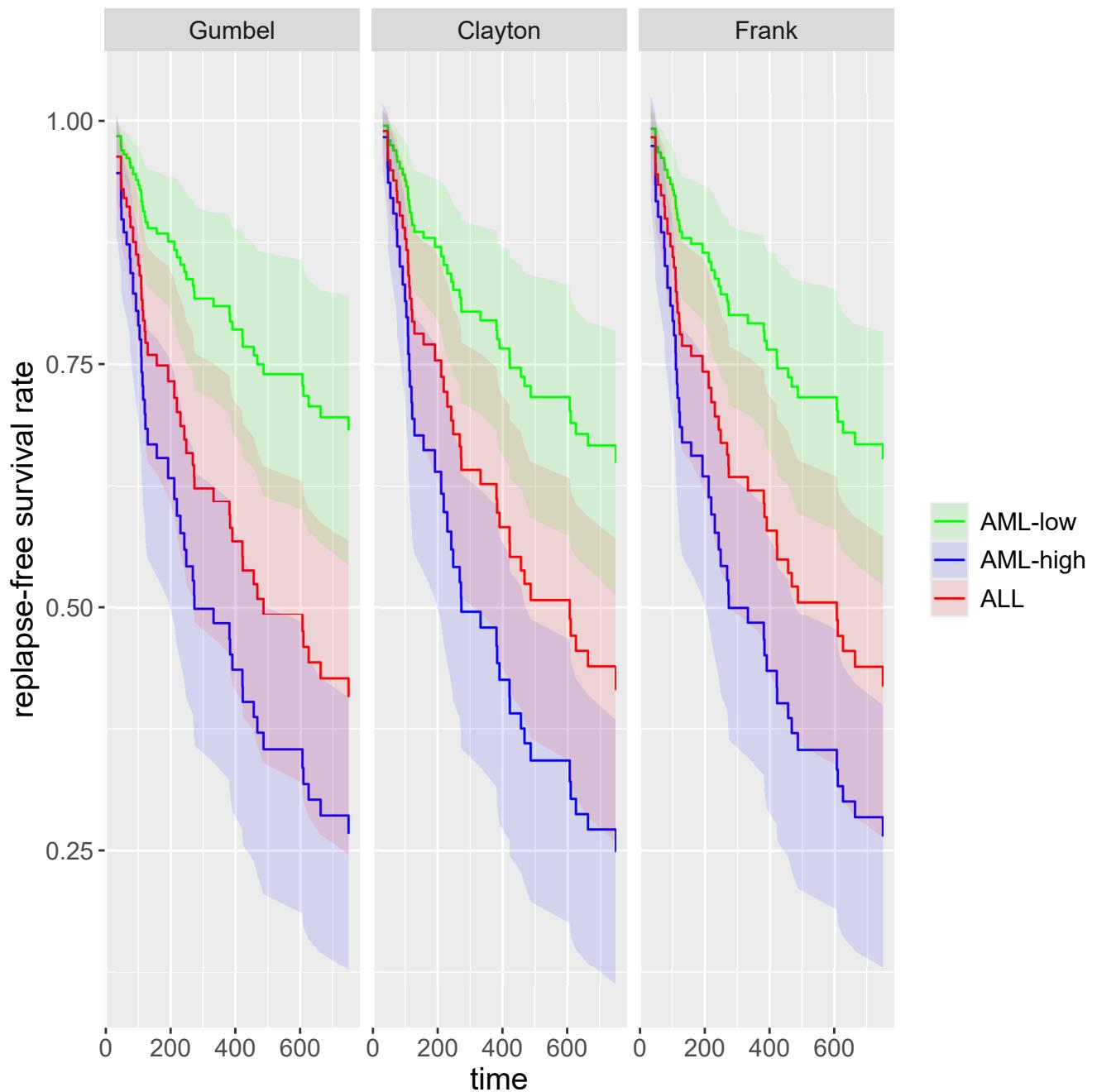


Figure 1: PMLE estimates of relapse-free survival probabilities  $\mathbb{P}(T > t)$  for each disease group. The shaded areas indicate the 95% pointwise confidence intervals.

a semi-parametric model, it is likely to violate the model assumptions imposed on both  $T$  and  $D$ . For example, if one assumes that  $T$  and  $D$  both follow PH models, then the survival function of  $T^*$  does not follow a PH model and might not even have an explicit expression.

In the semiparametric copula model, the copula parameter could depend on the baseline covariates via a parametric specification. In the nonsurvival context, some works focused on nonparametrically modeling the functional relationship between the copula parameter and covariates, e.g., Acar et al. (2011) and Abegaz et al. (2012). Specifically, let  $W$  be a continuous covariate, and it is connected with the copula parameter  $\alpha$  via  $\eta(w) = \phi^{-1}(\alpha)$ , where  $\phi^{-1}$  is the inverse of the link function  $\phi$  introduced in Remark 2. In these papers,  $\eta(w)$  was approximated by a Taylor expansion of polynomial of degree  $q$  with the coefficient parameters  $\gamma = (\gamma_0, \dots, \gamma_q)^\top$ . Geerdens et al. (2018) extended this model to right-censored clustered survival data. In these articles, the function  $\eta(w)$  was estimated by maximizing a kernel-smoothed pseudo-log-likelihood function:  $\ell(\gamma) = \sum_{i=1}^n \ell_i(\gamma) K_h(W_i - w)$ , where  $\ell_i$  is the local pseudo log-likelihood contribution from subject  $i$ , and  $K_h(\cdot)$  is a kernel function. Since our estimation procedure is also likelihood-based, it can potentially accommodate such a nonparametric specification  $\eta(w)$  for the copula parameter. In the above existing works, the marginal survival functions were estimated separately. However, for semi-competing risks data,  $\eta(w)$  would be estimated along with the parameters associated with the marginal distribution of  $T$ .

## acknowledgements

This work is supported by the grant DMS-2210481 from the National Science Foundation.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *Journal of Multivariate Analysis*, 110:43–73.

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics*, 67(2):445–453.

Acute Respiratory Distress Syndrome Network (2000). Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *New England Journal of Medicine*, 342(18):1301–1308. PMID: 10793162.

Chen, X., Harhay, M. O., Tong, G., and Li, F. (2024). A Bayesian machine learning approach for estimating heterogeneous survivor causal effects: Applications to a critical care trial. *The Annals of Applied Statistics*, 18(1):350–374.

Chen, Y.-H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2):235–251.

Chen, Y.-H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis*, 18:36–57.

Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88(4):907–919.

Fletcher, R. (2000). *Practical Methods of Optimization*. John Wiley & Sons, Chichester, UK.

Geerdens, C., Acar, E. F., and Janssen, P. (2018). Conditional copula models for right-censored clustered event time data. *Biostatistics*, 19(2):247–262.

Geyer., C. J. (2020). *trust: Trust Region Optimization*. R package version 0.1-8.

Ghosh, D. (2006). Semiparametric inferences for association with semi-competing risks data. *Statistics in Medicine*, 25(12):2059–2070.

Hsieh, J.-J., Wang, W., and Ding, A. A. (2008). Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(1):3–20.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, NY.

Lakhal, L., Rivest, L.-P., and Abdous, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics*, 64(1):180–188.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, NY, 2 edition.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27(30):6393–6406.

Peng, L. and Fine, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics*, 63(1):96–108.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399.

Sklar, M. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Annales de l'ISUP*, 8(3):229–231.

Sun, T., Li, Y., Xiao, Z., Ding, Y., and Wang, X. (2023). Semiparametric copula method for semi-competing risks data subject to interval censoring and left truncation: Application to disability in elderly. *Statistical Methods in Medical Research*, 32(4):656–670.

Sun, T., Liang, W., Zhang, G., Yi, D., Ding, Y., and Zhang, L. (2024). Penalised semi-parametric copula method for semi-competing risks data: application to hip fracture in elderly. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 73(1):241–256.

Varadhan, R., Xue, Q.-L., and Bandeen-Roche, K. (2014). Semicompeting risks in aging research: methods, issues and needs. *Lifetime data analysis*, 20(4):538–562.

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1):257–273.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640.

Zhou, R., Zhu, H., Bondy, M., and Ning, J. (2016). Semiparametric model for semi-competing risks data with application to breast cancer study. *Lifetime data analysis*, 22:456–471.

Zhu, H., Lan, Y., Ning, J., and Shen, Y. (2021). Semiparametric copula-based regression modeling of semi-competing risks data. *Communications in Statistics-Theory and Methods*, 51(22):7830–7845.