



Article

Partially Linear Additive Hazards Regression for Bivariate Interval-Censored Data

Ximeng Zhang ¹, Shishun Zhao ¹, Tao Hu ^{2,*} and Jianguo Sun ³¹ Center for Applied Statistical Research, College of Mathematics, Jilin University, Changchun 130012, China² School of Mathematical Sciences, Capital Normal University, Beijing 100048, China³ Department of Statistics, University of Missouri, Columbia, MO 65211, USA

* Correspondence: hutao@cnu.edu.cn

Abstract: In this paper, we discuss regression analysis of bivariate interval-censored failure time data that often occur in biomedical and epidemiological studies. To solve this problem, we propose a kind of general and flexible copula-based semiparametric partly linear additive hazards models that can allow for both time-dependent covariates and possible nonlinear effects. For inference, a sieve maximum likelihood estimation approach based on Bernstein polynomials is proposed to estimate the baseline hazard functions and nonlinear covariate effects. The resulting estimators of regression parameters are shown to be consistent, asymptotically efficient and normal. A simulation study is conducted to assess the finite-sample performance of this method and the results show that it is effective in practice. Moreover, an illustration is provided.

Keywords: Archimedean copula model; Bernstein polynomials; bivariate interval-censored data; partly linear model

MSC: 62N02; 62H05; 62G05; 62G20



Citation: Zhang, X.; Zhao, S.; Hu, T.; Sun, J. Partially Linear Additive Hazards Regression for Bivariate Interval-Censored Data. *Axioms* **2023**, *12*, 198. <https://doi.org/10.3390/axioms12020198>

Academic Editors: Hsien-Chung Wu, Jiajuan Liang and Kaitai Fang

Received: 22 December 2022

Revised: 7 February 2023

Accepted: 11 February 2023

Published: 13 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this article, we discuss a regression analysis using the marginal additive hazards model on bivariate interval-censored data. Interval-censored data refer to failure times that are observed to only belong to an interval rather than being known with absolute certainty. These types of data frequently occur in various areas, including biomedical and epidemiological investigations [1]. It is easy to see that the most commonly studied right-censored data can be seen as a special case of interval-censored data, and it is important to note that the analysis of interval-censored data is typically far more difficult than right-censored data research. Bivariate interval-censored data occur when there exist two correlated failure times of interest and the observed times of both failure events suffer interval censoring. Clinical trials or medical studies on several events from the same individual such as eye disease studies often yield bivariate interval-censored data. The purpose of this article is to propose a flexible regression model which can process bivariate interval-censored data when the main interest is on the risk differential or excess risk.

There are several methods available for regression analysis of univariate interval-censored data arising from the additive hazards model. Refs. [2,3], for example, investigated the problem for case I interval-censored or current state data, a special form of interval-censored data in which the observed time contains zero or infinite. The former discussed an estimating equation approach and the latter considered an efficient estimation approach. More recently, Ref. [4] studied the same issue as Ref. [2] but with some inequality constraints, and Ref. [5] proposed a sieve maximum likelihood approach. Moreover, Ref. [6] developed several inverse probability weight-based and reweighting-based estimation procedures for the situation with missing covariates, while Ref. [7] presented an efficient approach for general situations.

A large number of approaches have been established for modeling bivariate interval-censored survival data and three types of methods are generally used. One is a marginal method that relies on the working independence assumption [8,9]. Another commonly used approach is the frailty-model-based method, which employs the frailty or latent variable to build the correlation among the associated failure times [10,11]. Ref. [12] proposed a frailty model method for multivariate interval-censored data with informative censoring. The third type of method is the copula-based approach, which gives a different, specific way to model two dependent failure times. One advantage of the approach is that it directly connects the two marginal distributions through a copula function to construct the joint distribution and uses the copula parameter to determine the correlation. This distinguishing characteristic makes it possible to represent the margins independent of the copula function. This advantage is appealing for the points of both modeling and interpretation views. Among others, Ref. [13] discussed this approach based on the marginal transformation model with the two-parameter Archimedean copula. Ref. [14] proposed a copula link-based additive model for right-censored event time data. Moreover, Ref. [15] proposed a copula-based model to deal with bivariate survival data with various censoring mechanisms. In the following, we discuss the regression analysis of bivariate cases by employing the method in [13]. More specifically, we propose a kind of semiparametric partly linear additive hazards model.

Partly linear models are becoming more and more common since they combine the flexibility of nonparametric modeling with the simplicity and ease of interpretation of parametric modeling. It is presumptive that the marginally conditional hazard function has nonlinear relationships with some covariates but linear relationships with others [16–18]. In practice, nonlinear covariate effects are typical. For instance, in some medication research, the influence of the dosage of a particular medication may reach a peak at a certain dosage level and be maintained at the peak level, or it may diminish after the dosage level. Although there is a fair amount of literature in this field, to the best of our knowledge, there does not seem to exist a study considering this for bivariate interval-censored survival data.

The presented model involves two nonparametric functions, one identifies the baseline cumulative hazard function and another describes the nonlinear effects of a continuous covariate. In the following, the two-parameter Archimedean copula model is employed for the dependence and more comments on this is given below. Moreover, a sieve maximum likelihood estimation approach is developed to approximate two involved nonparametric nuisance functions by Bernstein polynomials. The proposed method has several desirable features: (a) it allows both time-dependent covariates and the covariates that may have nonlinear effects; (b) the two-parameter Archimedean copula model can flexibly handle dependence structures on both upper and lower tails and the strength of the dependence can be quantified via Kendall's τ ; (c) the sieve maximum likelihood estimation approach by Bernstein polynomials can be easily implemented with the use of some existing software; (d) as is seen in the simulation study, the computation is both stable and efficient. Note that the method given by [13] only allows for linear covariate effects.

More specifically, in Section 2, after recommending some notation, assumptions and models that would be used throughout the paper, the resulting likelihood function is presented. In Section 3, we first describe the proposed sieve maximum likelihood estimation approach and then present the asymptotic properties of the proposed estimators. Section 4 presents a simulation study for the assessment of the finite sample performance of the proposed estimation approach, and the results indicate that it works well as expected. In Section 5, an illustration is provided by using a set of data arising from the Age-Related Eye Disease Study (AREDS), and Section 6 gives the conclusion with some discussion and concluding remarks.

2. Assumptions and Likelihood Function

Consider a study consisting of n independent subjects. Define T_{ij} as the failure time of interest associated with the i th subject of the j th failure event. Suppose that for T_{ij} , the

two observation times are given by $(U_{ij}, V_{ij}]$ such that $U_{ij} < T_{ij} \leq V_{ij}$. In addition, suppose that for the i th subject, the p -dimensional covariate vectors are possibly time-dependent and denoted by X_{i1} and X_{i2} and the single continuous covariates Z_{i1} and Z_{i2} are related to T_{i1} and T_{i2} , respectively. More details on them are given below. Here, we assume that X_{i1} and X_{i2} or Z_{i1} and Z_{i2} could be the same, entirely different, or they also could have some common components. Then, the observed data are as follows:

$$O = \{O_i = (U_{ij}, V_{ij}, X_{ij}, Z_{ij}); i = 1, \dots, n, j = 1, 2\}.$$

Note that when $V_{ij} = \infty$, T_{ij} is right-censored and when $U_{ij} = 0$, T_{ij} is left-censored. Moreover, it is assumed that given the covariates, the interval censoring is independent of the failure times T_{ij} [1].

Given some covariates, define $S_{ij}(t_{ij}|X_{ij}, Z_{ij}) = P(T_{ij} > t_{ij}|X_{ij}, Z_{ij})$ as the marginal survival function of T_{ij} , and

$$S(t_{i1}, t_{i2}|X_{i1}, X_{i2}, Z_{i1}, Z_{i2}) = P(T_{i1} > t_{i1}, T_{i2} > t_{i2}|X_{i1}, X_{i2}, Z_{i1}, Z_{i2}),$$

is the joint survival function of T_{i1} and T_{i2} . For the covariate effects, suppose that given X_{ij} and Z_{ij} , the marginal hazard function of T_{ij} is defined as follows:

$$h_j(t|X_{ij}, Z_{ij}) = \lambda_j(t) + \beta^T X_{ij}(t) + g(Z_{ij}),$$

where $\lambda_j(t)$ denotes an unknown baseline hazard function, β is an unknown regression coefficient vector, and $g(\cdot)$ is an unknown, smooth nonlinear regression function. That is, T_{ij} follows a partially linear additive hazards model, in which Z_{ij} represents the covariate that may have nonlinear effects on T_{ij} . Correspondingly, T_{ij} has the cumulative hazard

$$H_j(t|X_{ij}, Z_{ij}) = \int_0^t h_j(s|X_{ij}, Z_{ij})ds = \Lambda_j(t) + \beta^T W_{ij}(t) + g(Z_{ij})t,$$

where $\Lambda_j(t) = \int_0^t \lambda_j(s)ds$, $W_{ij}(t) = \int_0^t X_{ij}(s)ds$, and we have that

$$\begin{aligned} S_{ij}(t|X_{ij}, Z_{ij}) &= \exp\left[-\int_0^t h_j(s|X_{ij}, Z_{ij})ds\right] = \exp[-H_j(t; X_{ij}, Z_{ij})] \\ &= \exp\left[-\{\Lambda_j(t) + \beta^T W_{ij}(t) + g(Z_{ij})t\}\right], \quad i = 1, \dots, n, j = 1, 2, \end{aligned} \tag{1}$$

It is worth noting that for the simplicity of the expressions and calculations, we assume that both linear and nonlinear covariates effects are the same for the two associated failure times of interest [19,20]. It is simple to extend the following method to the situation where the covariate effects are different.

It follows from Sklar’s theorem [21] that if the marginal survival functions $S_{ij}(\cdot)$ are continuous, there exists a unique copula function $C_{\xi}(\cdot, \cdot)$ on $[0, 1]^2$ such that $C_{\xi}(s_1, 0) = C_{\xi}(0, s_2) = 0$, $C_{\xi}(s_1, 1) = s_1$ and $C_{\xi}(1, s_2) = s_2$, and it gives

$$S(t_{i1}, t_{i2}|X_{i1}, X_{i2}, Z_{i1}, Z_{i2}) = C_{\xi}(S_{i1}(t_{i1}|X_{i1}, Z_{i1}), S_{i2}(t_{i2}|X_{i2}, Z_{i2})), t_{i1}, t_{i2} \geq 0.$$

Here, the parameter ξ generally denotes the correlation or dependence between T_{i1} and T_{i2} . As mentioned above, a significant advantage of the copula representation above is that it separates the correlation from the two marginal distributions [22]. There exist many copula functions and among others, one type of the most commonly used copula functions

for bivariate data is perhaps the Archimedean copula family. By following [13], we focus on the flexible two-parameter Archimedean copula model given by

$$\begin{aligned}
 C_{\varphi,\omega}(s_1, s_2) &= \left[1 + \left\{ \left(s_1^{-1/\omega} - 1 \right)^{1/\varphi} + \left(s_2^{-1/\omega} - 1 \right)^{1/\varphi} \right\}^\varphi \right]^{-\omega} \\
 &= \mu \left(\mu^{-1}(s_1) + \mu^{-1}(s_2) \right), \quad \varphi \in (0, 1], \omega \in (0, \infty),
 \end{aligned}
 \tag{2}$$

where $\mu(s) = \mu_{\varphi,\omega}(s) = (1 + s^\varphi)^{-\omega}$ and μ^{-} is the generalized inverse of μ , which is defined as $\mu^{-}(y) = \inf\{x \in \mathbb{R} : y \leq \mu(x)\}$, $y \in \mathbb{R}$. The detailed derivation and more comments can be found in Chapter 5 of [23].

As mentioned before, the two parameters φ and ω in the copula function above are the association parameters, representing the correlation in both the upper and lower tails. In particular, when $\varphi = 1$, the copula model above is equal to the Clayton copula [24], while if $\omega \rightarrow \infty$, the copula model becomes the Gumbel copula [25]. In other words, the two-parameter copula model is more flexible and has the Clayton or Gumbel copula as special cases. It is well-known that another commonly used measure for the correlation is Kendall’s τ , and it has an explicit connection with φ, ω as

$$\tau = 1 - 2\varphi\omega / (2\omega + 1).$$

Under the assumptions above, the observed likelihood function has the form

$$\begin{aligned}
 L_n(S_1, S_2, \varphi, \omega | O) &= \prod_{i=1}^n P(U_{i1} < T_{i1} \leq V_{i1}, U_{i2} < T_{i2} \leq V_{i2} | Y_{i1}, Y_{i2}) \\
 &= \prod_{i=1}^n \{ P(T_{i1} > U_{i1}, T_{i2} > U_{i2} | Y_{i1}, Y_{i2}) - P(T_{i1} > U_{i1}, T_{i2} > V_{i2} | Y_{i1}, Y_{i2}) \\
 &\quad - P(T_{i1} > V_{i1}, T_{i2} > U_{i2} | Y_{i1}, Y_{i2}) + P(T_{i1} > V_{i1}, T_{i2} > V_{i2} | Y_{i1}, Y_{i2}) \} \\
 &= \prod_{i=1}^n \{ S(U_{i1}, U_{i2} | Y_{i1}, Y_{i2}) - S(U_{i1}, V_{i2} | Y_{i1}, Y_{i2}) - S(V_{i1}, U_{i2} | Y_{i1}, Y_{i2}) \\
 &\quad + S(V_{i1}, V_{i2} | Y_{i1}, Y_{i2}) \} \\
 &= \prod_{i=1}^n \{ C_{\varphi,\omega}[S_1(U_{i1} | Y_{i1}), S_2(U_{i2} | Y_{i2})] - C_{\varphi,\omega}[S_1(U_{i1} | Y_{i1}), S_2(V_{i2} | Y_{i2})] \\
 &\quad - C_{\varphi,\omega}[S_1(V_{i1} | Y_{i1}), S_2(U_{i2} | Y_{i2})] + C_{\varphi,\omega}[S_1(V_{i1} | Y_{i1}), S_2(V_{i2} | Y_{i2})] \},
 \end{aligned}
 \tag{3}$$

where $Y_{i1} = (X_{i1}, Z_{i1}), Y_{i2} = (X_{i2}, Z_{i2})$. Let $\eta = (\beta, \varphi, \omega, \Lambda_1, \Lambda_2, g)$, all unknown parameters. The next section proposes a sieve approach for the estimation of η .

3. Sieve Maximum Likelihood Estimation

It is well-known that directly maximizing the likelihood function $L_n(\eta | O) = L_n(S_1, S_2, \varphi, \omega | O) = \sum_{i=1}^n L(\eta; O_i)$ can provide an estimate of η . On the other hand, it is inappropriate for this situation, as it involves infinite-dimensional functions $\Lambda_j(t)$ and $g(\cdot)$. For this, according to [26] and others, we suggest using the sieve method to approximate these functions based on Bernstein’s polynomial first, and then maximize the likelihood function.

Specifically, define $\Theta = \mathcal{A} \otimes \mathcal{M} \otimes \mathcal{M} \otimes \mathcal{G}$, the parameter space, and

$$\mathcal{A} = \left\{ (\beta, \varphi, \omega) \in R^p \times R^{(0,1]} \times R^+, \|\beta\| + \|\varphi\| + \|\omega\| \leq K \right\},$$

in which \otimes denotes the Kronecker product, and K is a nonnegative constant. Additionally, \mathcal{M} denotes the subset of all bounded and continuous, nondecreasing, nonnegative functions within $[c_1, d_1], 0 \leq c_1 < d_1 < \infty$. Similarly, \mathcal{G} denotes the collection of all bounded and continuous functions within $[c_2, d_2]$, the support of Z_{ij} . In practice, $[c_1, d_1]$ is generally

valued as the large and minimum values of all observation times. In the following, define the sieve parameter space

$$\Theta_n = \left\{ \eta_n = \left(\beta^T, \varphi, \omega, \Lambda_{n1}, \Lambda_{n2}, g_n \right)^T \in \mathcal{A} \otimes \mathcal{M}_n \otimes \mathcal{M}_n \otimes \mathcal{G}_n \right\}.$$

In the above,

$$\mathcal{M}_n = \left\{ \Lambda_{nj}(t) = \sum_{k=0}^{m_1} \phi_{jl}^* B_l(t, m_1, c_1, d_1) : \sum_{l=0}^{m_1} \phi_{jl}^* \leq M_n; 0 \leq \phi_{j0}^* \leq \dots \leq \phi_{jm_1}^*; j = 1, 2 \right\},$$

and

$$\mathcal{G}_n = \left\{ g_n(z) = \sum_{k=0}^{m_2} \alpha_k B_k(z, m_2, c_2, d_2) : \sum_{k=0}^{m_2} |\alpha_k| \leq G_n \right\}$$

with

$$B_k(x, m, c, d) = \binom{m}{k} \left(\frac{x-c}{d-c} \right)^k \left(1 - \frac{x-c}{d-c} \right)^{m-k}, k = 0, \dots, m,$$

the Bernstein basis polynomial with the degrees $m_1 = O(n^{\nu_1})$ and $m_2 = O(n^{\nu_2})$ for some fixed $\nu_1, \nu_2 \in (0, 1)$, and M_n and G_n being some positive constants. To estimate η , define $\hat{\eta}_n = (\hat{\beta}_n, \hat{\varphi}_n, \hat{\omega}_n, \hat{\Lambda}_{n1}, \hat{\Lambda}_{n2}, \hat{g})$ as the value of η by maximizing the log-likelihood function

$$\ell_n(\eta; O) = \log L_n(\eta; O) = \sum_{i=1}^n \log L(\eta; O_i). \tag{4}$$

Note that one of the main advantages of Bernstein polynomials is that they can easily implement the nonnegativity and monotonicity properties of $\Lambda_j(t)$ by the reparameterization $\phi_{j0}^* = e^{\phi_{j0}}$, $\phi_{jl}^* = \sum_{i=0}^l e^{\phi_{ji}}$, $\forall 1 \leq l \leq m_1$ [26]. In addition, of all approximation polynomials, they have the optimal shape-preserving properties [27]. By the way, they are easy to use because they do not need interior knots. In the above, for simplicity, the same basis polynomial is used for Λ_1 and Λ_2 . Moreover, note that this approach can be relatively easily implemented as discussed below, although it may seem to be complicated. In practice, one need to choose m_j ($j = 1, 2$). We suggest employing the Akaike information criterion (AIC) defined as

$$AIC = -2\ell_n(\hat{\eta}_n) + 2(p + m_1 + m_2 + 2 + 2).$$

In the above, the form in the second bracket is the number of unknown parameters in the model [28], where p denotes the dimension of β ; $m_1 + 1$ and $m_2 + 1$ represent the degree of the Bernstein polynomials (i.e., $m_1 + m_2 + 2$); and the last number denotes the dimension of correlation parameters φ and ω .

For the maximization of the log-likelihood function $\ell_n(\eta; O)$ over the sieve space Θ_n or the determination of $\hat{\eta}_n = (\hat{\beta}_n^T, \hat{\varphi}_n, \hat{\omega}_n, \hat{\Lambda}_{n1}, \hat{\Lambda}_{n2}, \hat{g})^T$, we suggest first determining the initial estimate of η and then applying the Newton–Raphson approach to maximize $\ell_n(\beta, \varphi, \omega, \Lambda_{n1}, \Lambda_{n2}, g_n; O)$. In the numerical study section, one can use the R function *nlm* for the maximization. For the determination of the initial estimates, the following procedure can be applied.

- Step 1: obtain $(\hat{\beta}^{(0)}, \hat{\Lambda}_{n1}^{(0)}, \hat{g}^{(0)})$ by maximizing the log marginal likelihood function under the observation data on the T_{i1} 's;
- Step 2: obtain $\hat{\Lambda}_{n2}^{(0)}$ by maximizing the log marginal likelihood function under the observation data on the T_{i2} 's;
- Step 3: obtain the initial estimates $(\hat{\varphi}^{(0)}, \hat{\omega}^{(0)})$ of φ and ω by maximizing the joint sieve log likelihood function $\ell_n(\hat{\beta}^{(0)}, \varphi, \omega, \hat{\Lambda}_{n1}^{(0)}, \hat{\Lambda}_{n2}^{(0)}, \hat{g}_n^{(0)})$.

Now, we establish the asymptotic properties of $\hat{\eta}_n$. Define $\eta^1 = (\beta^1, \varphi^1, \omega^1, \Lambda_1^1, \Lambda_2^1, g^1) \in \Theta_n$ and $\eta^2 = (\beta^2, \varphi^2, \omega^2, \Lambda_1^2, \Lambda_2^2, g^2) \in \Theta_n$, and their distance has the form

$$d(\eta^1, \eta^2) = \left\{ \|\beta^1 - \beta^2\|^2 + \|\varphi^1 - \varphi^2\|^2 + \|\omega^1 - \omega^2\|^2 + \|\Lambda_1^1 - \Lambda_1^2\|_2^2 + \|\Lambda_2^1 - \Lambda_2^2\|_2^2 + \|g^1 - g^2\|_2^2 \right\}^{1/2}.$$

In the above, $\|\Lambda_j\|_2^2 = \int [(\Lambda_j(u))^2 + (\Lambda_j(v))^2] dF_j(u, v)$ with $F_j(u, v)$ denoting the joint distribution function of the U_j 's and V_j 's, $j = 1, 2$. Let the true value of η be denoted $\eta_0 = (\beta_0, \varphi_0, \omega_0, \Lambda_{10}, \Lambda_{20}, g_0)$.

Theorem 1 (Consistency). *Suppose that Conditions 1–4 given in Appendix A hold. Then, we have $d(\hat{\eta}_n, \eta_0) \rightarrow 0$ almost surely as $n \rightarrow \infty$.*

Theorem 2 (Convergence rate). *Suppose that Conditions 1–5 given in the Appendix A hold. Then,*

$$d(\hat{\eta}_n, \eta_0) = O_p(n^{-\min\{qv_1/2, (1-v_1)/2, rv_2/2, (1-v_2)/2\}}),$$

where $v_1 \in (0, 1)$ and $v_2 \in (0, 1)$ such that $m_1 = o(n^{v_1})$, $m_2 = o(n^{v_2})$, and q and r are defined in Condition 4 of Appendix A.

Theorem 3 (Asymptotic normality). *Suppose that Conditions 1–5 given in Appendix A hold. Then, we have*

$$n^{1/2}(\hat{b}_n - b_0) \rightarrow_d N\{0, I^{-1}(b_0)\},$$

where $\hat{b}_n = (\hat{\beta}_n^T, \hat{\varphi}_n, \hat{\omega}_n)^T$, $b_0 = (\beta^{(0)T}, \varphi_0, \omega_0)^T$,

$$I(b_0) = \mathbb{P}\left\{ \dot{\ell}_b(\eta_0) - \dot{\ell}_{\Lambda_1}(\eta_0)[h_{\Lambda_1}^*] - \dot{\ell}_{\Lambda_2}(\eta_0)[h_{\Lambda_2}^*] - \dot{\ell}_g(\eta_0)[h_g^*] \right\}^{\otimes 2},$$

$w^{\otimes 2} = ww^T$ for an arbitrary vector w , and $\dot{\ell}_b(\eta_0)$, $\dot{\ell}_{\Lambda_1}(\eta_0)[h_{\Lambda_1}]$, $\dot{\ell}_{\Lambda_2}(\eta_0)[h_{\Lambda_2}]$, $\dot{\ell}_g(\eta_0)[h_g]$ are the score statistics defined in Appendix A.

We sketch the proofs of the theorems above in Appendix A. Note that based on Theorem 2, one can get the optimal rate of convergence $\min(n^{q/2(1+q)}, n^{r/2(1+r)})$ with $v_1 = 1/(1+q)$ and $v_2 = 1/(1+r)$. Specifically, it becomes $n^{1/3}$ with $r = 2$ or $q = 2$ and increases while q and r increases. Theorem 3 demonstrates that the suggested estimator of the regression parameter is nonetheless asymptotically normal and effective even when the total convergence rate is less than $n^{1/2}$. It is clear that to give the inference of the suggested estimators, we need to estimate the variance or covariance matrix of β , φ and ω . However, it can be seen from Appendix A that obtaining their consistent estimators would be difficult. As a result, we propose using the simple nonparametric bootstrap approach [29], which is well-known for offering a direct and simple tool for estimating covariances when no explicit formula is given. It looks to work well, according to the numerical analysis below.

4. A Simulation Study

In this section, we present some simulation studies conducted to evaluate the finite sample performance of the sieve maximum likelihood estimation approach suggested in the preceding sections. Three scenarios for covariates were taken into account in the study. The first one was to generate the single covariates X_i 's to follow the Bernoulli distribution with the success probability 0.5 and Z_i 's to follow the uniform distribution over $(0, 1)$. In scenario two, we considered the same Z_i 's as above but generated a two-dimensional vector of the covariate $(X_1, X_2)^T$ with $X_1 \sim \text{Bernoulli}(0.5)$ and $X_2 \sim U(0, 1)$. For scenario three, we first generated covariates X_i 's and Z_i 's as in scenario one and then replaced X_i by $X_i \exp(t)$. That is, we had time-dependent covariates.

To generate the true bivariate failure times (T_{i1}, T_{i2}) under model (1), one needed to generate u_{i2} and w_i from the uniform distribution over $(0, 1)$ independently for the first step

and solve u_{i1} from the equation $w_i = h(u_{i1}, u_{i2}) = \partial C_{\xi}^{-1}(u_{i1}, u_{i2}) / \partial u_{i2}$ for a given copula function C_{ξ} . Then, the two dependent survival times t_{i1} and t_{i2} were obtained based on $t_{i1} = S_1^{-1}(u_{i1})$ and $t_{i2} = S_2^{-1}(u_{i2})$, respectively, with $h_j(t|X, Z) = 2 + \beta X + \sin(\frac{1}{2}\pi Z)$ or $h_j(t|X, Z) = 2 + \beta X + Z^2, j = 1, 2$. In order to generate the censoring intervals or the observed data, we assumed that each subject was assessed at discrete time points, and the length of two contiguous observation times followed the standard exponential distribution. Then, for every subject, U_{ij} was valued as the last assessment time before T_{ij} , and V_{ij} was equal to the first assessment time behind T_{ij} . The length of the study was determined to yield about a 20% right-censoring rate. On the Bernstein polynomial approximation, by following [13], we set $m_1 = m_2 = 3$ and $\phi_{1k} = \phi_{2k}$ and also took a Kendall's τ equal to 0.3 for the weak dependency and 0.6 for the strong dependency. The results given below are based on 1000 replications and 100 bootstrap samples for the variance estimation.

Tables 1 and 2 present the results based on $n = 20, 50, 100, 300, 400$ given by the proposed estimation procedure on the estimation of the regression parameter β and the dependence parameter τ with the true $\beta_0 = 0$ or 0.5 and $g(z) = \sin(\frac{1}{2}\pi Z)$ or $g(z) = z^2$, respectively. The table consists of the bias of estimates (Bias) determined by the difference value between the mean of the estimates and the real value, the sample standard error (SSE) of the estimates based on 1000 replications, the mean of the estimated standard errors (ESE) (for one replication, we obtained the estimated standard errors based on 100 bootstrap samples and computed the average of the 1000 estimated standard errors), as well as the 95% empirical coverage probabilities (CP). The tables show that the proposed estimator was seemingly unbiased, and the variance estimation procedure seemed to perform well. Moreover, all empirical coverage probabilities were close to the nominal level 95% when the sample sizes were increasing, indicating that the normal approximation to the distribution of the proposed estimator seemed reasonable. Moreover, as expected, the results got better in general with the increasing sample size.

Table 1. Estimation results on β and τ with $g(z) = \sin(\frac{1}{2}\pi z)$.

τ	n	β	$\hat{\beta}$				$\hat{\tau}$				
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	
0.3	20	0	-0.047	1.611	2.702	1.000	-0.145	0.477	2.262	0.991	
		0.5	0.029	1.747	3.164	0.998	-0.119	0.541	1.842	0.987	
	50	0	0.034	0.692	0.831	0.978	-0.061	0.276	0.304	0.982	
		0.5	0.021	0.642	0.806	0.988	-0.048	0.286	0.303	0.977	
	100	0	-0.008	0.468	0.477	0.952	-0.022	0.145	0.172	0.978	
		0.5	0.009	0.469	0.498	0.965	-0.015	0.157	0.179	0.975	
	300	0	-0.004	0.254	0.254	0.949	-0.008	0.072	0.079	0.959	
		0.5	-0.007	0.271	0.263	0.940	0.003	0.073	0.081	0.961	
	400	0	0.003	0.219	0.220	0.951	-0.011	0.063	0.065	0.956	
		0.5	-0.005	0.237	0.228	0.940	-0.000	0.061	0.067	0.958	
	0.6	20	0	-0.080	1.718	3.595	1.000	-0.051	0.413	2.797	0.979
			0.5	0.035	1.751	3.564	1.000	-0.029	0.403	2.639	0.979
50		0	-0.023	0.578	0.749	0.983	0.015	0.177	0.233	0.980	
		0.5	0.025	0.576	0.799	0.990	-0.012	0.202	0.236	0.965	
100		0	-0.012	0.367	0.393	0.962	0.013	0.103	0.132	0.954	
		0.5	0.020	0.365	0.418	0.973	0.007	0.114	0.137	0.958	
300		0	-0.007	0.201	0.206	0.956	-0.009	0.064	0.066	0.933	
		0.5	0.015	0.206	0.215	0.962	0.005	0.064	0.069	0.931	
400		0	0.006	0.178	0.177	0.949	-0.008	0.055	0.056	0.932	
		0.5	0.011	0.188	0.184	0.946	0.001	0.055	0.057	0.945	

Table 2. Estimation results on β and τ with $g(z) = z^2$.

τ	n	β	$\hat{\beta}$				$\hat{\tau}$				
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	
0.3	20	0	−0.045	1.737	2.162	1.000	−0.138	0.485	2.536	0.983	
		0.5	0.074	1.922	2.941	0.997	−0.108	0.435	1.260	0.987	
	50	0	0.045	0.732	0.919	0.983	−0.052	0.215	0.249	0.993	
		0.5	0.076	0.767	0.961	0.989	−0.046	0.209	0.248	0.985	
	100	0	−0.010	0.509	0.531	0.956	−0.024	0.122	0.149	0.971	
		0.5	0.027	0.511	0.552	0.967	−0.019	0.120	0.145	0.979	
	300	0	−0.009	0.280	0.285	0.950	−0.013	0.060	0.066	0.956	
		0.5	−0.013	0.296	0.296	0.947	−0.005	0.057	0.064	0.968	
	400	0	−0.004	0.247	0.245	0.948	−0.010	0.052	0.055	0.960	
		0.5	−0.016	0.252	0.253	0.951	−0.003	0.056	0.057	0.956	
	0.6	20	0	0.157	1.767	2.243	1.000	−0.046	0.361	1.798	0.974
			0.5	0.195	2.134	2.373	0.997	−0.034	0.329	0.636	0.972
50		0	0.034	0.608	0.815	0.989	−0.029	0.158	0.208	0.984	
		0.5	0.025	0.632	0.843	0.979	−0.018	0.170	0.237	0.984	
100		0	−0.008	0.403	0.435	0.965	−0.020	0.096	0.116	0.969	
		0.5	0.028	0.417	0.449	0.963	0.014	0.095	0.118	0.970	
300		0	−0.003	0.222	0.225	0.956	−0.016	0.052	0.057	0.960	
		0.5	0.020	0.231	0.234	0.953	−0.010	0.057	0.058	0.956	
400		0	−0.001	0.199	0.195	0.957	−0.018	0.046	0.048	0.936	
		0.5	0.011	0.195	0.201	0.946	−0.006	0.050	0.052	0.952	

The estimation results obtained under scenario two for the covariates are given in Table 3; the sample size was $n = 200$ or 400 with the true $(\beta_1, \beta_2)^T = (0, 0)^T$ or $(0.1, 0.1)^T$ and $g(z) = \sin(\frac{1}{2}\pi z)$. Table 4 contains the estimation results obtained with the time-dependent covariates based on $\beta_0 = 0, 1$ or 0.5 as well as $n = 200$ or 400 , and the other values being the same as in Table 1. They again indicated that the proposed method seemed to perform well for the estimation of the regression and association parameters. Furthermore, the results generally became better when the sample size increased.

Table 3. Estimation results on β and τ with $g(z) = \sin(\frac{1}{2}\pi z)$ and two covariates.

Param	True Value	$n = 200$				$n = 400$			
		Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
β_1	0	0.010	0.457	0.433	0.950	0.007	0.318	0.299	0.948
β_2	0	0.010	0.571	0.550	0.947	0.007	0.392	0.380	0.947
τ	0.3	−0.009	0.089	0.103	0.963	−0.003	0.064	0.065	0.936
β_1	0.1	−0.007	0.316	0.321	0.959	0.006	0.218	0.221	0.953
β_2	0.1	−0.009	0.532	0.557	0.954	0.006	0.386	0.385	0.951
τ	0.3	−0.005	0.092	0.105	0.967	−0.003	0.063	0.066	0.943
β_1	0	−0.015	0.244	0.258	0.960	−0.009	0.176	0.177	0.948
β_2	0	−0.015	0.461	0.448	0.938	−0.004	0.300	0.308	0.947
τ	0.6	−0.001	0.081	0.083	0.931	−0.006	0.056	0.058	0.949
β_1	0.1	−0.003	0.257	0.261	0.955	0.003	0.183	0.179	0.945
β_2	0.1	−0.003	0.434	0.452	0.964	0.003	0.316	0.309	0.941
τ	0.6	0.002	0.080	0.084	0.931	0.001	0.055	0.056	0.938

Table 4. Estimation results on β and τ with the time-dependent covariates.

n	τ	β	$\hat{\beta}$				$\hat{\tau}$			
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
200	0.3	0	−0.016	0.265	0.270	0.946	−0.025	0.098	0.113	0.974
		0.5	0.027	0.279	0.286	0.948	−0.012	0.098	0.117	0.970
		1	0.023	0.301	0.310	0.954	−0.007	0.102	0.123	0.963
	0.6	0	0.025	0.231	0.234	0.942	−0.016	0.090	0.095	0.940
		0.5	0.046	0.233	0.243	0.957	−0.008	0.092	0.094	0.948
		1	0.042	0.259	0.265	0.945	−0.010	0.093	0.095	0.936
400	0.3	0	0.006	0.222	0.204	0.945	−0.004	0.060	0.070	0.957
		0.5	0.026	0.203	0.198	0.944	−0.003	0.068	0.071	0.959
		1	0.008	0.207	0.214	0.954	0.003	0.070	0.074	0.953
	0.6	0	0.020	0.159	0.165	0.953	−0.010	0.062	0.062	0.941
		0.5	0.038	0.155	0.163	0.952	−0.003	0.060	0.062	0.943
		1	0.028	0.182	0.183	0.946	0.005	0.061	0.063	0.936

To assess the performance of the proposed method on the estimation of the non-linear function g , we repeated the study given in Table 1 with four different g functions, $g_1(z) = \sin(2\pi z)$, $g_2(z) = \cos(2\pi z)$, $g_3(z) = -\frac{5}{3} + 5z^2$, and $g_4(z) = -z + 5z^3$, and $n = 200$. Figures 1 and 2 show the average of the estimated g for each of the four cases with $\beta = 0$ and $\tau = 0.3$ or $\tau = 0.6$, respectively. The solid red lines represent the genuine functions, while the dashed blue lines represent the estimations. They show that the proposed method based on Bernstein polynomials seemed to perform reasonably well for the different Kendall’s τ considered, including a weak or strong dependency. We also took into account various configurations and got comparable outcomes.

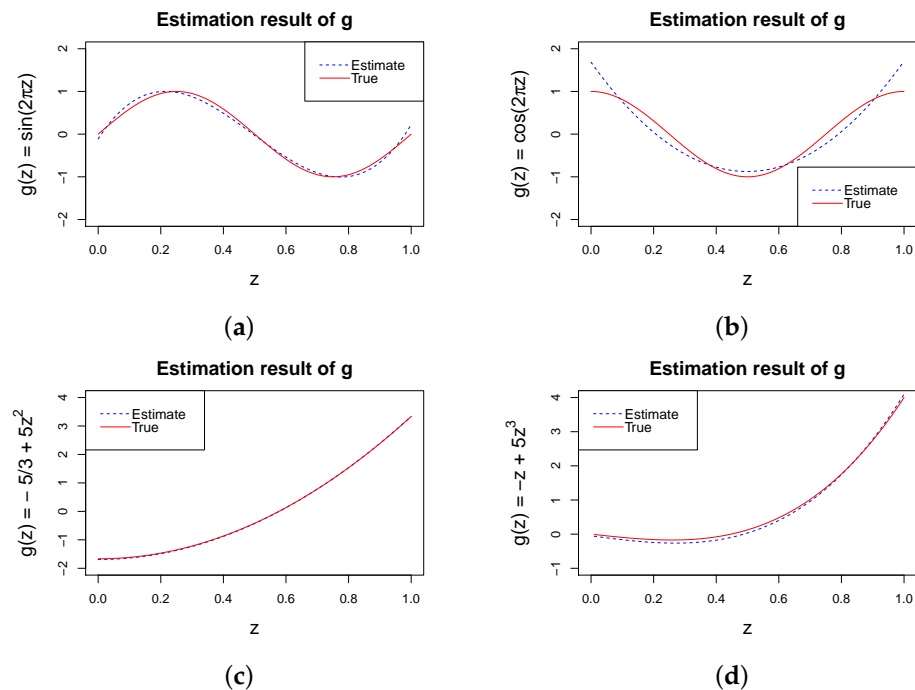


Figure 1. Estimated g with $\beta = 0$ and $\tau = 0.3$; the solid red lines represent the real functions and the dashed blue lines show the estimated functions. (a) $g(z) = \sin(2\pi z)$; (b) $g(z) = \cos(2\pi z)$; (c) $g(z) = -\frac{5}{3} + 5z^2$; (d) $g(z) = -z + 5z^3$.

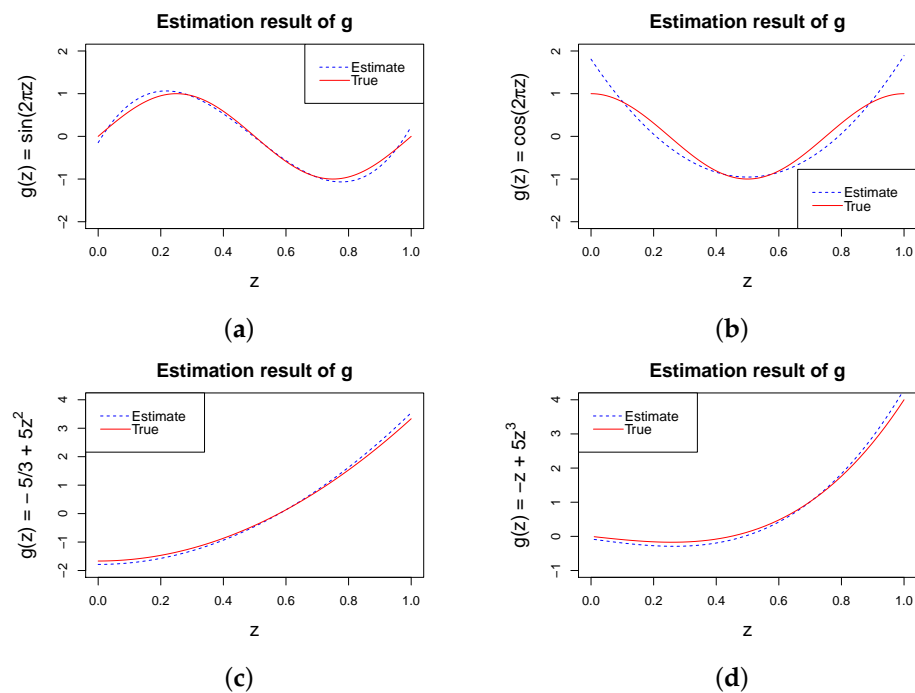


Figure 2. Estimated g with $\beta = 0$ and $\tau = 0.6$; the solid red lines represent the real functions and the dashed blue lines show the estimated functions. (a) $g(z) = \sin(2\pi z)$; (b) $g(z) = \cos(2\pi z)$; (c) $g(z) = \frac{5}{3} - 5z^2$; (d) $g(z) = -z + 5z^3$.

5. An Illustration

In this section, we illustrate the proposed procedure by using the data from the Age-related Eye Disease Study (AREDS) [30], a clinical experiment tracking the development of a bilateral eye disease, age-related macular degeneration (AMD), provided the CopulaCenR package in version 4.2.0 of R software [31]. Each participant in the research was monitored every six months (during the first six years) or once a year (after year 6) for about 12 years. At each appointment, each participant’s eyes were given a severity score on a range of 1 to 12 (a higher number signifying a more serious condition). Moreover, the time-to-late AMD, which is the interval between the baseline visit and the first appointment at which the severity score reached nine or above, was computed for each eye of these subjects. Either interval censoring or right censoring was applied to the observations at both periods.

The data set consisted of $n = 629$ subjects and in the analysis below, we focused on the effects of two covariates, SevScaleBL for the baseline AMD severity score (a value between one and eight with a higher value indicating more severe AMD) and rs2284665 for a genetic variant (zero, one and two for GG, GT and TT) of the AMD progression. In order to use the proposed method, we set rs2284665 to be X and SevScaleBL to be Z since the latter could be viewed as continuous. For the identifiability of the model, both X and Z were standardized and thus had the support $[0, 1]$. Moreover, various degrees of Bernstein polynomials were considered, such as $m_1 = m_2 = 3, 4, 5, 6$.

Table 5 gives the analysis results obtained by the suggested estimation approach. They include the estimated effect of the covariate rs2284665 and Kendall’s τ along with the estimated standard errors and the p -value for testing the effect to be zero. One can see that the results were consistent with respect to the degree of the Bernstein polynomials and suggested that the minor allele (TT) had a significantly “harmful” effect on AMD progression. Moreover, the estimated Kendall’s τ was 0.389, suggesting a moderate dependence of the AMD progression between the two eyes. Figure 3 gives the estimated effect of the SevScaleBL, which indeed seemed to be nonlinear. To be more specific, the increased risk of AMD patients was associated with higher severity scores. The findings reached here were in line with those of other researchers who examined this subject [15]. It is worth pointing

out that the conclusion of [15] was obtained under the proportional odds model and they could not visualize nonlinear covariate effects as in Figure 3.

Table 5. Analysis results for the AREDS data.

Bernstein Polynomials	β			τ	
	EST	ESE	p-Value	EST	ESE
$m_1 = 3, m_2 = 3$	0.041	0.009	<0.000	0.473	0.043
$m_1 = 4, m_2 = 4$	0.046	0.008	<0.000	0.416	0.056
$m_1 = 5, m_2 = 5$	0.034	0.008	<0.000	0.389	0.048
$m_1 = 6, m_2 = 6$	0.024	0.009	0.011	0.388	0.056

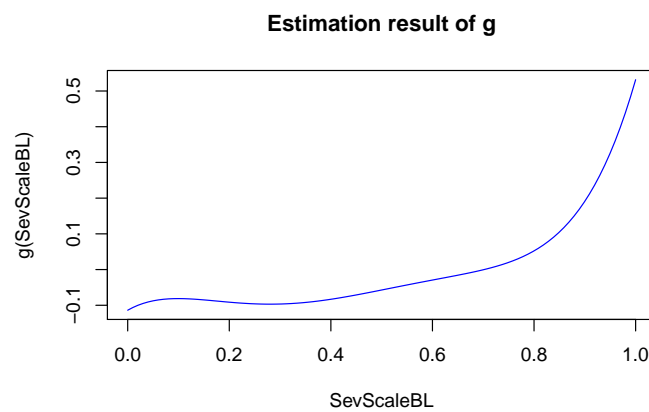


Figure 3. Estimated nonlinear effect of SevScaleBL for the illustration.

6. Concluding Remarks

In the preceding sections, the regression analysis of bivariate interval-censored survival data was considered under a family of copula-based, semiparametric, partly linear additive hazards models. As discussed above, one significant advantage of the models was that they only needed to handle the two marginal distributions via a copula function with the copula parameter determining the dependence. For inference, a sieve maximum likelihood estimation procedure with the use of Bernstein polynomials were provided, and it was shown that the resulting estimators of the regression parameters were consistent and asymptotically efficient. Furthermore, the simulation studies suggested that the recommended approach worked effectively in practical situations.

It is worth emphasizing that the main reason for using Bernstein polynomials to approximate the infinite-dimensional cumulative hazard function $\Lambda_j(t)$ and nonlinear covariate effects $g(\cdot)$ was its simplicity. Other smoothing functions, such as spline functions [20], could also be used, and estimation procedures similar to the one described above could be developed. In order to implement the approach, we needed to choose the degrees m_1 and m_2 . As discussed above, a common method is to consider different values and compare the resulting estimators. Certainly, developing a data-driven approach for their selections would also be helpful.

In this paper, we focused on the additive hazards model. Other models, for example, the proportional hazards model, are sometimes more popular. It is well-known that the latter is more suitable for the situation where one is interested in the hazard ratio, whereas the former fits better if the excess risk or the risk differential is what is most important. However, there seems to be little literature on how to choose the better model or develop some model-checking methods, which is attractive research for the future.

Author Contributions: Conceptualization, T.H. and S.Z.; methodology, X.Z. and T.H.; software, X.Z.; validation, X.Z. and T.H.; resources, T.H. and S.Z.; writing—original draft preparation, X.Z.; writing—review and editing, T.H. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Beijing Natural Science Foundation Z210003, Technology Developing Plan of Jilin Province (No. 20200201258JC) and National Nature Science Foundation of China (grant nos. 12171328, 12071176 and 11971064).

Data Availability Statement: Research data are available in the R package CopulaCenR.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike information criterion
AREDS	Age-related Eye Disease Study
AMD	Age-related macular degeneration
a.s.	almost surely

Appendix A. Proof of Asymptotic Properties

This appendix begins by describing the necessary regular conditions, which are similar to those generally used in the interval-censored data literature [32,33], and then sketch the proof of the asymptotic results given in Theorems 1–3.

Condition 1. There exists a positive number δ_j such that $P(V_j - U_j \geq \delta_j) = 1, j = 1, 2$.

Condition 2. (i) There exist $0 < c_1 < d_1 < \infty$ such that $P(c_1 \leq U_{ij} < V_{ij} \leq d_1) = 1, j = 1, 2$. (ii) The covariate X_j 's are bounded; in other words, there exists $x_0 > 0$ that makes $P(|X| \leq x_0) = 1$, where $j = 1, 2$. The distribution of the X_j 's is not focused on any proper affine subspace of R^p . (iii) For some positive constant K , given the set of all observation times \mathcal{V} , suppose the conditional density of Z_j is twice continuously differentiable and has a bound within $[K^{-1}, K]$ on $[c_2, d_2]$ a.s.

Condition 3. After substituting Λ_j with h_j , the likelihood function can be rewritten as $L(\beta, \varphi, \omega, h_1, h_2, g)$. Define

$$\mu^T \dot{L}(\beta, \varphi, \omega, h_1, h_2, g) = \mu_1^T \frac{\partial L}{\partial \beta} + \mu_2 \frac{\partial L}{\partial \varphi} + \mu_3 \frac{\partial L}{\partial \omega} + \mu_4 \frac{\partial L}{\partial h_1} + \mu_5 \frac{\partial L}{\partial h_2} + \mu_6 \frac{\partial L}{\partial g}$$

with $\mu = (\mu_1^T, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)^T$. There exist $l_j^*, r_j^* \in [c_1, d_1]$ where there exist $p + 5$ various sets of (X_1, X_2) so that when $\mu^T \dot{L}(\beta_0, \varphi_0, \omega_0, \Lambda_{10}, \Lambda_{20}, g_0; O^*) = 0$ in which $O^* = \{u_j^*, v_j^*, X_j, Z_j\}$ for every $p + 5$ sets of values, one can conclude $v = 0_{(p+5) \times 1}$.

Condition 4. There exist $0 < m_1 < m_2 < \infty$ satisfying $m_1 < \Lambda_{j0}(c_1) < \Lambda_{j0}(d_1) < m_2$, and Λ_{j0} is strictly increasing and continuously differentiable until a q -order within $[c_1, d_1]$, $j = 1, 2$. In addition, g_0 is continuously differentiable until an r -order within $[c_2, d_2]$. Additionally, $(\beta^{(0)T}, \varphi_0, \omega_0)^T$ denotes an interior point in $\mathcal{A} \subseteq \mathbb{R}^p \times \mathbb{R}^{(0,1]} \times \mathbb{R}^+$.

Condition 5. There exists $\varepsilon > 0$, for every $|\eta - \eta_0| < \varepsilon$, $P\{\ell_n(\eta; O) - \ell_n(\eta_0; O)\} \preceq -d^2(\eta, \eta_0)$, in which $\ell_n(\eta; O)$ is the log-likelihood function acquired before, and \preceq denotes that “the left-hand side is smaller than the right, until a constant time”.

It is worth noting that Conditions 1–5, except Condition 3, are usually employed in interval-censored failure time research [26,34,35]. Condition 3 helps to ensure the parameters’ identifiability and ensure the effective Fisher information matrix is always positive [26,35].

Proof of Theorem 1. Define $\ell(\eta; O_i)$ as the log-likelihood for only one observation O_i , and $\Theta_n = \mathcal{A} \otimes \mathcal{M}_n \otimes \mathcal{M}_n \otimes \mathcal{H}_n$. Let $\mathcal{L} = \{\ell(\eta; O_i) : \eta \in \Theta\}$ denote a class of functions, and let P_n and P denote the empirical and true probability measures, respectively. Then, following a similar calculation as in Lemma 1 in [19], the bracketing number of \mathcal{L} is (until a constant

time) in a bound of $(1/\epsilon)^{2m_1+m_2+p+2}$. Therefore, following Theorem 2.4.1 in [36], we can get \mathcal{L} is a Glivenko–Cantelli class. Thus,

$$\sup_{\eta \in \Theta_n} |(P_n - P)\ell(\eta; O)| \rightarrow 0 \text{ a.s.}$$

Let $H(\eta; O) = -\ell(\eta; O)$, for any $\epsilon > 0$, define $K_\epsilon = \{\eta : d(\eta, \eta_0) \geq \epsilon, \eta \in \Theta_n\}$, $\kappa_{1n} = \sup_{\eta \in \Theta_n} |(P_n - P)H(\eta; O)|$, and $\kappa_{2n} = P_n H(\eta_0; O) - PH(\eta_0; O)$. Hence, we can conclude that

$$\inf_{K_\epsilon} PH(\eta; O) = \inf_{K_\epsilon} \{PH(\eta; O) - P_n H(\eta; O) + P_n H(\eta; O)\} \leq \kappa_{1n} + \inf_{K_\epsilon} P_n H(\eta; O). \tag{A1}$$

If $\hat{\eta}_n \in K_\epsilon$, we can get

$$\inf_{K_\epsilon} P_n H(\eta; O) = P_n H(\hat{\eta}_n; O) \leq P_n H(\eta_0; O) = \kappa_{2n} + PH(\eta_0; O). \tag{A2}$$

Then, we can prove $\delta_\epsilon > 0$ based on the proof by contradiction under Condition 3. Combining (A1) and (A2), we have

$$\inf_{K_\epsilon} PH(\eta; O) \leq \kappa_{1n} + \kappa_{2n} + PH(\eta_0; O) = \kappa_n + PH(\eta_0; O),$$

in which $\kappa_n = \kappa_{1n} + \kappa_{2n}$, so that $\kappa_n \geq \delta_\epsilon$, where $\delta_\epsilon = \inf_{K_\epsilon} PH(\eta; O) - PH(\eta_0; O)$. This gives $\{\hat{\eta}_n \in K_\epsilon\} \subseteq \{\kappa_n \geq \delta_\epsilon\}$. Based on Condition 1 and together with the strong law of large numbers, we have $\kappa_{1n} + \kappa_{2n} \rightarrow 0$ a.s. Hence, $\cup_{k=1}^\infty \cap_{n=k}^\infty \{\hat{\eta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\kappa_n \geq \delta_\epsilon\}$, which proves that $d(\hat{\eta}_n, \eta_0) \rightarrow 0$ a.s. The proof of Theorem 1 is complete. \square

Proof of Theorem 2. We verify Conditions C1–C3 in [37] to derive the rate of convergence. Define $\|u\|$ as the Euclidean norm of a vector u , $\|h\|_\infty = \sup_x |h(x)|$ is the supremum norm of a function h , and $\|h\|_{L_2(P)} = (\int |h|^2 dP)^{1/2}$. Moreover, let P denote a probability measure. After that, for the convenience of understanding the proof as follows, we define $\Theta^{qr} = \mathcal{A} \otimes \mathcal{M}^q \otimes \mathcal{M}^q \otimes \mathcal{G}^r$, where \mathcal{M}^q is the sets of Λ_j , and \mathcal{G}^r is the sets of g . q and r are as defined in Condition 4. It is noteworthy that Θ^{qr} is completely the same as Θ except for the notation. Similarly, Θ_n^{qr} is the corresponding sieve space containing \mathcal{M}_n^q and \mathcal{G}_n^r . First, as a result of Condition 5, we can verify that Condition C1 directly stands. That is, $P\{\ell(\eta_0; O) - \ell(\eta; O)\} \preceq -d^2(\eta, \eta_0)$ for any $\eta \in \Theta_n^{qr}$. Second, we verify Condition C2 in [37]. Based on Conditions 1–4, one can easily find that for every $\eta \in \Theta_n^{qr}$,

$$\begin{aligned} P\{\ell(\eta; O) - \ell(\eta_0; O)\}^2 &\preceq |b - b_0|^2 + P\left[\{\Lambda_1(l_1) - \Lambda_{10}(l_1)\}^2 + \{\Lambda_1(r_1) - \Lambda_{10}(r_1)\}^2\right] \\ &\quad + P\left[\{\Lambda_2(l_2) - \Lambda_{20}(l_2)\}^2 + \{\Lambda_2(r_2) - \Lambda_{20}(r_2)\}^2\right] \\ &\quad + P\{g(Z_1) - g_0(Z_1)\}^2 + P\{g(Z_2) - g_0(Z_2)\}^2 = d^2(\eta, \eta_0), \end{aligned}$$

in which $b = (\beta^T, \varphi, \omega)^T$, and this implies that for any $\eta \in \Theta_n^{qr}$, $\sup_{d(\eta, \eta_0) \leq \epsilon} \text{var}\{\ell(\eta_0; O) - \ell(\eta; O)\} \leq \sup_{d(\eta, \eta_0) \leq \epsilon} P\{\ell(\eta_0; O) - \ell(\eta; O)\}^2 \preceq \epsilon^2$. Thus, Condition C2 from [37] holds when the sign β in their paper is equal to one. In the end, one needs to verify Condition C3 of [37]. Define the class of functions $\mathcal{N}_n = \{\ell(\eta; O) - \ell(\eta_{n0}; O) : \eta \in \Theta_n^{qr}\}$ and let $N_{[\cdot]}(\epsilon, \mathcal{N}_n, L_\infty)$ denote the ϵ -bracketing number related to L_∞ norm of \mathcal{N}_n . Then, we have $N_{[\cdot]}(\epsilon, \mathcal{N}_n, L_\infty) \preceq (1/\epsilon)^{c_1 m_1 + c_2 m_2 + p + 2}$ by following similar arguments as in Lemma A3 of [13], where $k_1, k_2 > 0$, and $p + 2$ is the dimensionality of b . By following the fact that the covering number is always smaller than the bracketing number, we have $\log N_{[\cdot]}(\epsilon, \mathcal{N}_n, L_\infty) \leq (k_1 m_1 + k_2 m_2 + p + 2) \log(1/\epsilon) \preceq n^{\nu_1 + \nu_2} \log(1/\epsilon)$. Therefore, Condition C3 in [37] is satisfied under $2r_0 = \nu$, and $r = 0^+$ in their sign. Hence, τ of Theorem 1 in [37] on page 584 may be equal to $(1 - \nu_1)/2 - \{\log(\log n)\}/(2 \log n)$. Because the part behind the minus is close to zero with $n \rightarrow 0$, one can set a $\tilde{\nu}_1$ a little bigger than ν_1 so as

to get $(1 - \tilde{v}_1)/2 \leq \tau$ with a large n . Let \tilde{v}_1 replace v_1 but keep the same notation with v_1 , then the new constant $\tau' = (1 - v_1)/2$.

Notice that from Theorem 1.6.2 in [38], there are Bernstein polynomials $\Lambda_{jn0} \in \mathcal{M}_n^q$ that make $\|\Lambda_{jn0} - \Lambda_{j0}\|_\infty = O(n^{-q/2})$, $j = 1, 2$. Similarly, there also exists a function $g_{n0} \in \mathcal{G}_n^r$ satisfying $\|g_{n0} - g_0\|_\infty = O(n^{-r/2})$. Then, the sieve approximate error $\rho(\pi_n \eta_0, \eta_0)$ in [37] is $O(n^{-qv_1/2})$. Therefore, applying the Taylor expansion to $P\{\ell(\eta_0; O) - \ell(\eta; O)\}$ surrounding η_0 , then plugging in $\eta_{n0} = (\beta_0^T, \varphi_0, \omega_0, \Lambda_{1n0}, \Lambda_{2n0}, g_{n0})^T$, the Kullback–Leibler pseudodistance of η_0 and η_{n0} follows

$$\begin{aligned} K(\eta_{n0}, \eta_0) &= -P\{\ell(\eta_{n0}; O) - \ell(\eta^{(0)}; O)\} \\ &= -\frac{1}{2}P\{\ddot{\ell}_{\Lambda_1 \Lambda_1}(\eta_0; O)[\Lambda_{1n0} - \Lambda_{10}, \Lambda_{1n0} - \Lambda_{10}] + \ddot{\ell}_{\Lambda_2 \Lambda_2}(\eta_0; O)[\Lambda_{2n0} - \Lambda_{20}, \Lambda_{2n0} - \Lambda_{20}] \\ &\quad + 2\ddot{\ell}_{\Lambda_1 \Lambda_2}(\eta_0; O)[\Lambda_{1n0} - \Lambda_{10}, \Lambda_{2n0} - \Lambda_{20}]\} - P\{\ddot{\ell}_{gg}(\eta_0; O)[g_{n0} - g_0, g_{n0} - g_0]\} \\ &\quad + o(d^2(\eta_{n0}, \eta_0)) \\ &\leq \|\Lambda_{1n0} - \Lambda_{10}\|_2^2 + \|\Lambda_{2n0} - \Lambda_{20}\|_2^2 + o(\|\Lambda_{1n0} - \Lambda_{10}\|_2^2 + \|\Lambda_{2n0} - \Lambda_{20}\|_2^2) \\ &\quad + \|g_{n0} - g_0\|_2^2 + o(\|g_{n0} - g_0\|_2^2) \\ &\leq O(n^{-qv_1}) + O(n^{-rv_2}) = O(n^{-\min(qv_1, rv_2)}). \end{aligned}$$

The first equality holds due to the first derivative of $P\ell(\eta; O)$ at η_0 being equal to zero. As for the penultimate inequality, it holds because all the derivatives and second-order derivatives of the log-likelihood are bounded. Furthermore, since $\|\Lambda_{jn0} - \Lambda_{j0}\|_2 \leq \|\Lambda_{jn0} - \Lambda_{j0}\|_\infty = O(n^{-qv_1/2})$, and $\|g_{n0} - g_0\|_2 \leq \|g_{n0} - g_0\|_\infty = O(n^{-rv_2/2})$, we can get the last inequality, so that $K^{1/2}(\eta_0, \eta_{n0}) = O(n^{-qv_1/2}) + O(n^{-rv_2/2}) = O(n^{-\min(qv_1/2, rv_2/2)})$. Hence, by Theorem 1 in [37], the convergence rate of $\hat{\eta}_n$ is

$$\begin{aligned} d(\hat{\eta}_n, \eta_0) &= O_p\left\{\max\left(n^{-(1-v_1)/2}, n^{-qv_1/2}, n^{-(1-v_2)/2}, n^{-rv_2/2}\right)\right\} \\ &= O_p\left(n^{-\min\{qv_1/2, (1-v_1)/2, rv_2/2, (1-v_2)/2\}}\right). \end{aligned}$$

The proof of Theorem 2 is complete. \square

Proof of Theorem 3. Let us sketch the proof of Theorem 3 in five steps as follows.

Step 1. We first calculate the derivatives regarding $\eta = (\beta^T, \varphi, \omega, \Lambda_1, \Lambda_2, g)^T$, such that $\dot{\ell}_{\Lambda_j}(\eta; O)[h_{\Lambda_j}]$, $\dot{\ell}_g(\eta; O)[h_g]$ and so on; now, we omit $(\eta; O)$ in the following formula for convenience in Step 1.

To obtain the score functions of $\Lambda_j, j = 1, 2$. Let $y_{\Lambda_j}(t_j) \in H_{\Lambda_j} = \{y_{\Lambda_j} : y_{\Lambda_j} = \frac{\partial \Lambda_{j\tilde{\xi}}}{\partial \tilde{\xi}}|_{\tilde{\xi}=0}, \Lambda_{j\tilde{\xi}} \in \mathcal{M}^q\}$ denote an arbitrary parametric submodel of Λ_j , in which $y_{\Lambda_j}(t_j)$ satisfies the Fréchet derivative $\lim_{\tilde{\xi} \rightarrow 0} \{\Lambda_j(t_j + \tilde{\xi}) - \Lambda_j(t_j) - y_{\Lambda_j}(t_j)\tilde{\xi}\} / \tilde{\xi} = 0$. Similarly, we can also define a submodel of g noted by $y_g(t_j) \in H_g$. Moreover, note

$$\begin{aligned} \ell(\eta; t_1, t_2, Y) &= \log S(t_1, t_2 | Y) \\ &= -\omega \log \left\{ 1 + \left[\left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1 \right)^{\frac{1}{\varphi}} + \left(\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1 \right)^{\frac{1}{\varphi}} \right]^\varphi \right\}, \end{aligned}$$

and

$$J = 1 + \left[\left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1 \right)^{1/\varphi} + \left(\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1 \right)^{1/\varphi} \right]^\varphi,$$

where $Y = (X, Z)$ and $h(t; Y) = \beta^T W(t) + g(Z)t$. The score function along $y_{\Lambda_j}(t_j)$ is

$$\dot{\ell}_{\Lambda_j}(\eta; t_1, t_2, Y)[y_{\Lambda_j}] = \frac{-1}{J} \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_{j'}(t_{j'}) + h(t_{j'}; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_j(t_j) + h(t_j; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_j(t_j) + h(t_j; Y))\right) \times y_{\Lambda_j}(t_j),$$

with $j, j' \in \{1, 2\}$, $h_{\Lambda_j}(t_j) \in \mathcal{M}^{q-1}$ and $\Lambda_j \in \mathcal{M}^q$. Analogously, we have the derivatives with respect to g as

$$\begin{aligned} \dot{\ell}_g(\eta; t_1, t_2, Y) = & \frac{-1}{J} \left\{ \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) t_1 \right. \\ & \left. + \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) t_2 \right\} \times y_g, \end{aligned}$$

with $g(z) \in \mathcal{G}^r$, and $y_g(t_j) \in \mathcal{G}^{r-1}$.

The second-order derivatives of $\ell(\eta; t_1, t_2, Y)$ have the form

$$\begin{aligned} \ddot{\ell}_{\Lambda_1 \Lambda_1}(\eta; t_1, t_2, Y)[y_{\Lambda_1}, \tilde{y}_{\Lambda_1}] &= \left\{ \frac{-1}{J} \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) \right. \\ &+ \frac{1}{J} \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-2} \exp\left(\frac{2}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) \left(1 - \frac{1}{\varphi}\right) \\ &\times \left(\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1 \right)^{1/\varphi} \left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1 \right)^{-1-1/\varphi} \\ &\left. + \frac{1}{J^2} \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{2\varphi-2} \exp\left(\frac{2}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) \right\} \\ &\times \frac{1}{\omega} y_{\Lambda_1}(t_1) \tilde{y}_{\Lambda_1}(t_1), \end{aligned}$$

$$\begin{aligned} \ddot{\ell}_{\Lambda_1 \Lambda_2}(\eta; t_1, t_2, Y)[y_{\Lambda_1}, h_{\Lambda_2}] &= \ddot{\ell}_{\Lambda_2 \Lambda_1}(\eta; t_1, t_2, Y)[h_{\Lambda_2}, h_{\Lambda_1}] \\ &= \left\{ \frac{-1}{J} \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-2} \frac{1}{\left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1 \right)} \right. \\ &\times \left(1 - \frac{1}{\varphi}\right) + \frac{1}{J^2} \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{2\varphi-2} \left. \right\} \\ &\times \frac{1}{\omega} \exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y) + \Lambda_2(t_2) + h(t_2; Y))\right) y_{\Lambda_1}(t_1) y_{\Lambda_2}(t_2) \\ &\times \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi-1}. \end{aligned}$$

$$\begin{aligned}
 \ddot{\ell}_{\Lambda_1 g}(\eta; t_1, t_2, Y)[y_{\Lambda_1}, y_g] = & \\
 & \left\{ \frac{1}{j} \left[\left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y)) - 1\right) \right)^{1/\varphi} + \left(\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y)) - 1\right) \right)^{1/\varphi} \right]^{\varphi-1} \right. \\
 & \times \left[\left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y)) - 1\right) \right)^{1/\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y)) \times t_1 \right. \right. \\
 & \left. \left. + \left(\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y)) - 1\right) \right)^{1/\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y)) \times t_2 \right) \right] \right. \\
 & - \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{-1} \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \\
 & \times \left[\left(\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y)) - 1\right) \right) \exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y)) \times t_2 \right) \right. \\
 & \left. + \left(\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y)) - 1\right) \right) \exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y)) \times t_1 \right) \right] \left(1 - \frac{1}{\varphi} \right) \left. \right\} \\
 & \times \left[1 + \left(\frac{\exp\left(\frac{1}{\omega}(\Lambda_2(t_2) + h(t_2; Y))\right) - 1}{\exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) - 1} \right)^{1/\varphi} \right]^{\varphi-1} \exp\left(\frac{1}{\omega}(\Lambda_1(t_1) + h(t_1; Y))\right) \\
 & \times \frac{1}{j} \frac{1}{\omega} y_{\Lambda_1}(t_1) y_g,
 \end{aligned}$$

Similarly, we can derive $\ddot{\ell}_{bb}$, $\ddot{\ell}_{\Lambda_j b}[y_{\Lambda_j}]$ and $\ddot{\ell}_{gb}[y_g]$ as, respectively, the derivatives of $\dot{\ell}_b$, $\dot{\ell}_{\Lambda_j}[y_{\Lambda_j}]$ and $\dot{\ell}_g[y_g]$ with respect to b .

$\ddot{\ell}_{b\Lambda_j}[y_{\Lambda_j}]$, $\ddot{\ell}_{\Lambda_j \Lambda_{j'}}[y_{\Lambda_j}, y_{\Lambda_{j'}}]$, $\ddot{\ell}_{\Lambda_j \Lambda_{j'}}[y_{\Lambda_j}, y_{\Lambda_{j'}}]$ and $\ddot{\ell}_{g\Lambda_j}[y_g, y_{\Lambda_j}]$ are, respectively, the derivatives of $\dot{\ell}_b$, $\dot{\ell}_{\Lambda_j}$, $\dot{\ell}_{\Lambda_j}$ and $\dot{\ell}_g[y_g]$ with respect to Λ_j , $j, j' \in \{1, 2\}$.

$\ddot{\ell}_{bg}[y_g]$, $\ddot{\ell}_{\Lambda_j g}[y_{\Lambda_j}, y_g]$ and $\ddot{\ell}_{gg}[y_g, y_g]$ are, respectively, the derivatives of $\dot{\ell}_b$, $\dot{\ell}_{\Lambda_j}[y_{\Lambda_j}]$ and $\dot{\ell}_g[y_g]$ with respect to g .

Step 2. Consider the classes of functions $\mathcal{D}_1 = \{\dot{\ell}_b(\eta) : d(\eta, \eta_0) \leq \delta\}$, $\mathcal{D}_2 = \{\dot{\ell}_{\Lambda_j}(\eta)[y_{\Lambda_j}] : d(\eta, \eta_0) \leq \delta\}$ and $\mathcal{D}_3 = \{\dot{\ell}_g(\eta)[y_g] : d(\eta, \eta_0) \leq \delta\}$. We need to show these three function classes are Donsker for any $\delta > 0$. We determine the bracketing number of \mathcal{D}_1 in order to demonstrate that it is Donsker. In accordance with [37], we have $\log N_{[\cdot]}(\epsilon, \mathcal{D}_1, L_2) \preceq \max(m_1, m_2) \log(\delta/\epsilon)$ for $0 < \epsilon < \delta$. This results in a finite-valued bracketing integral according to Theorem 2.8.4 of [36]. Hence, the class \mathcal{D}_1 is Donsker. Similar justifications support that \mathcal{G}_2 and \mathcal{G}_3 are also Donsker.

Step 3. Following similar arguments as in Lemma 2 of [19] and the properties of the score statistic, there exist $y_{\Lambda_j}^* \in \mathcal{M}^{q-1}$ and $y_g^* \in \mathcal{G}^{r-1}$ satisfying

$$P\dot{\ell}_b(\eta_0) = 0, P\dot{\ell}_{\Lambda_j}(\eta_0)[y_{\Lambda_j}^*] = 0, P\dot{\ell}_g(\eta_0)[y_g^*] = 0. \tag{A3}$$

Let $\hat{\eta}_n$ denote the estimators of the sieve log-likelihood and $y_{\Lambda_j, n}^*$ is the projection of $y_{\Lambda_j}^*$ onto \mathcal{M}_n , $j = 1, 2$. We get

$$\begin{aligned}
 P_n \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j}^*] = & P_n \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j, n}^*] + P \dot{\ell}_{\Lambda_j}(\eta_0)[y_{\Lambda_j}^* - y_{\Lambda_j, n}^*] + (P_n - P) \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j}^* - y_{\Lambda_j, n}^*] \\
 & + P \left\{ \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j}^* - y_{\Lambda_j, n}^*] - \dot{\ell}_{\Lambda_j}(\eta_0)[y_{\Lambda_j}^* - y_{\Lambda_j, n}^*] \right\} \\
 = & (I) + (II) + (III) + (IV).
 \end{aligned} \tag{A4}$$

Following the discussion about the proof for Theorem 5.3 of [34], we can derive that part (I) is equal to $o_p(n^{-1/2})$. In addition, (II) is also equal to $o_p(n^{-1/2})$ based on

(A3). We can acquire (III) as $o_p(n^{-1/2})$ due to \mathcal{D}_2 being Donsker. As for the fourth term (IV), on account of Theorem 2 and employing the first-order linear expansion of $\dot{\ell}_{\Lambda_j}(\hat{\eta}_n)$ around η_0 , one can get (IV) is $o_p(n^{-1/2})$ as well. Summating the four terms, we have $P_n \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j}^*] = o_p(n^{-1/2})$. Likewise, we have the property of $P_n \dot{\ell}_g(\hat{\eta}_n)[y_g^*]$. Hence, we have

$$P_n \dot{\ell}_b(\hat{\eta}_n) = o_p(n^{-1/2}), P_n \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j}^*] = o_p(n^{-1/2}), P_n \dot{\ell}_g(\hat{\eta}_n)[y_g^*] = o_p(n^{-1/2}). \tag{A5}$$

Step 4. Combining (A3) and (A5), we can easily show that

$$\begin{aligned} P_n \{ \dot{\ell}_b(\hat{\eta}_n) - \dot{\ell}_b(\eta_0) \} &= -(P_n - P) \dot{\ell}_b(\eta_0) + o_p(n^{-1/2}), \\ P_n \{ \dot{\ell}_{\Lambda_j}(\hat{\eta}_n)[y_{\Lambda_j}^*] - \dot{\ell}_{\Lambda_j}(\eta_0)[y_{\Lambda_j}^*] \} &= -(P_n - P) \dot{\ell}_{\Lambda_j}(\eta_0)[y_{\Lambda_j}^*] + o_p(n^{-1/2}), \\ P_n \{ \dot{\ell}_g(\hat{\eta}_n)[y_g^*] - \dot{\ell}_g(\eta_0)[y_g^*] \} &= -(P_n - P) \dot{\ell}_g(\eta_0)[y_g^*] + o_p(n^{-1/2}). \end{aligned} \tag{A6}$$

Furthermore, based on some arguments in the proof of Theorem 3.2 in [13], there exists a neighborhood of $(b_0, \Lambda_{10}, \Lambda_{20}, g_0)$ as $\{(b, \Lambda_1, \Lambda_2, g) : |b - b_0| + \|\Lambda_1 - \Lambda_{10}\|_2 + \|\Lambda_2 - \Lambda_{20}\|_2 + \|g - g_0\|_2 \leq Cn^{-\xi}\}$, where $\xi = \min\{(1 - \nu_1)/2, q\nu_1/2, (1 - \nu_2)/2, r\nu_2/2\}$. Then, applying the Taylor expansion for $\ell_{\Lambda_j}(\eta)[y_j^*]$ yields

$$\begin{aligned} P(\ell_{\Lambda_j}(\eta)[y_j^*] - \ell_{\Lambda_j}(\eta_0)[y_j^*] - \ddot{\ell}_{\Lambda_j b}(\eta_0)[h_j^*](b - b_0) - \ddot{\ell}_{\Lambda_j \Lambda_j}(\eta_0)[y_j^*, \Lambda_j - \Lambda_{j0}] \\ - \ddot{\ell}_{\Lambda_j \Lambda_{j'}}(\eta_0)[y_j^*, \Lambda_{j'} - \Lambda_{j'0}] - \ddot{\ell}_{\Lambda_j g}(\eta_0)[y_j^*, g - g_0]) = o_p(n^{-1/2}), \end{aligned} \tag{A7}$$

where $j, j' \in \{1, 2\}$. Likewise, it is also easy to get the property of $\ell_b(\eta)$ and $\ell_g(\eta)[y_g^*]$. Note that the derivatives of the score statistics are bounded. After applying Taylor series expansions about η_0 to (A6), and combining the Equations (A7), we have

$$\begin{aligned} P \dot{\ell}_{bb}(\eta_0)(\hat{b}_n - b_0) + P \ddot{\ell}_{b\Lambda_1}(\eta_0)[\hat{\Lambda}_{1,n} - \Lambda_{10}] + P \ddot{\ell}_{b\Lambda_2}(\eta_0)[\hat{\Lambda}_{2,n} - \Lambda_{20}] + P \dot{\ell}_{bg}(\eta_0)[\hat{g}_n - g_0] \\ = -P_n \dot{\ell}_b(\eta_0) + o_p(n^{-1/2}), \\ P \ddot{\ell}_{\Lambda_j b}(\eta_0)[y_j^*](\hat{b}_n - b_0) + P \ddot{\ell}_{\Lambda_j \Lambda_j}(\eta_0)[y_j^*, \hat{\Lambda}_{j,n} - \Lambda_{j0}] + P \ddot{\ell}_{\Lambda_j \Lambda_{j'}}(\eta_0)[y_j^*, \hat{\Lambda}_{j',n} - \Lambda_{j'0}] \\ + P \ddot{\ell}_{\Lambda_j g}(\eta_0)[y_j^*, \hat{g}_n - g_0] = -P_n \dot{\ell}_{\Lambda_j}(\eta_0)[y_j^*] + o_p(n^{-1/2}), \\ P \ddot{\ell}_{gb}(\eta_0)[y_g^*](\hat{b}_n - b_0) + P \ddot{\ell}_{g\Lambda_1}(\eta_0)[y_g^*, \hat{\Lambda}_{1,n} - \Lambda_{10}] + P \ddot{\ell}_{g\Lambda_2}(\eta_0)[y_g^*, \hat{\Lambda}_{2,n} - \Lambda_{20}] + P \dot{\ell}_{gg}(\eta_0)[y_g^*, \hat{g}_n - g_0] \\ = -P_n \dot{\ell}_g(\eta_0)[y_g^*] + o_p(n^{-1/2}). \end{aligned} \tag{A8}$$

Taking the first equality in (A8) and subtracting the second and third equalities, we have

$$\begin{aligned} P(\ddot{\ell}_{bb}(\eta_0) - \ddot{\ell}_{\Lambda_1 b}(\eta_0)[h_{\Lambda_1}^*] - \ddot{\ell}_{\Lambda_2 b}(\eta_0)[h_{\Lambda_2}^*] - \ddot{\ell}_{gb}(\eta_0)[h_g^*])(\hat{b}_n - b_0) = -P_n \{ \dot{\ell}_b(\eta_0) \\ - \dot{\ell}_{\Lambda_1}(\eta_0)[y_{\Lambda_1}^*] - \dot{\ell}_{\Lambda_2}(\eta_0)[y_{\Lambda_2}^*] - \dot{\ell}_g(\eta_0)[y_g^*] \} + o_p(n^{-1/2}). \end{aligned} \tag{A9}$$

Step 5. Define $Q = -P(\ddot{\ell}_{bb}(\eta_0; O) - \ddot{\ell}_{\Lambda_1 b}(\eta_0; O)[y_{\Lambda_1}^*] - \ddot{\ell}_{\Lambda_2 b}(\eta_0; O)[y_{\Lambda_2}^*] - \ddot{\ell}_{gb}(\eta_0; O)[y_g^*])$ and $B = P\ell^*(\eta_0; O)^{\otimes 2} = P\{\ell^*(\eta_0; O)\ell^*(\eta_0; O)^T\}$; then, we have

$$\begin{aligned} Q &= P(\dot{\ell}_b(\eta_0; O) - \dot{\ell}_{\Lambda_1}(\eta_0; O)[y_{\Lambda_1}^*] - \dot{\ell}_{\Lambda_2}(\eta_0; O)[y_{\Lambda_2}^*] - \dot{\ell}_g(\eta_0)[y_g^*])^{\otimes 2} \\ &= P\ell^*(\eta_0; O)^{\otimes 2} = B, \end{aligned}$$

where $\ell^*(\eta_0; O) = \dot{\ell}_b(\eta_0; O) - \dot{\ell}_{\Lambda_1}(\eta_0; O)[y_{\Lambda_1}^*] - \dot{\ell}_{\Lambda_2}(\eta_0; O)[y_{\Lambda_2}^*] - \dot{\ell}_g(\eta_0; O)[y_g^*]$. Next, we need to verify Q is nonsingular. If Q is a nonsingular matrix, then we can conclude

$v = (v_1^T, v_2, v_3)^T = 0$ from $v^T Q v = v^T P \ell^*(\eta_0; O) \otimes^2 v = 0$. Moreover, one is enough to show if $v^T \ell^*(\eta_0; O) = 0$, then $v = 0$. Thus, we have

$$\begin{aligned}
 v^T \ell^*(\eta_0; O) &= v^T \left\{ \dot{\ell}_b(\eta_0; O) - \dot{\ell}_{\Lambda_1}(\eta_0; O)[y_{\Lambda_1}^*] - \dot{\ell}_{\Lambda_2}(\eta_0; O)[y_{\Lambda_2}^*] - \dot{\ell}_g(\eta_0; O)[y_g^*] \right\} \\
 &= \left\{ v_1^T \frac{\partial L}{\partial \beta}(\eta_0; O) + v_2 \frac{\partial L}{\partial \varphi}(\eta_0; O) + v_3 \frac{\partial L}{\partial \omega}(\eta_0; O) + \frac{\partial L}{\partial \Lambda_1}(\eta_0; O)[-v^T y_{\Lambda_1}^*] \right. \\
 &\quad \left. + \frac{\partial L}{\partial \Lambda_2}(\eta_0; O)[-v^T y_{\Lambda_2}^*] + \frac{\partial L}{\partial g}(\eta_0; O)[-v^T y_g^*] \right\} \frac{1}{L(\eta_0; O)}, \tag{A10}
 \end{aligned}$$

where $\eta_0 = (\beta^{(0)T}, \varphi_0, \omega_0, \Lambda_{10}, \Lambda_{20}, g_0)$ and $L(\eta_0; O)$ is the likelihood function. Under our Condition 3, (A10) is equal to zero only if $v = 0$. As a consequence, we have verified Q is nonsingular.

Substitute Q into (A9), we get

$$-Q(\hat{b}_n - b_0) = -P_n l^*(\eta_0; O) + o_p(n^{-1/2}).$$

This implies that

$$\sqrt{n}(\hat{b}_n - b_0) = Q^{-1} n^{1/2} P_n l^*(\eta_0; O) + o_p(1) \rightarrow_d N\{0, Q^{-1} B(Q^{-1})^T\}. \tag{A11}$$

Since $Q = B = P \ell^*(\eta_0) \otimes^2$, we obtain $Q^{-1} B(Q^{-1})^T = Q^{-1} \triangleq I^{-1}(b_0)$. Thus, $n^{1/2}(\hat{b}_n - b_0) \rightarrow_d N\{0, I^{-1}(b_0)\}$, with $I(b_0) = P \ell^*(\eta_0) \otimes^2$ and $\ell^*(\eta_0)$ being the efficient score function of b_0 . Now, we complete the proof of Theorem 3. \square

References

1. Sun, J. *The Statistical Analysis of Interval-Censored Failure Time Data*; Springer: Berlin/Heidelberg, Germany, 2006.
2. Lin, D.Y.; Oakes, D.; Ying, Z. Additive hazards regression with current status data. *Biometrika* **1998**, *85*, 289–298. [\[CrossRef\]](#)
3. Martinussen, T.; Scheike, T.H. Efficient estimation in additive hazards regression with current status data. *Biometrika* **2002**, *89*, 649–658. [\[CrossRef\]](#)
4. Feng, Y.; Sun, J.; Sun, L. Estimation of the additive hazards model with linear inequality restrictions based on current status data. *Commun. Stat.-Theory Methods* **2022**, *51*, 68–81. [\[CrossRef\]](#)
5. Wang, P.; Zhou, Y.; Sun, J. A new method for regression analysis of interval-censored data with the additive hazards model. *J. Korean Stat. Soc.* **2020**, *49*, 1131–1147. [\[CrossRef\]](#)
6. Li, H.; Zhang, H.; Zhu, L.; Li, N.; Sun, J. Estimation of the additive hazards model with interval-censored data and missing covariates. *Can. J. Stat.* **2020**, *48*, 499–517. [\[CrossRef\]](#)
7. Wang, T.B.; Yopadhyay, D.; Sinha, S. Efficient estimation of the additive risks model for interval-censored data. *arXiv* **2022**, arXiv:2203.09726.
8. Tong, X.; Chen, M.H.; Sun, J. Regression analysis of multivariate interval-censored failure time data with application to tumorigenicity experiments. *Biom. J.* **2008**, *50*, 364–374. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Yin, G.; Cai, J. Additive hazards model with multivariate failure time data. *Biometrika* **2004**, *91*, 801–818. [\[CrossRef\]](#)
10. Liu, P.; Song, S.; Zhou, Y. Semiparametric additive frailty hazard model for clustered failure time data. *Can. J. Stat.* **2022**, *50*, 549–571. [\[CrossRef\]](#)
11. Zeng, D.; Cai, J. Additive transformation models for clustered failure time data. *Lifetime Data Anal.* **2010**, *16*, 333–352. [\[CrossRef\]](#)
12. Yu, M.; Du, M. Regression analysis of multivariate interval-censored failure time data under transformation model with informative censoring. *Mathematics* **2022**, *10*, 3257. [\[CrossRef\]](#)
13. Sun, T.; Ding, Y. Copula-based semiparametric regression method for bivariate data under general interval censoring. *Biostatistics* **2021**, *22*, 315–330. [\[CrossRef\]](#)
14. Marra, G.; Radice, R. Copula link-based additive models for right-censored event time data. *J. Am. Stat. Assoc.* **2020**, *115*, 886–895. [\[CrossRef\]](#)
15. Petti, D.; Eletti, A.; Marra, G.; Radice, R. Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme. *Comput. Stat. Data Anal.* **2022**, *175*, 107550. [\[CrossRef\]](#)
16. Cheng, G.; Zhou, L.; Huang, J.Z. Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data. *Bernoulli* **2014**, *20*, 141–163. [\[CrossRef\]](#)
17. Lu, X.; Song, P.X.K. Efficient estimation of the partly linear additive hazards model with current status data. *Scand. J. Stat.* **2015**, *42*, 306–328. [\[CrossRef\]](#)
18. Wang, X.; Song, Y.; Zhang, S. An efficient estimation for the parameter in additive partially linear models with missing covariates. *J. Korean Stat. Soc.* **2020**, *49*, 779–801. [\[CrossRef\]](#)

19. Lee, C.Y.; Wong, K.Y.; Lam, K.F.; Xu, J. Analysis of clustered interval-censored data using a class of semiparametric partly linear frailty transformation models. *Biometrics* **2022**, *78*, 165–178. [[CrossRef](#)] [[PubMed](#)]
20. Chen, W.; Ren, F. Partially linear additive hazards regression for clustered and right censored data. *Bulletin of Informatics and Cybernetics* **2022**, *54*, 1–14. [[CrossRef](#)]
21. Sklar, M. Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
22. Nelson, R.B. *An Introduction to Copulas*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2006.
23. Joe, H. *Multivariate Models and Dependence Concepts*; CRC Press: Boca Raton, FL, USA, 1997.
24. Clayton, D.G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151. [[CrossRef](#)]
25. Gumble, E.J. Bivariate exponential distributions. *J. Am. Statitital Assoc.* **1960**, *55*, 698–707. [[CrossRef](#)]
26. Zhou, Q.; Hu, T.; Sun, J. A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *J. Am. Stat. Assoc.* **2017**, *112*, 664–672. [[CrossRef](#)]
27. Carnicer, J.M.; Peña, J.M. Shape preserving representations and optimality of the bernstein basis. *Adv. Comput. Math.* **1993**, *1*, 173–196. [[CrossRef](#)]
28. Burnham, K.P.; Anderson, D.R.; Burnham, K.P.; Anderson, D.R. *Practical Use of the Information-Theoretic Approach*; Springer: Berlin/Heidelberg, Germany, 1993; pp. 75–117.
29. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [[CrossRef](#)]
30. Age-Related Eye Disease Study Research Group. The age-related eye disease study (AREDS): Design implications AREDS report no. 1. *Control. Clin. Trials* **1999**, *20*, 573. [[CrossRef](#)]
31. Sun, T.; Ding, Y. CopulaCenR: Copula based regression models for bivariate censored data in R. *R J.* **2020**, *12*, 266. [[CrossRef](#)]
32. Huang, J. Efficient estimation for the proportional hazards model with interval censoring. *Ann. Stat.* **1996**, *24*, 540–568. [[CrossRef](#)]
33. Zhang, Y.; Hua, L.; Huang, J. A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scand. J. Stat.* **2010**, *37*, 338–354. [[CrossRef](#)]
34. Huang, J.; Rossini, A. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *J. Am. Stat. Assoc.* **1997**, *92*, 960–967. [[CrossRef](#)]
35. Wen, C.C.; Chen, Y.H. A frailty model approach for regression analysis of bivariate interval-censored survival data. *Stat. Sin.* **2013**, *23*, 383–408. [[CrossRef](#)]
36. Van der Vaart, A.; Wellner, J. *Weak Convergence and Empirical Processes*; Springer: Berlin/Heidelberg, Germany, 1996.
37. Shen, X.; Wong, W.H. Convergence rate of sieve estimates. *Ann. Stat.* **1994**, *22*, 580–615. [[CrossRef](#)]
38. Lorentz, G.G. *Bernstein Polynomials*; American Mathematical Society: Providence, RI, USA, 2013.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.