Mathematics
https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(11): 27755–27774.
DOI: 10.3934/math.20251220
Received: 23 September 2025
Revised: 13 November 2025
Accepted: 20 November 2025
Published: 27 November 2025

Research article

Efficient sieve estimation of semiparametric probit model with doubly censored data

Lanxin Cui[1], **Shishun Zhao**[1] **and Shuwei Li**[2,*]

[1] School of Mathematics, Jilin University, Changchun 130000, China
[2] School of Economics and Statistics, Guangzhou University, Guangzhou 510000, China

* **Correspondence:** Email: seslishuw@gzhu.edu.cn.

Abstract: Semiparametric probit model serves as a valuable alternative to the popular proportional hazards/odds model in survival analysis partly due to the use of a standard normal distributed random error. This feature can facilitate developing an efficient inference and may render a better fit for the real world data than other models. In this work, we concern regression analysis of doubly censored data with a spline-based probit regression model and provide an efficient maximum likelihood estimation procedure. A novel and reliable expectation-maximization algorithm is proposed to identify the sieve estimator. Asymptotic properties of the proposed estimator are established. Simulation studies suggest that the proposed method works well in finite samples and obviously outperforms the direct sieve maximum likelihood method, which is accomplished with some existing optimization algorithm in the software. An application to a real data set is also provided.

Keywords: double censoring; maximum likelihood estimation; monotone splines; semiparametric model; survival analysis
Mathematics Subject Classification: 62N02, 62G08

1. Introduction

Doubly censored data frequently arises in various fields, particularly in medical research, where the response of interest (e.g., a failure time) may not be observed exactly due to the presence of either left or right censoring [1,2]. In such data, the exact response of interest T is only available within the interval $(L, R]$. Left censoring arises if the response of a subject is smaller than L, while right censoring happens when the subject has not experienced the response by R. For instance, in a clinical trial comparing two drug treatments for AIDS [3], the HIV-1 RNA level in plasma is an important index for assessing the treatment efficacy. The NucliSens assay used to measure the HIV-1 RNA level has detection limits in the sense that it can only provide an accurate measure if the RNA level is between 400 and 750,000 copies per milliliter of plasma. Otherwise, the RNA level cannot be measured exactly and is treated as being left censoring at 400 or right censoring by 750,000. Thus, doubly censored data on the RNA level occurs.

The analysis of doubly censored data has gained considerable attention, and a growing number of methods have been proposed to estimate the survival function, perform two-sample tests, and conduct regression analysis. For example, in the univariate case, estimation of the survival distribution typically involves solving self-consistent equations [4, 5]. Gehan [1] and Hughes [6] proposed comparative analyses of two samples with doubly censored data. In the context of regression analysis, techniques based on M-estimation for the linear regression model have been advanced by Zhang and Li [7] and Ren and Gu [8]. Kim et al. [9] and Shen [10] investigated maximum likelihood estimation (MLE) methods along with asymptotic properties for the proportional hazards (PH), transformation and Cox-Aalen models, respectively. Cai and Cheng [11] proposed an unbiased estimating equation approach for the linear transformation model. Ji et al. [12] developed a martingale-based estimation method for the quantile regression model. Li et al. [13] explored a shared frailty modeling approach for multivariate doubly censored data.

Notably, most existing regression methods for analyzing doubly censored data are built upon the PH or proportional odds (PO) assumptions, which may be restrictive in practice. The semiparametric probit model provides an useful alternative by modeling the transformed failure time with a linear combination of covariates and a standard normal distributed random error. This feature can facilitate developing an efficient inference and may render a better fit for the real world data than other models. In particular, let T denote the failure time of interest, and $\boldsymbol{X}$ be a p-dimensional covariate vector. Given $\boldsymbol{X}$, the probit regression model posits that the cumulative distribution function of T takes the form

$$F(t \mid \boldsymbol{X}) = \Phi\{\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}\}, \quad \forall t \in [0, \infty), \tag{1.1}$$

where $\Phi(\cdot)$ denotes the distribution function of the standard normal random variable, $\boldsymbol{\beta}$ is the vector of regression parameters, and $\alpha(\cdot)$ is an unspecified increasing function with $\alpha(0) = -\infty$ and $\alpha(\infty) = \infty$. As commented by Deng et al. [14] and others, model (1.1) can be rewritten as $\alpha(T) = -\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} + \epsilon$, where ϵ is the random error term following the standard normal distribution. That is, the probit regression model evaluates covariate effects on the transformed failure time of interest. In addition, under model (1.1), the hazard function of T given $\boldsymbol{X}$ is

$$\lambda(t \mid \boldsymbol{X}_i) = \alpha'(t)\phi\{\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i\}\left[1 - \Phi\{\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i\}\right]^{-1},$$

where $\alpha'(t) = \mathrm{d}\alpha(t)/\mathrm{d}t$ and $\phi(\cdot)$ is the density function of a standard normal random variable. Thus, in contrast to the PH model, the probit model allows a time-varying hazard ratio regarding two different covariate values and may be less restrictive.

To make inference on the probit regression model (1.1), a growing number of frequentist and Bayesian methods have been developed under various situations [14–20]. For instance, Huang and Cai [21] conducted a mediation analysis for right censored data. Lin and Wang [16] and Liu and Qin [17] investigated regression analysis of current status data with the Bayesian and MLE procedures, respectively. Du et al. [18] proposed a sieve MLE method for analyzing current status data with informative censoring. Regarding the general interval-censored data, Lin and Wang [15] introduced a Bayesian estimation method, while Deng et al. [14] developed a pseudo MLE approach as well as a variable selection procedure. Despite the aforementioned method developments, to our knowledge, estimating the semiparametric probit model (1.1) under doubly censored data has not been investigated. This is partly because the coexistence of left and right censoring complicates the likelihood construction and limits the applicability of existing estimation procedures. Filling in this methodological gap constitutes the main goal of the current study.

In this work, we introduce an efficient sieve MLE approach for semiparametric probit model under doubly censored data. Specifically, we approximate the nuisance function $\alpha(t)$ of model (1.1) with monotone splines. An expectation-maximization (EM) algorithm that utilizes normal and multinomial distributed latent variables is proposed to maximize the intractable observed data likelihood. The proposed algorithm is easy to implement, robust to initial values and reliable, making it a reliable and valuable tool for the analysis of doubly censored data. Asymptotic properties of the proposed sieve estimator, including the consistency, asymptotic normality and semiparametric efficiency, are established with the empirical process techniques and sieve estimation theory. Furthermore, thanks to the desirable estimation performance of the proposed method, we can readily obtain the variance estimates through a numerical profile likelihood method instead of the intensive bootstrapping used in Li et al. [3].

The remainder of this paper is organized as follows. Section 2 introduces the notation, data structure and observed data likelihood function. Section 3 describes the proposed MLE procedure along with an EM algorithm. In Section 4, we discuss the asymptotic properties of the resultant sieve estimator. Simulation studies are carried out in Section 5 to evaluate the proposed method's empirical performance, followed by an application in Section 6. Section 7 gives some discussions and remarks. Appendix A provides a summary table of symbols used in this paper, and Appendix B contains the technical proofs of the asymptotic properties outlined in Section 4.

2. Notation, likelihood and monotone splines

Consider a failure time study giving doubly censored data for n independent individuals. In this case, the failure time of interest T can only be observed exactly within an interval $(L, R]$. Otherwise, the failure time suffers from left censoring if $T < L$ and right censoring if $T > R$. Define $\delta_1 = I(T \leq L)$, $\delta_2 = I(L < T \leq R)$ and $\delta_3 = I(T > R)$, where $I(\cdot)$ is the indicator function. In what follows, notations with subscript i represents their sample analogues. Then the observed data consist of $\mathcal{D} = \{(\widetilde{T}_i, \delta_{i1}, \delta_{i2}, \delta_{i3}, \boldsymbol{X}_i); i = 1, \ldots, n\}$, where $\widetilde{T}_i = \max\{L_i, \min(T_i, R_i)\}$.

Assuming that T_i and $(L_i, R_i]$ are conditionally independent given the covariates, which is essential for the likelihood construction, the observed data likelihood function associated with model (1.1) is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \alpha) &= \prod_{i=1}^{n} F(\widetilde{T}_i \mid \boldsymbol{X}_i)^{\delta_{i1}} f(\widetilde{T}_i \mid \boldsymbol{X}_i)^{\delta_{i2}} \{1 - F(\widetilde{T}_i \mid \boldsymbol{X}_i)\}^{\delta_{i3}} \\ &= \prod_{i=1}^{n} \Phi\{\alpha(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\}^{\delta_{i1}} [\phi\{\alpha(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\} \alpha'(\widetilde{T}_i)]^{\delta_{i2}} [1 - \Phi\{\alpha(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\}]^{\delta_{i3}}, \qquad (2.1) \end{aligned}$$

where $F(t \mid \boldsymbol{X}_i) = \Phi\{\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\}$ (1.1) and $f(t \mid \boldsymbol{X}_i) = \mathrm{d}F(t \mid \boldsymbol{X}_i)/\mathrm{d}t$ is the density function of $\widetilde{T}_i$ given $\boldsymbol{X}_i$ for $i = 1, \ldots, n$.

Notably, in likelihood (2.1), $\alpha(\cdot)$ is infinite-dimensional. A routine way in survival analysis is to approximate $\alpha(\cdot)$ with some smooth function. Herein, we propose to use the monotone splines, taking the form

$$\alpha_n(t) = \xi + \sum_{k=1}^{K_n} \gamma_k b_k(t),$$

where $b_k(t)$'s are integrated spline basis functions, ξ is the intercept, and γ_k's are non-negative spline coefficients to ensure a monotone increasing function [22]. Before using the monotone splines, one needs to specify the interior knots, a sequence of q_n increasing points, and to determine the spline degree $\tilde{l}$. By setting $\tilde{l} = 1$, 2 and 3, we can obtain linear, quadratic, and cubic splines, respectively. The $K_n = q_n + \tilde{l}$ basis functions can be fully determined if the interior knots and degree are specified. In practice, one can try to use linear, quadratic or cubic splines and place a series of interior knots at equally spaced quantiles of a certain time interval formed by the observed data. Then some commonly used model selection criterion, such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), can be used to determine the optimal model.

Through the above approximation, the observed data likelihood can be expressed as

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \prod_{i=1}^{n} & \Phi\{\xi + \sum_{k=1}^{K_n} \gamma_k b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\}^{\delta_{i1}} \\ & \times \left[\phi\{\xi + \sum_{k=1}^{K_n} \gamma_k b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\} \sum_{k=1}^{K_n} \gamma_k m_k(\widetilde{T}_i) \right]^{\delta_{i2}} \qquad (2.2) \\ & \times \left[1 - \Phi\{\xi + \sum_{k=1}^{K_n} \gamma_k b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\} \right]^{\delta_{i3}}, \end{aligned}$$

where $\boldsymbol{\zeta} = (\xi, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{K_n})^{\mathrm{T}}$, and $m_k(t)$ is the first derivative of $b_k(t)$ with respect to t for $k = 1, \ldots, K_n$.

To estimate all unknown parameters, one could consider performing a direct maximization of (2.2) with some existing optimization algorithm in the software. However, due to the intractable form of (2.2), such an approach is cumbersome and unreliable as manifested by the simulations in Section 5. To render an accurate and easy-to-implement estimation, in the next section, we develop an EM algorithm which utilizes some normal and multinomial distributed random variables to proceed with the data augmentation.

3. Estimation algorithm

The proposed EM algorithm involves a two-stage data augmentation procedure, i.e., E-step and M-step. We first construct the complete data likelihood with a tractable form through data augmentation. Motivated by Lin and Wang [15], we introduce a normally distributed latent variable Z_i as follows

$$Z_i = N(\alpha_n(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i, 1), \quad i = 1, \ldots, n,$$

where $Z_1, \ldots, Z_n$ are independent and the required constraints are $Z_i > 0$ if $\delta_{i1} = 1$, $Z_i < 0$ if $\delta_{i3} = 1$ and $Z_i = 0$ if $\delta_{i2} = 1$ for $1, \ldots, n$. With the fact $F(t \mid \boldsymbol{X}_i) = \Phi\{\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\} = 1 - \int_{-\infty}^{0} \phi\{z_i - (\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i)\} \mathrm{d}z_i$, we can obtain the augmented likelihood

$$L_1(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \prod_{i=1}^{n} \phi\{Z_i - (\alpha_n(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i)\} \alpha_n'(\widetilde{T}_i)^{\delta_{i2}} \{1_{(Z_i>0)}\}^{\delta_{i1}} \{1_{(Z_i=0)}\}^{\delta_{i2}} \{1_{(Z_i<0)}\}^{\delta_{i3}},$$

where $\alpha_n'(t) = \mathrm{d}\alpha_n(t)/\mathrm{d}t$. Based on the spline representations of $\alpha_n(\cdot)$ and $\alpha_n'(\cdot)$, for each i, we introduce a multinomial latent vector $\boldsymbol{u}_i = (u_{i1}, \cdots, u_{iK_n})^{\mathrm{T}} \sim M(1, \boldsymbol{p}_i)$, a multinomial distribution with the total count equal to 1 and $\boldsymbol{p}_i = (1/K_n, \ldots, 1/K_n)$. Then we have the proposed complete data likelihood, which takes the form

$$L_c(\boldsymbol{\beta}, \boldsymbol{\zeta}) = \prod_{i=1}^{n} \frac{1}{2\pi} \exp\left[-\frac{1}{2} \{Z_i - (\xi + \sum_{k=1}^{K_n} \gamma_k b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i)\}^2 \right]$$
$$\times \left[\prod_{k=1}^{K_n} \{\gamma_k m_k(\widetilde{T}_i)\}^{\delta_{i2} u_{ik}} \right] \{1_{(Z_i>0)}\}^{\delta_{i1}} \{1_{(Z_i=0)}\}^{\delta_{i2}} \{1_{(Z_i<0)}\}^{\delta_{i3}}.$$

Note that unlike Lin and Wang [15], we have exactly observed failure times in doubly censored data and thus additionally introduce a multinomial latent vector for each subject in the above data augmentation procedure, rendering $L_c(\boldsymbol{\beta}, \boldsymbol{\zeta})$ with a simple form.

The E-step involves calculating the conditional expectation of $\log L_c(\boldsymbol{\beta}, \boldsymbol{\zeta})$ given the observed data $\mathcal{D}$, the constraints of Z_i's in $L_c(\boldsymbol{\beta}, \boldsymbol{\zeta})$ and the dth parameter updates, $\boldsymbol{\beta}^{(d)}$ and $\boldsymbol{\zeta}^{(d)}$, with $d \geq 0$. In particular, we define $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\zeta}^{(0)}$ as the initial values of $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$, respectively. For notational simplicity, in what follows, we will ignore all the conditional arguments in the expressions of conditional expectations. This step leads to

$$Q(\boldsymbol{\beta}, \boldsymbol{\zeta}; \boldsymbol{\beta}^{(d)}, \boldsymbol{\zeta}^{(d)}) = -\frac{1}{2} \sum_{i=1}^{n} \{\xi + \sum_{k=1}^{K_n} \gamma_k b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\}^2 + \sum_{i=1}^{n} \sum_{k=1}^{K_n} E(u_{ik}) \log(\gamma_k) \delta_{i2}$$
$$+ \sum_{i=1}^{n} E(Z_i) \{\xi + \sum_{k=1}^{K_n} \gamma_k b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i\},$$

where $E(u_{ik}) = \frac{\gamma_k^{(d)} m_k(\widetilde{T}_i)}{\sum_{l=1}^{K_n} \gamma_l{}^{(d)} m_l(\widetilde{T}_i)}$,

$$E(Z_i) = \begin{cases} \nu_i^{(d)} + \phi\{\nu_i^{(d)}\}/[1 - \Phi\{\nu_i^{(d)}\}], & \text{if } \delta_{i1} = 1, \\ 0, & \text{if } \delta_{i2} = 1, \\ \nu_i^{(d)} - \phi\{\nu_i^{(d)}\}/\Phi\{\nu_i^{(d)}\}, & \text{if } \delta_{i3} = 1, \end{cases}$$

and $\nu_i^{(d)} = \xi^{(d)} + \sum_{k=1}^{K_n} \gamma_k^{(d)} b_k(\widetilde{T}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i^{(d)}$, for $k = 1, \ldots, K_n$ and $i = 1, \ldots, n$.

In the M-step, by setting $\xi = \xi^{(d)}$ and $\gamma_k = \gamma_k^{(d)}$ for $k = 1, \ldots K_n$, we first maximize $Q(\boldsymbol{\beta}, \boldsymbol{\zeta}^{(d)}; \boldsymbol{\beta}^{(d)}, \boldsymbol{\zeta}^{(d)})$ with respect to $\boldsymbol{\beta}$ and obtain the following least square solution

$$\boldsymbol{\beta}^{(d+1)} = (\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^{\mathrm{T}})^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \left\{ \xi^{(d)} + \sum_{k=1}^{K_n} \gamma_k^{(d)} b_k(\widetilde{T}_i) - E(Z_i) \right\}. \tag{3.1}$$

Next, we maximize $Q(\boldsymbol{\beta}^{(d+1)}, \xi, \boldsymbol{\gamma}^{(d)}; \boldsymbol{\beta}^{(d)}, \boldsymbol{\zeta}^{(d)})$ with respect to ξ and obtain an explicit expression

$$\xi^{(d+1)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i^{(d+1)} + E(Z_i) - \sum_{k=1}^{K_n}\gamma_k^{(d)} b_k(\widetilde{T}_i)\right\}. \tag{3.2}$$

To update each γ_k, we suppose that $\gamma_{k'}^{(d+1)}$ has been available for all $k' < k$. Let $Q_k^{(d)}(\gamma_k)$ be the updated $Q(\boldsymbol{\beta}^{(d+1)}, \xi^{(d+1)}, \boldsymbol{\gamma}; \boldsymbol{\beta}^{(d)}, \boldsymbol{\gamma}^{(d)})$ with $\gamma_{k'} = \gamma_{k'}^{(d+1)}$ for $k' < k$ and $\gamma_{k'} = \gamma_{k'}^{(d)}$ for $k' > k$, the derivative of $Q_k^{(d)}(\gamma_k)$ with respect to γ_k is given by

$$\mathrm{d}Q_k^{(d)}(\gamma_k)/\mathrm{d}\gamma_k = \sum_{i=1}^{n}\Bigg[-b_k(\widetilde{T}_i)\Bigg\{\xi^{(d+1)} + \sum_{k'<k}\gamma_{k'}^{(d+1)} b_{k'}(\widetilde{T}_i) + \sum_{k'>k}\gamma_{k'}^{(d)} b_{k'}(\widetilde{T}_i) - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i^{(d+1)} - E(Z_i)\Bigg\} + \frac{E(u_{ik})}{\gamma_k}\Bigg].$$

Examining $\mathrm{d}Q_k^{(d)}(\gamma_k)/\mathrm{d}\gamma_k = 0$ leads to the following quadratic equation of γ_k

$$\gamma_k^2\sum_{i=1}^{n} b_k^2(\widetilde{T}_i) + \gamma_k\sum_{i=1}^{n}\left[b_k(\widetilde{T}_i)\left\{\xi + \sum_{k'\neq k}\gamma_{k'} b_{k'}(\widetilde{T}_i) - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i - E(Z_i)\right\}\right] - \sum_{i=1}^{n} E(u_{ik}) = 0.$$

Let

$$a_k = \sum_{i=1}^{n} b_k^2(\widetilde{T}_i),\ e_k^{(d)} = \sum_{i=1}^{n}\delta_i E(u_{ik}),$$

and

$$c_k^{(d)} = \sum_{i=1}^{n} b_k(\widetilde{T}_i)\left\{\xi^{(d+1)} + \sum_{k'<k}\gamma_{k'}^{(d+1)} b_{k'}(\widetilde{T}_i) + \sum_{k'>k}\gamma_{k'}^{(d)} b_{k'}(\widetilde{T}_i) - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i^{(d+1)} - E(Z_i)\right\}$$

for $k = 1, \ldots, K_n$, we have

$$\gamma_k^{(d+1)} = \frac{-c_k^{(d)} + \sqrt{c_k^{(d)2} + 4a_k e_k^{(d)}}}{2a_k}. \tag{3.3}$$

In summary, detailed iteration processes of the proposed EM algorithm are given in **Algorithm 1**.

Algorithm 1

```
1) Let $d = 0$ and initialize $\boldsymbol{\beta}^{(0)}, \xi^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$.
2) Calculate the conditional expectations $E(Z_i)$ and $E(u_{ik})$ based on the observed data
   and $\{\boldsymbol{\beta}^{(d)}, \xi^{(d)}, \boldsymbol{\gamma}^{(d)}\}$ for $i = 1, \ldots, n$ and $k = 1, 2, \ldots, K_n$.
3) Obtain $\boldsymbol{\beta}^{(d+1)}$ with the closed-from solution (3.1).
4) Calculate $\xi^{(d+1)}$ with (3.2).
5) Calculate $\gamma_k^{(d+1)}$ with (3.3) for $k = 1, \ldots, K_n$, and let $d = d + 1$.
6) Repeat Step 2 to Step 5 until the convergence is achieved.
```

Our experiences show that the performance of the proposed EM algorithm is insensitive to the choices of the initial values, $\boldsymbol{\beta}^{(0)}, \xi^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$. In practice, one can simply set the initial values of each regression parameter and each spline coefficient to 0 and 1, respectively. The algorithm can be identified as convergence when the maximum absolute difference between successive parameter updates is less than a small positive threshold (e.g., 10^{-7}).

4. Asymptotic properties

Let $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}(\cdot)$ be the proposed estimators of $\boldsymbol{\beta}$ and $\alpha(\cdot)$, respectively. Let $\boldsymbol{\beta}_0$ and $\alpha_0(\cdot)$ be the true values of $\boldsymbol{\beta}$ and $\alpha(\cdot)$, respectively. Define $\Theta = \mathcal{B}\times\mathcal{A}$ as the parameter space of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha)$, where $\boldsymbol{\beta} \in \mathcal{B}$, a compact set in R^p, $\alpha \in \mathcal{A} = \{\alpha(t) : \mathrm{d}\alpha(t)/\mathrm{d}t > 0, -M_\alpha \le \alpha(t) \le M_\alpha \text{ for } t \in [\tau_u, \tau_v]\}$, $[\tau_u, \tau_v]$ denotes the union support of random variables T, L, and R with $0 < \tau_u < \tau_v < \infty$, and both τ and M_α are positive constants. Consider a knot sequence $\tau_u = t_1 = \cdots = t_{\tilde{l}} < t_{\tilde{l}+1} < \cdots < t_{\tilde{l}+q_n} < t_{\tilde{l}+q_n+1} = \cdots = t_{q_n+2\tilde{l}} = \tau_v$, where $q_n = O(n^\nu)$ with $0 < \nu < 0.5$. Define the sieve space $\Theta_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\beta}, \alpha_n); \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n\}$, where $\mathcal{A}_n = \{\alpha_n(t) = \xi + \sum_{k=1}^{K_n} \gamma_k b_k(t); b_k(t) \in [0,1], t \in [\tau_u, \tau_v], -M_\alpha \le \xi \le M_\alpha, 0 \le \gamma_k \le M_\alpha\}$. The first derivative of $\alpha_n(t)$ with respect to t is expressed as $\alpha_n'(t) = \sum_{k=1}^{K_n} \gamma_k m_k(t)$, where $m_k(t)$ is the first derivative of $b_k(t)$ with respect to t for $k = 1, \ldots, K_n$. Let $\|\cdot\|$ denote the Euclidean norm. For a vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\|\boldsymbol{\beta}\| = (\sum_{i=1}^p \beta_i^2)^{1/2}$. Let $\|\cdot\|_2$ denote the L_2-norm with respect to a probability measure. For any two $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \alpha_1)$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \alpha_2)$ in the parameter space, define

$$\rho(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\|\alpha_1 - \alpha_2\|_2^2 + \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2)^{1/2},$$

where $\|\alpha_1 - \alpha_2\|_2^2 = \int_{\tau_u}^{\tau_v} \{\alpha_1(s) - \alpha_2(s)\}^2 \mathrm{d}F(s)$ with $F(\cdot)$ being a distribution function.

To establish asymptotic properties of $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}(\cdot)$, like Chen et al. [23], Li et al. [3] and others, we use the following standard regularity conditions.

(A1) (i). $P(L < R \mid \boldsymbol{X}) = 1$; (ii). There exists a finite $\tau_1 > 0$ such that $P(R \le \tau_1 \mid \boldsymbol{X}) = 1$; (iii). $P(L \in (0, \tau_3) \mid \boldsymbol{X}) = 0$ and $P(L \le \tau_2 \mid \boldsymbol{X}) = 1$, where $0 < \tau_3 \le \tau_2 \le \tau_1$.

(A2) T and (L, R) are conditionally independent given the covariates.

(A3) The covariate vector $\boldsymbol{X}$ takes values in a bounded subset of R^p.

(A4) The minimum spacing between knots satisfies $\Delta_\alpha = \max_{\tilde{l}+1 \le k \le \tilde{l}+q_n+1} |t_k - t_{k-1}| = O(n^{-\nu})$ with $0 < \nu < 0.5$. Additionally, the ratio $\Delta_\alpha/\delta_\alpha$ remains bounded, where $\delta_\alpha = \min_{\tilde{l}+1 \le k \le \tilde{l}+q_n+1} |t_k - t_{k-1}|$ represents the knots' minimum spacing.

(A5) (i). The υth derivative of $\alpha(t)$ with respect to t, denoted by $\alpha^{(\upsilon)}(t)$, exists; (ii). There exists a positive constant $M > 0$ such that $\alpha^{(\upsilon)}(t)$ satisfies the order-η Lipschitz condition:

$$\left|\alpha^{(\upsilon)}(s) - \alpha^{(\upsilon)}(t)\right| \le M|s - t|^\eta \quad \text{for any } s, t \in [\tau_u, \tau_v];$$

(iii). Let r be a nonnegative integer and $\eta \in (0, 1]$ such that $r = \upsilon + \eta > 0.5$.

(A6) The true regression parameter $\boldsymbol{\beta}_0$ belongs to the interior of $\mathcal{B}$, a compact set in R^p. The true value of α denoted by α_0 is in $\mathcal{A}$.

(A7) If $\alpha(t) + \boldsymbol{\beta}^\top \boldsymbol{X} = 0$ for any $t \in [\tau_u, \tau_v]$ with probability one, then $\boldsymbol{\beta} = 0$ and $\alpha(t) = 0$ for $t \in [\tau_u, \tau_v]$.

Conditions (A1)–(A3) are standard assumptions regarding the data structure and censoring mechanism. Conditions (A4) and (A5) provide smoothness and spline approximation requirements. Conditions (A6) and (A7) guarantee parameter space compactness and model identifiability. The following two theorems summarize the asymptotic properties of the proposed estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}(\cdot)$. Detailed proofs of the theorems are sketched in the Appendix B.

Theorem 1. *Suppose that conditions (A1)–(A6) hold, as n → ∞, we have*

$$\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O(n^{-(1-\nu)/2} + n^{-r\nu}).$$

In particular, by taking $\nu = 1/(2r+1)$, $\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O\left(n^{-r/(2r+1)}\right)$.

As commented by Huang [24] and others, the smoothness assumption of $\alpha(t)$ is often needed. Usually, in many situations, one can set r defined in (A6) to 1 (i.e., $\upsilon = 0$ and $\eta = 1$) or 2 (i.e., $\upsilon = 1$ and $\eta = 1$), which correspond to assuming that $\alpha(t)$ has bounded first-order derivative or bounded second-order derivative. Notably, if $v = 1/(2r+1)$, the convergence rate of $\hat{\boldsymbol{\theta}}_n$ is $n^{r/(2r+1)}$, which is the same as the optimal rate in nonparametric regression setting. The following theorem states that the convergence rate of $\hat{\boldsymbol{\beta}}$ achieves $\sqrt{n}$.

Theorem 2. *Suppose that conditions (A1)–(A7) hold, as $n \to \infty$, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\beta}_0)), \tag{4.1}$$

where "$\xrightarrow{d}$" denotes the convergence in distribution and $\mathbf{I}(\boldsymbol{\beta}_0)$ is defined in the Appendix B.

To make inference about the regression vector $\boldsymbol{\beta}$, the parameter of interest, it is apparent that one needs to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ as suggested by Theorem 2. Herein, by following Zeng et al. [25] and others, we suggest to employ a numerical profile likelihood approach. Specifically, define $pl_n(\boldsymbol{\beta}) = \max_{\boldsymbol{\zeta}} \log L(\boldsymbol{\beta}, \boldsymbol{\zeta})$, where $L(\boldsymbol{\beta}, \boldsymbol{\zeta})$ is given in (2.2). The limiting covariance matrix of $\hat{\boldsymbol{\beta}}$ can be estimated by $(n\hat{\mathbf{I}}_n)^{-1}$

$$\hat{\mathbf{I}}_n = n^{-1} \sum_{i=1}^{n} \left\{ \frac{pl_i(\hat{\boldsymbol{\beta}} + w_n e_j) - pl_i(\hat{\boldsymbol{\beta}})}{w_n} \right\} \left\{ \frac{pl_i(\hat{\boldsymbol{\beta}} + w_n e_j) - pl_i(\hat{\boldsymbol{\beta}})}{w_n} \right\}^{\mathrm{T}},$$

where $pl_i(\boldsymbol{\beta})$ is the ith subject's contribution to $pl_n(\boldsymbol{\beta})$, $w_n = cn^{-1/2}, c$ is a positive constant, and, for $j = 1, \ldots, p$, e_j is p-dimensional vector whose jth element is 1 and 0 elsewhere. To calculate $pl_n(\boldsymbol{\beta})$, one can directly use a simplified version of the above EM algorithm to maximize $L(\boldsymbol{\beta}, \boldsymbol{\zeta})$, which only involves updating $\boldsymbol{\zeta}$ given a fixed $\boldsymbol{\beta}$.

5. Simulation studies

In this section, simulation studies were conducted to evaluate the finite-sample performance of the proposed method. The failure time of interest T was generated from the probit regression model: $F(t \mid X_1, X_2) = \Phi(\alpha(t) + \beta_1 X_1 + \beta_2 X_2)$, where $X_1 \sim Uniform(0, 1)$, $X_2 \sim Bernoulli(0.5)$, $\alpha(t) = \log(t)$ or $\left(t^{0.1} - 1\right)/0.1$, and (β_1, β_2) equals $(0, 0.5)$, $(0.5, -0.5)$ or $(1, 1)$. We generated the left and right endpoints of the interval $(L, R]$ from $Uniform(0, 1)$ and $Uniform(2\tau/3, \tau)$, respectively, where τ was chosen to produce different right censoring rates. Given the simulated interval $(L, R]$, we had exact failure time T if $T \in (L, R]$. Otherwise, we had a left censored observation on T if $T < L$, and T suffered from right censoring if $T > R$. Herein, the right censoring rates were 24% – 43% and 46% – 65% by setting $\tau = 5$ or $\tau = 3$, respectively. We set $n = 200$ or 500, and used 500 replicates.

To implement the proposed method, we used monotone splines with degree 2 and 5 interior knots. In particular, the interior knots were placed at equally spaced quantiles of the interval formed by $\{\widetilde{T}_i; i = 1, \ldots, n\}$. We set the initial value of each regression parameter to 0, and initialized each spline coefficient with 1. To manifest the proposed EM algorithm's advantage, we also performed direct maximization of the observed data likelihood (2.2) using the `optim()` function in the R software. For this direct maximization procedure, we investigated two choices of initial values. The first choice was to use the same initial values as the proposed EM algorithm, which was abbreviated as "Direct maximization (a)", and the second choice was to set initial values to the true parameter values, which was abbreviated as "Direct maximization (b)".

Table 1 summarized the results of the regression coefficient estimates with the proposed EM algorithm and $\tau = 5$. The evaluation criteria include Bias (the estimation bias calculated by the average of 500 point estimates minus the true value), SSE (the sample standard error of 500 point estimates), SEE (the average of 500 standard error estimates), and CP95 (95% coverage probability formed by the normal approximation). The simulation results indicate that the proposed EM algorithm performs reasonably well across various true values of regression parameters, true $\alpha(\cdot)$ and sample sizes. The biases are virtually small, SSE is in well accordance with SEE, implying that the standard error estimators accurately reflect the true variations, and each CP95 is close to the nominal value 95%. The algorithm consistently converged within a similar number of iterations and yielded stable estimates, suggesting that the proposed EM procedure is insensitive to initialization and exhibits reliable convergence behavior.

Table 1. Simulation results with $\tau = 5$. Results include the estimation bias (Bias), the sample standard error of 500 point estimates (SSE), the average of 500 standard error estimates (SEE), and 95% coverage probability (CP95).

				Proposed EM algorithm				Direct maximization (a)				Direct maximization (b)			
n	$\alpha(t)$	(β_1,β_2)	Par.	Bias	SSE	SEE	CP95	Bias	SSE	SEE	CP95	Bias	SSE	SEE	CP95
200	$\log(t)$	(0,0.5)	β_1	-0.005	0.145	0.150	0.950	0.089	0.221	0.212	0.960	0.012	0.237	0.248	0.940
			β_2	-0.010	0.260	0.266	0.948	-0.097	0.233	0.221	0.860	0.005	0.222	0.223	0.950
		(0.5,-0.5)	β_1	-0.004	0.160	0.173	0.954	-0.016	0.208	0.196	0.950	0.069	0.222	0.198	0.955
			β_2	0.014	0.255	0.274	0.944	0.098	0.212	0.199	0.890	-0.023	0.204	0.190	0.960
		(1,1)	β_1	-0.032	0.161	0.169	0.946	-0.053	0.229	0.206	0.950	0.011	0.162	0.153	0.970
			β_2	-0.027	0.252	0.247	0.944	-0.117	0.199	0.195	0.880	-0.026	0.179	0.180	0.950
	$\frac{t^{0.1}-1}{0.1}$	(0,0.5)	β_1	-0.002	0.140	0.149	0.954	0.071	0.188	0.184	0.960	-0.009	0.175	0.177	0.940
			β_2	-0.001	0.261	0.264	0.944	-0.085	0.191	0.195	0.860	0.004	0.180	0.185	0.950
		(0.5,-0.5)	β_1	-0.013	0.154	0.162	0.946	-0.049	0.185	0.198	0.960	-0.038	0.210	0.214	0.950
			β_2	0.003	0.266	0.273	0.948	0.134	0.218	0.216	0.890	-0.044	0.197	0.214	0.960
		(1,1)	β_1	-0.017	0.158	0.166	0.958	-0.116	0.199	0.197	0.870	0.002	0.199	0.196	0.940
			β_2	-0.023	0.248	0.241	0.950	-0.125	0.180	0.188	0.880	0.012	0.186	0.182	0.950
500	$\log(t)$	(0,0.5)	β_1	-0.001	0.094	0.094	0.946	0.079	0.166	0.162	0.900	-0.006	0.167	0.161	0.960
			β_2	-0.011	0.165	0.166	0.956	-0.065	0.160	0.161	0.920	0.011	0.163	0.162	0.970
		(0.5,-0.5)	β_1	-0.006	0.094	0.100	0.950	-0.042	0.166	0.169	0.930	0.034	0.171	0.178	0.960
			β_2	0.000	0.164	0.156	0.946	0.034	0.178	0.169	0.880	-0.028	0.176	0.176	0.940
		(1,1)	β_1	-0.031	0.100	0.105	0.952	-0.121	0.175	0.161	0.800	-0.009	0.167	0.161	0.960
			β_2	-0.028	0.156	0.153	0.950	-0.158	0.151	0.156	0.840	-0.032	0.168	0.152	0.930
	$\frac{t^{0.1}-1}{0.1}$	(0,0.5)	β_1	0.001	0.091	0.093	0.958	0.020	0.161	0.162	0.900	0.024	0.159	0.156	0.950
			β_2	-0.006	0.161	0.165	0.948	-0.080	0.166	0.163	0.880	-0.027	0.176	0.179	0.940
		(0.5,-0.5)	β_1	-0.003	0.096	0.100	0.956	-0.060	0.152	0.140	0.920	-0.066	0.175	0.171	0.920
			β_2	0.013	0.160	0.156	0.940	0.073	0.174	0.186	0.910	0.090	0.169	0.163	0.900
		(1,1)	β_1	-0.024	0.100	0.103	0.952	-0.086	0.158	0.153	0.850	0.015	0.155	0.142	0.910
			β_2	-0.020	0.156	0.150	0.950	-0.137	0.186	0.183	0.810	-0.008	0.162	0.166	0.940

* Note: "Direct maximization (a)" refers to the method that maximizes the likelihood directly with the existing software, in which the initial values were set to be the same as those used in the proposed EM algorithm. "Direct maximization (b)" refers to the method that maximizes the likelihood directly with existing software, in which the initial values were set to the true parameter values.

To further evaluate the normality of the proposed estimator, we drew quantile-quantile (Q-Q) plots of the standardized estimators against the standard normal distribution. For example, Figure 1 displayed the Q-Q plots for β_1 and β_2 with $n = 200$, $(\beta_1,\beta_2) = (0.5, -0.5)$, $\alpha(t) = \log(t)$ and $\tau = 5$. The almost linear plots suggest that the used normal approximation is reasonable. We also investigated the estimation performance of the proposed method for the baseline survival function $S(t) = 1 - \Phi\{\alpha(t)\}$. In particular, Figure 2 presents the average of 500 baseline survival function estimates as well as the true $\alpha(t)$ under the setup of $\alpha(t) = \log(t)$, $(\beta_1, \beta_2) = (0.5, -0.5)$ and $\tau = 5$. The two plots in Figure 2 show that the average function estimates are close to the true curves, reflecting a satisfactory estimation performance.

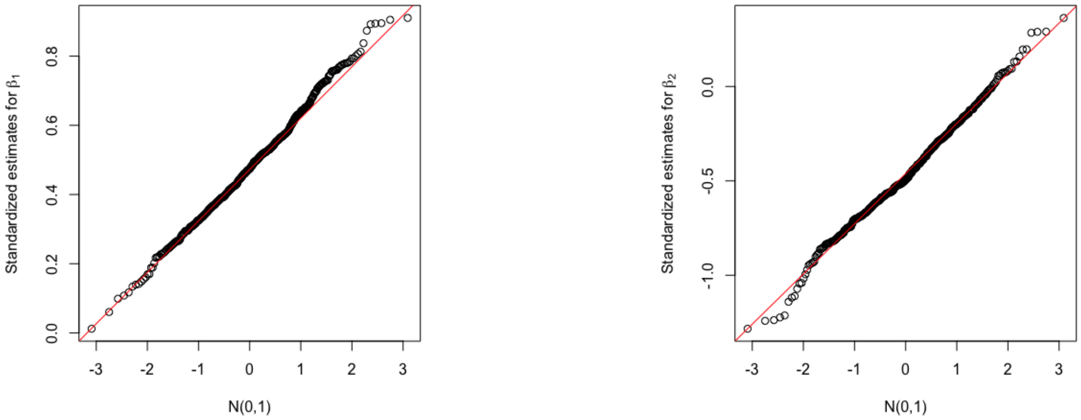

Figure 1. Quantile plots of the standardized, proposed estimates with $n = 200$, $(\beta_1, \beta_2) = (0.5, -0.5)$, $\alpha(t) = \log(t)$ and $\tau = 5$

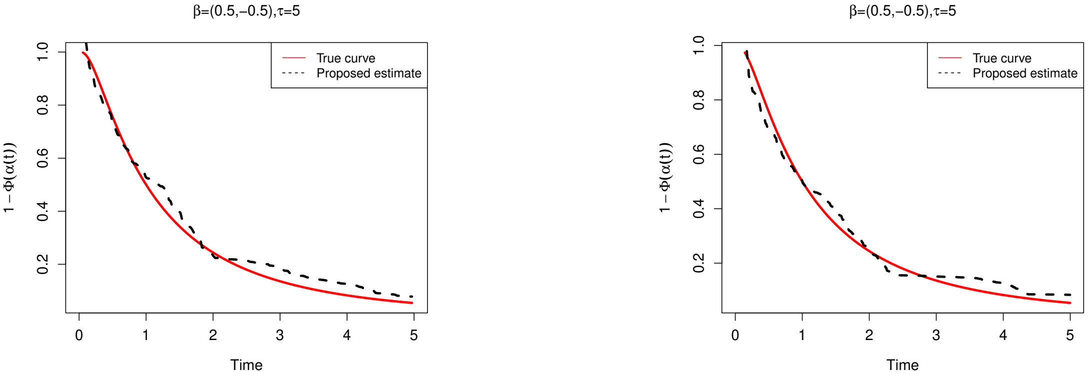

Figure 2. Simulation results for the estimation of the baseline survival function $S(t) = 1 - \Phi\{\alpha(t)\}$ with $\alpha(t) = \log(t)$, $(\beta_1, \beta_2) = (0.5, -0.5)$ and $\tau = 5$

Moreover, we presented in Tables 1 and 2 the simulation results obtained by "Direct maximization (a)" and "Direct maximization (b)". The results indicate that, by using the same initial values as our proposed EM algorithm, "Direct maximization (a)" often exhibits large estimation bias, while the performance of "Direct maximization (b)" improves obviously due to the use of true parameter values as the initial values. These comparative results demonstrate that performing direct MLE with some existing optimization procedure is sensitive to the initial value selection, a severe limitation that is nonexistent in the proposed approach. The results with $\tau = 3$ (high censoring rate) were given in Table 2, from which one can observe same findings as above.

Table 2. Simulation results with $\tau = 3$. Results include the estimation bias (Bias), the sample standard error of 500 point estimates (SSE), the average of 500 standard error estimates (SEE), and 95% coverage probability (CP95).

				Proposed EM algorithm				Direct maximization (a)				Direct maximization (b)			
n	$\alpha(t)$	(β_1,β_2)	Par.	Bias	SSE	SEE	CP95	Bias	SSE	SEE	CP95	Bias	SSE	SEE	CP95
200	$\log(t)$	(0,0.5)	β_1	0.005	0.154	0.156	0.946	0.055	0.218	0.203	0.920	0.016	0.212	0.221	0.910
			β_2	-0.011	0.274	0.273	0.944	0.005	0.188	0.185	0.970	-0.009	0.188	0.191	0.950
		(0.5,-0.5)	β_1	0.001	0.154	0.151	0.938	-0.053	0.179	0.167	0.920	-0.005	0.207	0.203	0.930
			β_2	-0.005	0.258	0.236	0.946	0.055	0.220	0.221	0.820	-0.007	0.181	0.179	0.930
		(1,1)	β_1	-0.030	0.181	0.182	0.946	-0.099	0.190	0.209	0.850	0.039	0.211	0.222	0.920
			β_2	-0.034	0.276	0.274	0.950	-0.104	0.154	0.156	0.820	0.017	0.220	0.214	0.950
	$\frac{t^{0.1}-1}{0.1}$	(0,0.5)	β_1	-0.004	0.152	0.156	0.956	0.169	0.205	0.199	0.800	-0.009	0.152	0.161	0.930
			β_2	0.005	0.250	0.274	0.946	-0.018	0.218	0.185	0.970	0.004	0.180	0.172	0.940
		(0.5,-0.5)	β_1	0.008	0.160	0.174	0.950	-0.069	0.176	0.168	0.920	-0.005	0.219	0.198	0.920
			β_2	0.002	0.272	0.278	0.940	0.064	0.222	0.214	0.870	0.014	0.189	0.197	0.970
		(1,1)	β_1	-0.022	0.164	0.169	0.952	-0.095	0.196	0.183	0.840	0.005	0.206	0.190	0.970
			β_2	-0.010	0.268	0.252	0.946	-0.103	0.179	0.162	0.840	-0.019	0.243	0.206	0.950
500	$\log(t)$	(0,0.5)	β_1	0.001	0.097	0.098	0.948	0.096	0.160	0.153	0.920	-0.005	0.130	0.130	0.950
			β_2	0.011	0.167	0.172	0.950	-0.113	0.172	0.172	0.930	0.008	0.152	0.128	0.950
		(0.5,-0.5)	β_1	-0.005	0.098	0.101	0.944	-0.113	0.228	0.209	0.900	0.036	0.122	0.122	0.980
			β_2	0.008	0.157	0.161	0.950	0.238	0.179	0.188	0.940	-0.033	0.101	0.103	0.940
		(1,1)	β_1	-0.012	0.109	0.112	0.950	0.103	0.145	0.147	0.890	-0.006	0.169	0.142	0.950
			β_2	-0.026	0.177	0.166	0.948	0.047	0.180	0.179	0.970	-0.031	0.173	0.154	0.940
	$\frac{t^{0.1}-1}{0.1}$	(0,0.5)	β_1	0.002	0.097	0.097	0.946	0.110	0.186	0.186	0.900	0.014	0.139	0.139	0.920
			β_2	0.011	0.167	0.171	0.948	-0.102	0.190	0.195	0.930	-0.009	0.130	0.140	0.970
		(0.5,-0.5)	β_1	-0.004	0.098	0.101	0.942	-0.063	0.151	0.155	0.930	-0.014	0.142	0.158	0.950
			β_2	0.006	0.157	0.160	0.950	0.127	0.168	0.175	0.880	0.035	0.136	0.136	0.920
		(1,1)	β_1	-0.024	0.106	0.105	0.954	0.049	0.183	0.182	0.990	0.060	0.183	0.183	0.940
			β_2	-0.022	0.167	0.157	0.942	0.076	0.152	0.143	0.960	-0.068	0.176	0.177	0.930

[*] Note: "Direct maximization (a)" refers to the method that maximizes the likelihood directly with the existing software, in which the initial values were set to be the same as those used in the proposed EM algorithm. "Direct maximization (b)" refers to the method that maximizes the likelihood directly with existing software, in which the initial values were set to the true parameter values.

6. Application

We applied the proposed method to a set of real world data arising from the randomized clinical trial conducted by the AIDS Clinical Trial Group Protocol 320 (ACTG 320) [3]. Infected patients in this trial were randomly assigned to receive either a two-drug combined therapy (ZDV and 3TC) or a three-drug combined therapy (ZDV, 3TC, and RTV). HIV-1 RNA level in plasma (copies/ml), measured with the NucliSens assay, was used as a biomarker to evaluate treatment efficacy since it essentially reflects the viral load. However, as introduced in Section 1, the NucliSens assay used to measure the HIV-1 RNA level has detection limits because it can only provide an accurate measure if the RNA level is between 400 and 750,000 copies/ml. Otherwise, the RNA level cannot be measured exactly and is treated as left censoring at 400 or right censoring by 750,000.

In our analysis, we considered 838 patients who completed 24 weeks' follow-up and the primary objective is to conduct the treatment effect comparison through investigating the plasma HIV-1 RNA level. To this end, the response of interest is defined as the change in $\log_{10}$ RNA value between week 0 and week 24, $l_0 - l_{24}$, where l_0 and l_{24} denote the $\log_{10}$ RNA values at week 0 and 24, respectively. In this HIV data set, since the measurements of l_0 for all patients are within the limit of quantification. The corresponding detection limits on the log scale are $\log_{10}(400) = 2.60$ and $\log_{10}(750{,}000) = 5.88$. Therefore, $l_0 - l_{24}$ was either left-censored by $L = l_0 - 5.88$ or right-censored by $R = l_0 - 2.60$ if the measurement of l_{24} was out of the quantification limit. In other words, we had doubly censored data on $l_0 - l_{24}$, with the left and right censoring proportions being 1.67% and 29.12%, respectively. To conduct our regression analysis, the covariates used in the probit regression model (1.1) consisted of the treatment indicator (trt = 1 for the three-drug combination group and 0 otherwise) and the baseline $\log_{10}$ RNA value (l_0).

To implement the proposed method, we set the initial value of each regression parameter to 0, and initialized each spline coefficient with 1 as in the simulation studies. We considered linear, quadratic, and cubic monotone splines, and placed 3–20 interior knots at equally spaced quantiles of the interval formed by the realizations of $\widetilde{T}$, where $\widetilde{T} = \max\{L, \min(T, R)\}$. We used AIC and BIC to select the optimal model with respect to the spline degree and number of interior knots. The model selection results indicate that the optimal model under both AIC and BIC is the one with cubic splines and 12 interior knots.

The analysis results of the proposed method with the optimal probit model were summarized in Table 3, which includes the degree and interior knot number of the optimal model, AIC value, BIC value, the estimated covariate effect (Est), the estimated standard error (SE) and the p-value. Results show that both baseline RNA value (l_0) and treatment (trt) have significant influences on the change in $\log_{10}$ RNA value between week 0 and week 24 ($l_0 - l_{24}$). In particular, larger l_0 corresponds to a larger change of $\log_{10}$ RNA value. Compared to two-drug therapy, receiving three-drug therapy leads to a larger $l_0 - l_{24}$, implying that three-drug therapy is beneficial to reduce the RNA value in plasma. Notably, Li et.al [3] analyzed the same HIV dataset under the PH and PO models. Their results also indicated that the effect of baseline RNA level is significant, and the 3-drug combination (ZDV + 3TC + RTV) is more effective than the 2-drug combination (ZDV + 3TC) in reducing plasma HIV-1 RNA levels. In addition, we note that, under different specifications of degree and knot number considered here, the obtained conclusions are consistent and the point estimates are also very close with the largest absolute discrepancy less than 0.052.

Table 3. Analysis results of the HIV data

					l_0			trt		
Method	$\tilde{l}$	q_n	AIC	BIC	Est	SE	p-value	Est	SE	p-value
Proposed EM algorithm	3	12	1135.29	1210.98	-0.1878	0.0707	<0.001	-1.1720	0.0951	<0.001
Direct maximization (a)	3	4	1292.78	1330.63	-0.7324	0.0534	<0.001	-1.3334	0.0808	<0.001

* Note: "Direct maximization (a)" refers to the method that maximizes the likelihood directly with the existing software. d and q_n denote the degree and number of interior knots in monotone splines, respectively.

We also analyzed the HIV data by performing direct maximization of the observed data likelihood (2.2) with the existing optimization function `optim()` in the R software. This method is abbreviated as "Direct maximization (a)", which sets the initial values of each regression parameter and each spline coefficient to 0 and 1, respectively, as in the proposed method. As above, we used AIC and BIC to select the optimal combination of degree and interior knot number in monotone splines. Through the analysis, it turned out that the optimal model is given by the cubic splines with 4 interior knots. Results shown in Table 3 suggest that regression coefficient estimates were quite different from those of the proposed method, especially for the effect of l_0. The AIC and BIC values of "Direct maximization (a)" are consistently larger than those of the proposed method, This phenomenon suggests a lack-of-fit of "Direct maximization (a)" for the HIV data, and thus the resultant regression coefficient estimates are unreliable.

7. Discussion

Regarding doubly censored data, this work provided a computationally efficient sieve MLE method for the spline-based probit regression model. Tailored to monotonic splines' representation and doubly censored data structure, a stable and reliable EM algorithm was proposed to identify the sieve estimator. The key idea was to utilize some normal and multinomial distributed random variables to conduct data augmentation, offering a tractable complete data likelihood that is easy to maximize. The proposed algorithm is quite reliable and insensitive to the initialization, and we did not encounter the non-convergence cases in the above numerical studies. The proposed estimator was theoretically justified by investigating its asymptotic properties with empirical process techniques and sieve estimation theory. Numerical results obtained from simulations and real data analysis highlighted the proposed method's satisfactory performance and advantage over the method that performs direct maximization of the observed data likelihood. Overall, the spline-based probit modeling framework, combined with the proposed EM algorithm, provides an useful alternative for analyzing doubly censored data.

Our proposed method could be extended to several research directions. First, in many practical applications, some individuals may not experience the failure event of interest even though the follow-up is sufficiently long. These individuals constitute a cured subgroup in the whole population under study [26]. It is helpful to generalize the proposed method to estimate a mixture cure rate probit model. Second, the proposed method was built upon an independent censoring assumption, where the censoring and failure times are conditionally independent given the covariates. However, this assumption may be fragile since censoring may be driven by some response-related reasons. To accommodate informative censoring, one may consider extending the proposed method by using a frailty-based method [27] or a copula-based framework [28]. Furthermore, multivariate failure time data are also frequently confronted in survival analysis [29]. To our knowledge, fitting a probit regression model to multivariate failure time data under double censoring is still unexplored and warrants a future investigation.

Author contributions

All authors have worked equally to write and review the manuscript. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they did not use AI tools in this article.

Funding statement

This research was partially supported by the National Nature Science Foundation of China (Grant No. 12471251), the Nature Science Foundation of Guangdong Province of China (Grant No. 2025A1515012851) and the Science and Technology Program of Guangzhou of China (Grant No. 2025A03J3083).

Conflict of interest

The authors declare no conflict of interest.

References

1. E. A. Gehan, A generalized two-sample Wilcoxon test for doubly censored data, *Biometrika*, **52** (1965), 650–653. https://doi.org/10.2307/2333721
2. B. W. Turnbull, Nonparametric estimation of a survivorship function with doubly censored data, *J. Amer. Stat. Assoc.*, **69** (1974), 169–173. https://doi.org/10.1080/01621459.1974.10480146
3. S. Li, T. Hu, P. Wang, J. Sun, A class of semiparametric transformation models for doubly censored failure time data, *Scandinavian J. Stat.*, **45** (2018), 682–698. https://doi.org/10.1111/sjos.12319
4. M. N. Chang, Weak convergence of a self-consistent estimator of the survival function with doubly censored data, *Ann. Stat.*, **18** (1990), 391–404.
5. M. G. Gu, C. H. Zhang, Asymptotic properties of self-consistent estimators based on doubly censored data, *Ann. Stat.*, **21** (1993), 611–624.
6. M. D. Hughes, Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials, *Stat. Med.*, **19** (2000), 3171–3191. https://doi.org/10.1002/1097-0258(20001215)19:23¡3171::AID-SIM619¿3.0.CO;2-T
7. C. H. Zhang, X. Li, Linear regression with doubly censored data, *Ann. Stat.*, **24** (1996), 2720–2743. https://doi.org/10.1214/aos/1032181177
8. J. J. Ren, M. Gu, Regression M-estimators with doubly censored data, *Ann. Stat.*, **25** (1997), 2638–2664. https://doi.org/10.1214/aos/1030741089
9. Y. Kim, B. Kim, W. Jang, Asymptotic properties of the maximum likelihood estimator for the proportional hazards model with doubly censored data, *J. Multivar. Anal.*, **101** (2010), 1339–1351. https://doi.org/10.1016/j.jmva.2010.01.010
10. P. S. Shen, The Cox-Aalen model for doubly censored data, *Commun. Stat.-Theory Meth.*, **51** (2022), 8075–8092. https://doi.org/10.1080/03610926.2021.1887241
11. T. Cai, S. Cheng, Semiparametric regression analysis for doubly censored data, *Biometrika*, **91** (2004), 277–290. https://doi.org/10.1093/biomet/91.2.277
12. S. Ji, L. Peng, Y. Cheng, H. Lai, Quantile regression for doubly censored data, *Biometrics*, **68** (2012), 101–112. https://doi.org/10.1111/j.1541-0420.2011.01667.x
13. S. Li, T. Hu, T. Tong, J. Sun, Semiparametric regression analysis of multivariate doubly censored data, *Stat. Model.*, **20** (2020), 502–526. https://doi.org/10.1177/1471082X19859
14. Y. Deng, S. Li, L. Sun, X. Song, Semiparametric probit regression model with general interval-censored failure time data, *J. Comput. Graphical Stat.*, **33** (2024), 1413–1423. https://doi.org/10.1080/10618600.2024.2330523
15. X. Lin, L. Wang, A semiparametric probit model for case 2 interval-censored failure time data, *Stat. Med.*, **29** (2010), 972–981. https://doi.org/10.1002/sim.3832
16. X. Lin, L. Wang, Bayesian proportional odds models for analyzing current status data: univariate, clustered, and multivariate, *Commun. Stat.-Simul. Comput.*, **40** (2011), 1171–1181. https://doi.org/10.1080/03610918.2011.566971

17. H. Liu, J. Qin, Semiparametric probit models with univariate and bivariate current-status data, *Biometrics*, **74** (2018), 68–76. https://doi.org/10.1111/biom.12726
18. M. Du, T. Hu, J. Sun, Semiparametric probit model for informative current status data, *Stat. Med.*, **38** (2019), 2219–2227. https://doi.org/10.1002/sim.8106
19. L. Fang, S. Li, L. Sun, X. Song, Semiparametric probit regression model with misclassified current status data, *Stat. Med.*, **42** (2023), 4440–4457. https://doi.org/10.1002/sim.9869
20. L. Fang, T. Hu, S. Li, L. Wang, C. S. McMahan, J. M. Tebbs, Probit time-to-event regression for misclassified grouptesting data, *Stat. Sinica*, **37** (2024), 1–25.
21. Y. T. Huang, T. Cai, Mediation analysis for survival data using semiparametric probit models, *Biometrics*, **72** (2016), 563–574. https://doi.org/10.1111/biom.12445
22. J. Ramsay, Monotone regression splines in action, *Stat. Sci.*, **3** (1988), 425–441.
23. X. Chen, T. Hu, J. Sun, Sieve maximum likelihood estimation for the proportional hazards model under informative censoring, *Comput. Stat. Data Anal.*, **112** (2017), 224–234. https://doi.org/10.1016/j.csda.2017.03.006
24. J. Huang, Asymptotic properties of nonparametric estimation based on partly interval-censored data, *Stat. Sinica*, **9** (1999), 501–519.
25. D. Zeng, F. Gao, D. Y. Lin, Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data, *Biometrika*, **104** (2017), 505–525. https://doi.org/10.1093/biomet/asx029
26. J. Zhou, J. Zhang, W. Lu, Computationally efficient estimation for the generalized odds rate mixture cure model with interval-censored data, *J. Comput. Graphical Stat.*, **27** (2018), 48–58. https://doi.org/10.1080/10618600.2017.1349665
27. S. Li, T. Hu, P. Wang, J. Sun, Regression analysis of current status data in the presence of dependent censoring with applications to tumorigenicity experiments, *Comput. Stat. Data Anal.*, **110** (2017), 75–86. https://doi.org/10.1016/j.csda.2016.12.011
28. L. Ma, T. Hu, J. Sun, Sieve maximum likelihood regression analysis of dependent current status data, *Biometrika*, **102** (2015), 731–738. https://doi.org/10.1093/biomet/asv020
29. D. Zeng, Q. Chen, J. G. Ibrahim, Gamma frailty transformation models for multivariate survival times, *Biometrika*, **96** (2009), 277–291. https://doi.org/10.1093/biomet/asp008
30. A. W. Van Der Vaart, J. A. Wellner, Weak convergence, In: *Weak Convergence and Empirical Processes* , 1996, 16–28.
31. X. Shen, W. H. Wong, Convergence rate of sieve estimates, *Ann. Stat.*, **22** (1994), 580–615.
32. M. R. Kosorok, Introduction to empirical processes and semiparametric inference, Berlin: Springer, 2008. https://doi.org/10.1007/978-0-387-74978-5_15
33. P. J. Bickel, C. A. J. Klaassen, Y. Ritov, J. A. Wellner, Efficient and adaptive estimation for semiparametric models, Berlin: Springer, 1993.
34. C. Yuan, S. Zhao, S. Li, X. Song, Sieve maximum likelihood estimation of partially linear transformation models with interval-censored data, *Stat. Med.*, **43** (2024), 5765–5776. https://doi.org/10.1002/sim.10225

Appendix A: A summary table of symbols used in this paper

Symbol	Description
T	The failure time of interest.
$\widetilde{T}$	The observed time, defined as $\max\{L, \min(T, R)\}$.
$(L, R]$	The interval within which T can be observed exactly.
$\delta_1, \delta_2, \delta_3$	The indicators for left censoring, exact, and right censoring observations, respectively.
$I(\cdot)$	The indicator function.
$\mathcal{D}$	The observed data.
$\boldsymbol{X}$	The covariate vector.
p	The dimension of $\boldsymbol{X}$.
$[\tau_u, \tau_v]$	The union support of the random variables T, L, and R.
$F(t \mid \boldsymbol{X})$	The cumulative distribution function of T given $\boldsymbol{X}$.
$\lambda(t \mid \boldsymbol{X})$	The hazard function of T given $\boldsymbol{X}$.
$\Phi(\cdot)$	The cumulative distribution function of a standard normal variable.
$\phi(\cdot)$	The probability density function of a standard normal variable.
$\boldsymbol{\beta}$	The vector of regression parameters.
$\boldsymbol{\beta}_0$	The true value of the regression parameters vector.
$\hat{\boldsymbol{\beta}}$	The proposed estimator of $\boldsymbol{\beta}$.
$\alpha(\cdot)$	An unspecified, strictly increasing function with $\alpha(0) = -\infty$ and $\alpha(\infty) = \infty$.
$\alpha_0(\cdot)$	The true value of $\alpha(\cdot)$.
$\hat{\alpha}(\cdot)$	The proposed estimator of $\alpha(\cdot)$.
$\boldsymbol{\theta}$	The full parameter vector.
$\mathcal{B}$	The parameter space for $\boldsymbol{\beta}$.
$\mathcal{A}$	The function space for $\alpha(\cdot)$.
Θ	The parameter space for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha)$.
L	The observed data likelihood function.
L_1	The augmented likelihood function.
L_c	The complete data likelihood function (with latent variables).
$\alpha_n(\cdot)$	Monotone splines used to approximate $\alpha(\cdot)$.
$b_k(\cdot)$	The kth integrated spline basis function.
$m_k(\cdot)$	The first derivative of the kth integrated spline basis function.
$\tilde{l}$	The degree of the spline.
q_n	The number of interior knots.
K_n	The total number of basis functions, $K_n = q_n + \tilde{l}$.
ξ	The intercept term in the spline representation of α_n.
γ_k	The non-negative spline coefficient for the kth basis function.
$\boldsymbol{\gamma}$	The vector of spline coefficients, $(\gamma_1, \ldots, \gamma_{K_n})^{\mathrm{T}}$.
$\boldsymbol{\zeta}$	The full spline parameter vector, $(\xi, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$.
$\mathcal{A}_n$	The sieve space for $\alpha_n(\cdot)$.
Θ_n	The sieve parameter space for $\boldsymbol{\theta}_n = (\boldsymbol{\beta}, \alpha_n)$.
Z_i	A latent variable for the ith subject.
$\boldsymbol{u}_i$	A multinomial latent vector for the ith subject, $\boldsymbol{u}_i \sim M(1, \boldsymbol{p}_i)$.
$\|\| \cdot \|\|$	The Euclidean norm.
$\|\| \cdot \|\|_2$	The L_2-norm with respect to a probability measure.
$\rho(\cdot, \cdot)$	The metric combining L_2 and Euclidean norms.

Appendix B: Proofs of asymptotic properties

Define $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha)$, $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \alpha_0)$, $\boldsymbol{\theta}_{n0} = (\boldsymbol{\beta}, \alpha_{n0})$, where α_{n0} is true value of the spline function α_n, and $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}, \hat{\alpha}_n)$. Under the proposed semiparametric probit model and doubly censored data, the log-likelihood function for the ith individual is

$$\ell_i(\boldsymbol{\theta}) = \delta_{1i} \log\{\Phi(\alpha(t_i) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)\} + \delta_{2i} \log\{\phi(\alpha(t_i) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)\alpha'(t_i)\} + \delta_{3i} \log\{1 - \Phi(\alpha(t_i) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)\}. \quad \text{(B.1)}$$

Let the negative log-likelihood function be $h_i(\boldsymbol{\theta}) = -\ell_i(\boldsymbol{\theta})$. We define the empirical and expected negative log-likelihood functions as $H_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n h_i(\boldsymbol{\theta})$ and $H(\boldsymbol{\theta}) = E\,[h_i(\boldsymbol{\theta})]$, respectively. This transformation facilitates the application of standard optimization theory while preserving the equivalence between maximizing the log-likelihood $L_n(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$ and minimizing $H_n(\boldsymbol{\theta})$.

Additionally, for any $\varepsilon > 0$, let $\mathcal{N}_{[]}(\varepsilon, \mathcal{F}, \rho)$ denote the ε-bracketing number, and $\mathcal{N}(\varepsilon, \mathcal{F}, \rho)$ be the covering number, with respect to a metric (or semimetric) ρ for a function class $\mathcal{F}$. To prove Theorem 3.1, we need the following lemma.

Lemma 1. *Suppose that conditions (A1)–(A6) hold, we have* $\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = o_p(1)$

Proof. For any $\mu > 0$, let $\mathcal{N}_\mu = \{\boldsymbol{\theta}_n \in \Theta_n : \rho(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) \ge \mu\}$ denote the complement of a neighborhood around the true sieve approximation $\boldsymbol{\theta}_{n0}$. We decompose $H(\boldsymbol{\theta}_n)$ over $\mathcal{N}_\mu$ as

$$\inf_{\boldsymbol{\theta}_n \in \mathcal{N}_\mu} H(\boldsymbol{\theta}_n) = \inf_{\boldsymbol{\theta}_n \in \mathcal{N}_\mu} \{[H(\boldsymbol{\theta}_n) - H_n(\boldsymbol{\theta}_n)] + H_n(\boldsymbol{\theta}_n)\} \le \sup_{\boldsymbol{\theta}_n \in \Theta_n} |H_n(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}_n)| + \inf_{\boldsymbol{\theta}_n \in \mathcal{N}_\mu} H_n(\boldsymbol{\theta}_n).$$

If $\hat{\boldsymbol{\theta}}_n \in \mathcal{N}_\mu$, we can obtain

$$\inf_{\boldsymbol{\theta}_n \in \mathcal{N}_\mu} H_n(\boldsymbol{\theta}_n) = H_n(\hat{\boldsymbol{\theta}}_n) \le H_n(\boldsymbol{\theta}_{n0}) = [H_n(\boldsymbol{\theta}_{n0}) - H(\boldsymbol{\theta}_{n0})] + H(\boldsymbol{\theta}_{n0}).$$

Under Conditions (A1), we have

$$\inf_{\boldsymbol{\theta}_n \in \mathcal{N}_\mu} H(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}_{n0}) \ge \delta_\mu > 0,$$

where δ_μ is a positive constant. Combining these inequalities yields

$$\delta_\mu \le \sup_{\boldsymbol{\theta}_n \in \Theta_n} |H_n(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}_n)| + [H_n(\boldsymbol{\theta}_{n0}) - H(\boldsymbol{\theta}_{n0})].$$

Under Conditions (A1)–(A5), according to Lemma 3 of [23] and the law of large numbers, we have

$$\sup_{\boldsymbol{\theta}_n \in \Theta_n} |H_n(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}_n)| = o_p(1), \quad H_n(\boldsymbol{\theta}_{n0}) - H(\boldsymbol{\theta}_{n0}) = o_p(1).$$

Consequently, we obtain $\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) = o_p(1)$. Furthermore, Lemma 1 of [23] ensures that $\rho(\boldsymbol{\theta}_{n0}, \boldsymbol{\theta}_0) \to o_p(1)$, leading to $\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = o_p(1)$. □

Proof of Theorem 3.1. We use Theorem 3.4.1 of [30] to derive the convergence rate of the proposed estimators. First, For any $0 < \varepsilon \le \mu$, define the function classes

$$\mathcal{Q}_1 = \{\alpha(t) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} : \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n\},$$

$$
\begin{aligned}
\mathcal{Q}_2 &= \{\log\{\Phi[\alpha(t)+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}]\} : \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n\},\\
\mathcal{Q}_3 &= \{\phi[\alpha(t)+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}]\alpha'(t) : \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n\},\\
\mathcal{Q}_4 &= \{1-\Phi[\alpha(t)+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}] : \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n\}, \quad \text{and}\\
\mathcal{L}_\mu &= \{\ell_i(\boldsymbol{\theta}_n) - \ell_i(\boldsymbol{\theta}_{n0}) : \rho(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) \le \mu\}.
\end{aligned}
$$

Recall the function class $\mathcal{A}_n = \{\alpha_n(t) = \xi + \sum_{k=1}^{K_n} \gamma_k b_k(t); b_k(t) \in [0,1], t \in [\tau_u, \tau_v], -M_\alpha \le \xi \le M_\alpha, 0 \le \gamma_k \le M_\alpha\}$, where M_α is a positive constant. Following the calculations in [31], the logarithm of bracketing number satisfies $\log N_{[]}(\varepsilon, \mathcal{A}_n, \|\cdot\|_2) \le K_n \log(\mu/\varepsilon)$. Moreover, the neighborhood $\mathcal{B}_\mu = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \mu\}$ in R^d can be covered by $(\mu/\varepsilon)^d$ balls of radius ε.

There exists a finite constant $M > 0$ such that the logarithmic bracketing number of $\mathcal{Q}_1$ is $K_n \log(\mu/\varepsilon) + d\log(\mu/\varepsilon) \le M(K_n + d)\log(\mu/\varepsilon) \asymp MK_n \log(\mu/\varepsilon)$, in which $\asymp$ means both sides have same order. Analogously, the bracketing number bounds for the remaining function classes $\mathcal{Q}_2$, $\mathcal{Q}_3$, and $\mathcal{Q}_4$ also satisfy $N_{[]}(\varepsilon, \mathcal{Q}_i, \rho) \asymp MK_n \log(\mu/\varepsilon)$ for $i = 2, 3, 4$. By Lemma 9.25 of [32], we have $\log N_{[]}(\varepsilon, \mathcal{L}_\mu, \|\cdot\|_2) \le MK_n \log(\mu/\varepsilon)$. Hence, the bracketing integral of function class $\mathcal{L}_\mu$ is

$$
J_{[]}(\mu, \mathcal{L}_\mu, \rho) = \int_0^{\mu_0} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{L}_\mu, \rho)}\, d\varepsilon \le \widetilde{M} K_n^{1/2} \mu,
$$

where $\widetilde{M}$ is a positive constant. Then, by Lemma 3.4.2 of [30], we have

$$
\begin{aligned}
&E\left(\sup_{\rho(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) < \mu} \left| \sqrt{n}(L_n - L)(\boldsymbol{\theta}_n) - \sqrt{n}(L_n - L)(\boldsymbol{\theta}_{n0}) \right| \right)\\
\le & J_{[]}(\mu, \mathcal{L}_\mu, \rho)\left(1 + \frac{J_{[]}(\mu, \mathcal{L}_\mu, \rho)}{\mu^2 \sqrt{n}}\right) \le \widetilde{M} K_n^{1/2} \mu \left(1 + \frac{\widetilde{M} K_n^{1/2} \mu}{\mu^2 \sqrt{n}}\right) = O(K_n^{1/2}\mu + K_n/n^{1/2}).
\end{aligned}
$$

In particular, the key function $\phi(\mu)$ defined in Theorem 3.4.1 of [30] is given by $\phi_n(\mu) = K_n^{1/2}\mu + K_n/n^{1/2}$. It can be readily verified that $\phi_n(\mu)/\mu$ is monotonically decreasing in μ. Furthermore, for the scaling sequence $r_n = (n/K_n)^{1/2} = n^{(1-\nu)/2}$, we have:

$$
r_n^2 \phi_n(1/r_n) = r_n K_n^{1/2} + \frac{r_n^2 K_n}{n^{1/2}} \le O(n^{1/2}).
$$

The estimation error between the estimator $\hat{\boldsymbol{\theta}}_n$ and the projected true value $\boldsymbol{\theta}_{n0}$ satisfies

$$
\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) = (K_n/n)^{1/2}.
$$

Combining this with the spline approximation error $\rho\left(\boldsymbol{\theta}_{n0}, \boldsymbol{\theta}_0\right) = O\left(K_n^{-r}\right)$, where the parameter r is defined in Condition (A5), we have

$$
\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = \rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) + \rho(\boldsymbol{\theta}_{n0}, \boldsymbol{\theta}_0) = O(n^{-(1-\nu)/2} + n^{-r\nu}).
$$

When $\nu = 1/(2r+1)$, $\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O(n^{-r/(2r+1)})$, which completes the proof.

□

Proof of Theorem 3.2. For a single subject, the score function for $\boldsymbol{\beta}$ can be derived from the log-likelihood function in equation B.1 as

$$\dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{X}Q(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta}),$$

where

$$Q(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta}) = \delta_1 \frac{\phi[\alpha(\widetilde{T}) + \boldsymbol{\beta}^\top \boldsymbol{X}]}{\Phi[\alpha(\widetilde{T}) + \boldsymbol{\beta}^\top \boldsymbol{X}]} - \delta_2[\alpha(\widetilde{T}) + \boldsymbol{\beta}^\top \boldsymbol{X}] - \delta_3 \frac{\phi[\alpha(\widetilde{T}) + \boldsymbol{\beta}^\top \boldsymbol{X}]}{1 - \Phi[\alpha(\widetilde{T}) + \boldsymbol{\beta}^\top \boldsymbol{X}]}.$$

Consider the submodel $\alpha_\eta(t) = \alpha(t) + \eta z(t)$ for a small constant η around 0, where $z(\cdot)$ is a nondecreasing function. By differentiating $\ell(\boldsymbol{\theta})$ with respect to η and evaluating the derivative at $\eta = 0$, we can obtain the score function for $\alpha(\cdot)$, taking the form

$$\dot{\ell}_\alpha(\boldsymbol{\theta})[z] = z(\widetilde{T})Q(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta}).$$

By following the arguments of Theorem 3.4.1 of [33], one can derive the information bound for $\boldsymbol{\beta}$ using the efficient score method. First, projecting $\dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ onto the linear space spanned by the score functions $\dot{\ell}_\alpha(\boldsymbol{\theta})[z]$. Next, identify a function z^* such that the efficient score $\dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \dot{\ell}_\alpha(\boldsymbol{\theta})[z^*]$ is orthogonal to $\dot{\ell}_\alpha(\boldsymbol{\theta})[z]$ for all $z(\cdot)$ with bounded variation. This implies

$$E_0\{[\dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \dot{\ell}_\alpha(\boldsymbol{\theta})[z^*]]\dot{\ell}_\alpha(\boldsymbol{\theta})[z]\} = E_0\{[\boldsymbol{X} - z^*(\widetilde{T})]z(\widetilde{T})Q^2(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta})\} = 0.$$

Thus, the solution for z^* is

$$z^* = \frac{E_0[\boldsymbol{X}Q^2(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta})]}{E_0[Q^2(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta})]},$$

and the efficient score function for $\boldsymbol{\beta}$ is given by

$$\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left\{\boldsymbol{X} - \frac{E_0[\boldsymbol{X}Q^2(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta})]}{E_0[Q^2(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta})]}\right\} Q(\boldsymbol{X}, \widetilde{T}; \boldsymbol{\theta}).$$

By following the arguments of [34] and others, we have

$$\mathbb{G}_n[\tilde{\ell}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})] = n^{1/2}\mathbf{I}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O(n^{1/2}\rho(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0)),$$

where $\mathbf{I}(\boldsymbol{\beta}_0) = E[\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)^\top]$, and the remainder term is ignorable.

Thus, to prove the asymptotic normality of $\hat{\boldsymbol{\beta}}$, we need to verify the following two conditions:
(C1) The efficient score $\tilde{\ell}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ belongs to a Donsker class and converges to $\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0) = \dot{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$ -$\dot{\ell}_\alpha(\boldsymbol{\theta}_0)[z^*]$ in the $L_2(\mathbb{P})$-norm;
(C2) The matrix $\mathbf{I}(\boldsymbol{\beta}_0) = E[\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)^\top]$ is nonsingular.

To verify Condition (C1), analogous to the proof of Lemma 1, we can establish that, for any $\epsilon > 0$, the class

$$\{\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_n) : \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n, \rho(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \le \epsilon\}$$

has a bounded ε-bracketing number $M\log(\mu/\varepsilon)$, which implies that it is Donsker. Similarly, it can be shown that the classes

$$\{\dot{\ell}_\alpha(\boldsymbol{\theta}_n)[z^*] : \boldsymbol{\beta} \in \mathcal{B}, \alpha_n \in \mathcal{A}_n, \rho(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \le \epsilon\}$$

is also Donsker. By the preservation property of Donsker classes, $\tilde{\ell}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ belongs to a Donsker class. Applying Theorem 1, we conclude that $\tilde{\ell}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ converges to $\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$ in the L_2-norm. This implies that $\mathbb{G}_n[\tilde{\ell}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})]$ converges in distribution to a zero-mean d-variate normal random vector.

Finally, we demonstrate that $\mathbf{I}(\boldsymbol{\beta}_0)$ is nonsingular. Suppose that the matrix $\mathbf{I}(\boldsymbol{\beta}_0)$ is singular, there exists a nonzero vector $\tilde{\chi}$ such that

$$\tilde{\chi}^\top \mathbf{I}(\boldsymbol{\beta}_0)\tilde{\chi} = \tilde{\chi}^\top E\{\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\,\tilde{\ell}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)^\top\}\tilde{\chi} = 0.$$

This implies that the score function along the submodel $(\boldsymbol{\beta}_0 + \kappa\tilde{\chi}, \alpha_0 + \kappa\tilde{\chi}^\top z^*)$ is zero with probability 1, where κ is any constant, that is,

$$Q(\boldsymbol{X}, T; \boldsymbol{\theta}_0)[\tilde{\chi}^\top \boldsymbol{X} + \tilde{\chi}^\top z^*] = 0.$$

In particular, by considering $\delta_1 = 1$, we have $Q(\boldsymbol{X}, T; \boldsymbol{\theta}_0) = \frac{\phi[\alpha(\widetilde{T})+\boldsymbol{\beta}^\top \boldsymbol{X}]}{\Phi[\alpha(\widetilde{T})+\boldsymbol{\beta}^\top \boldsymbol{X}]} \in (0, \infty)$. Consequently, $\tilde{\chi}^\top \boldsymbol{X} + \tilde{\chi}^\top z^* = 0$ with probability 1. By Condition (A7), we have $\tilde{\chi} = 0$, which contradicts the assumption that $\tilde{\chi}$ is a nonzero vector. Therefore, $\mathbf{I}(\boldsymbol{\beta}_0)$ is nonsingular, and we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\beta}_0)), \tag{B.2}$$

which completes this proof. □

AIMS Press