SCIENCE CHINA
Mathematics

• ARTICLES •

August 2018 Vol. 61 No. 8: 1495–1518
https://doi.org/10.1007/s11425-017-9268-0

An embedded estimating equation for the additive risk model with biased-sampling data

Feipeng Zhang[1,*], Xingqiu Zhao[2] & Yong Zhou[3,4]

[1] *School of Mathematics and Statistics, Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education), Hunan Normal University, Changsha 410081, China;*
[2] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China;*
[3] *School of Statistics, East China Normal University, Shanghai 200241, China;*
[4] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

Email: zhangfp108@163.com, xingqiu.zhao@polyu.edu.hk, yzhou@amss.ac.cn

Received September 14, 2017; accepted Feburary 6, 2018; published online June 25, 2018

Abstract This paper presents a novel class of semiparametric estimating functions for the additive model with right-censored data that are obtained from general biased-sampling. The new estimator can be obtained using a weighted estimating equation for the covariate coefficients, by embedding the biased-sampling data into left-truncated and right-censored data. The asymptotic properties (consistency and asymptotic normality) of the proposed estimator are derived via the modern empirical processes theory. Based on the cumulative residual processes, we also propose graphical and numerical methods to assess the adequacy of the additive risk model. The good finite-sample performance of the proposed estimator is demonstrated by simulation studies and two applications of real datasets.

Keywords additive risk model, biased-sampling data, missing covariates, estimating equation, model checking

MSC(2010) 62N02, 62N03

Citation: Zhang F P, Zhao X Q, Zhou Y. An embedded estimating equation for the additive risk model with biased-sampling data. Sci China Math, 2018, 61: 1495–1518, https://doi.org/10.1007/s11425-017-9268-0

1 Introduction

The biased-sampling data are frequently encountered in economics, epidemiology, and medical studies. They arise from complex surveys due to design or the data collecting mechanisms. A prominent feature of biased data is that the observed subjects are selected with data-dependent sampling probabilities, rather than being randomly sampled from the population. For example, during the prevalent cohort sampling of heart patients, the data included only diseased subjects who have not experienced the failure events before the recruitment. Thus, the subjects who experienced a failure event before the recruitment time cannot be observed; therefore, the observed survival time is subject to left-truncated. Ignoring the left truncation may substantially bias estimation and falsify the inference. There is a rich literature on left-truncated data, including nonparametric estimators of the survivor function (see [21, 45, 47, 49, 54]), and semiparametric regression models (see [11, 22, 46, 53] and among others). One important sampling scheme

* Corresponding author

is the length-biased data, in which the selection probabilities of the sampled subjects are proportional to the lengths of their survival times. The selection bias in length-biased data, is usually handled by the nonparametric or semiparametric models. The nonparametric estimators of the survival function under length-biased sampling were developed by [3, 13, 30, 50], and among others. Other estimation procedures for length-biased data are based on the Cox model (see [10,15,38,39,44,52,53,57]), the accelerated failure time (AFT) model, and the transformation model (see [7,19,42]). The nonparametric and semiparametric regression models under length-biased sampling are comprehensively reviewed in [43].

Another common sampling scheme is the missing covariates in the data. This occurs when the study design inherently or accidentally excludes some components of the covariates. A special case of missing covariates is the case-cohort design with Bernoulli sampling, in which some data are observed to be missing. The complete-cohort analysis, which discards all the subjects with missing covariates, may be inefficient and yield biased estimators when the missing data mechanism depends on the outcome variables and the observed covariates. To improve the complete-cohort analysis, researchers have developed various methods that incorporate the partially incomplete data into the analysis. The estimation procedures for data with missing covariates have been investigated in the proportional hazards model (see [6, 31, 35, 37, 51, 55]), in the additive model (see [12, 27]), and in the AFT model and the transformation model (see [17, 18, 20, 28]).

The current paper considers the additive risk model, which captures the risk difference rather than the risk ratio in the Cox model. The additive risk model with right-censored data in various forms has been studied by several authors (see [1, 5, 8, 26]), and the semiparametric estimation methods for this model under biased-sampling data have also received considerable attention in the literature. In the existing literature, the additive risk model with left-truncated and right-censored data, have been estimated using a conditional estimating equation estimator by [26], the empirical likelihood estimator by [29] and a semiparametric estimation proposed by [14] via combining the marginal pairwise pseudo-score function and the conditional estimating equation. Ma et al. [32] and Zhao et al. [58] proposed some estimating equation estimators for the additive risk model with length-biased and right-censored data. Li [23] proposed a unified approach for the additive risk model with the general biased sampling data. They used the weight function proposed by [17] in the transformation model. However, their method assumes that the data are right-censored before the biased sampling. Such an assumption is not always realistic in practice. This paper proposes a new class of semiparametric estimating functions for the additive model with censored data that are collected by general biased-sampling. In these data, the censoring can occur either before or after the biased sampling. The proposed approach is based on [44], in which the biased sampling data in the Cox proportional hazard model have been investigated. Regardless, his pseudo-partial likelihood approach for the Cox model cannot be directly used to eliminate the nuisance function (i.e., the baseline hazard function) in estimating the regression coefficients. This is the major challenge while performing statistical inferences for the additive risk model.

The contribution of this paper is twofold. First, we propose an embedded estimating equation from the conditional estimating equation, via embedding the biased data into a left-truncated and right-censored model. Compared with the conditional estimating equation estimator, the embedded estimating equation estimator is more efficient because it incorporates the latent information in the biased-sampling data. The limiting distributions of the proposed estimator are established using the modern empirical processes theory. Second, we develop model checking techniques to assess the adequacy of the additive risk model. We propose a class of numerical and graphical methods based on the cumulative residual processes. Under the assumed model, the cumulative residual process weakly converges to a zero-mean Gaussian process, whose distribution can be approximated using a simulation technique. The observed processes pattern can be compared both visually and analytically to a large number of realizations from the approximated processes. These comparisons reveal the various purposes of model fitting evaluation, including the functional form of each covariate, the assumption of additive hazards with respect to each component, and the adequacy of the overall model.

The remainder of this paper is organized as follows. Section 2 introduces the embedded estimating equation estimator and derives its large-sample properties. The model checking techniques are discussed in Section 3. Section 4 evaluates the performance of the proposed method by conducting simulation studies. Section 5 applies the proposed method to the Shrub data and the Stanford heart transplant data, and Section 6 concludes the paper. The technical proofs are provided in Appendix A.

2 Estimation procedure

2.1 Biased-sampling data and model specification

Given a p-vector of time-dependent covariates $\boldsymbol{Z}(\cdot) = (Z_1(\cdot), \ldots, Z_p(\cdot))^{\mathrm{T}}$, we can assume that the latent lifetime T^* has the conditional distribution function $F(t \,|\, \boldsymbol{Z}(t))$ and the conditional density function $f(t \,|\, \boldsymbol{Z}(t))$. A biased-sampling dataset of $(T^*, \boldsymbol{Z}(\cdot))$ comprises n independent random samples, where the observation of the i-th sample $(T_i, \boldsymbol{Z}_i(\cdot))$ has the conditional density function

$$h(t \,|\, \boldsymbol{Z}_i) = \frac{W(t \,|\, \boldsymbol{Z}_i(t)) f(t \,|\, \boldsymbol{Z}_i(t))}{\mu(\boldsymbol{Z}_i)}, \tag{2.1}$$

where $W(t \,|\, \boldsymbol{Z}_i(t))$ is a known non-negative weight function, and $\mu(\boldsymbol{Z}_i)$ is a normalization constant such that $h(t \,|\, \boldsymbol{Z}_i)$ is a probability density function.

In biased-sampling data, random samples are obtained from f; however, the i-th observation $(T_i, \boldsymbol{Z}_i(T_i))$ is accepted with a probability that is proportional to its weight function $W(T_i \,|\, \boldsymbol{Z}_i(T_i))$. Different biased-sampling data are assigned different weight functions. Some biased-sampling data with their corresponding weight functions $W(t \,|\, \boldsymbol{Z}(t))$ are listed below (see [31]),

1. for the left-truncated data, $W(t \,|\, \boldsymbol{Z}(t)) = I(A \leqslant t)$, where A is the left truncation;
2. for the length-biased data, $W(t \,|\, \boldsymbol{Z}(t)) = t$;
3. in the data with the missing covariates, we can observe $(T, \delta, V, \boldsymbol{Z}(\cdot))$, where $T = \min(T^*, C)$, $\delta = I(T^* \leqslant C)$ and $V = I(\boldsymbol{Z}^m(t)\,\text{is observed})$. $\boldsymbol{Z}(\cdot) = (\boldsymbol{Z}^{m^{\mathrm{T}}}(\cdot), \boldsymbol{Z}^{c^{\mathrm{T}}}(\cdot)^{\mathrm{T}})$, where $\boldsymbol{Z}^c(\cdot)$ and $\boldsymbol{Z}^m(\cdot)$ are complete and missing components of the covariates $\boldsymbol{Z}(\cdot)$, respectively. The weight function is given by $W(T \,|\, \delta, \boldsymbol{Z}(\cdot)) = \mathrm{P}(V = 1 \,|\, T, \delta, \boldsymbol{Z}(\cdot))$. In particular, when the missing data are observed due to the case-cohort design with Bernoulli sampling, the weight function becomes $W(T_i \,|\, \boldsymbol{Z}_i^c(\cdot), \delta_i) = \delta_i + (1 - \delta_i) p_i$, where p_i is the probability of selecting the i-th subject in the subcohort.

Given the time-dependent covariates $\boldsymbol{Z}_i(t) = (Z_{i1}(t), \ldots, Z_{ip}(t))^{\mathrm{T}}$, we assume that T_i^* follows an additive risk model [26],

$$\lambda(t \,|\, \boldsymbol{Z}_i) = \lambda_0(t) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t), \tag{2.2}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients and $\lambda_0(t)$ is the baseline hazard function.

The main objective is to propose an estimator for the true coefficient $\boldsymbol{\beta}_0$. Throughout this paper, the latent and observed values will be denoted by a superscripted and a non-superscripted star, respectively.

2.2 Working estimating equation: Left-truncated and right-censored data

We first consider the left-truncated and right-censored data. Assume that the latent lifetime T^* has the distribution $F(\cdot)$, and the truncation time A^* and censoring time C^* have joint distribution function $G(\cdot, \cdot)$ with the joint density function $g(\cdot, \cdot)$. Given the covariates $\boldsymbol{Z}(\cdot)$, we assume that T^* and (A^*, C^*) are conditionally independent, and that $\mathrm{P}(T^* \geqslant A^*, C^* \geqslant A^*) > 0$. The left-truncated data are not sampled from the joint distribution, rather they are sampled from the condition distribution given the event, $\{T^* \geqslant A^*, C^* \geqslant A^*\}$. In a sample of n independent triples, (A_i, T_i^0, C_i) $(i = 1, \ldots, n)$, are obtained from this conditional distribution, i.e., each triplet (A_i, T_i^0, C_i) has the identical joint distribution as $(A_i^*, T_i^*, C_i^*) \,|\, T^* \geqslant A^*, C^* \geqslant A^*$. We denote our observed data left-truncated and right-censored sample data as $\{A_i, T_i, \delta_i\}_{i=1}^n$, where $T_i = \min(T_i^0, C_i)$, and $\delta_i = I(T_i \leqslant C_i)$.

To proceed, we introduce the counting process $N_i(t) = I(T_i \leqslant t, \delta_i = 1)$ and the risk process $Y_i(t) = I(A_i \leqslant t \leqslant T_i)$. Under Model (2.2), as shown by [2, Subsections 3.3 and 3.4], the intensity function of $N_i(t)$ is given by

$$Y_i(t) d\Lambda(t \,|\, \boldsymbol{Z}_i) = Y_i(t)[d\Lambda_0(t) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt],$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the cumulative baseline hazard function. Thus, the counting process $N_i(\cdot)$ can be uniquely decomposed such that for each t,

$$N_i(t) = M_i^{(L)}(t) + \int_0^t Y_i(u)[d\Lambda_0(u \mid \boldsymbol{Z}_i) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt], \tag{2.3}$$

where $M_i^{(L)}(\cdot)$ is a local square-integrable martingale (see [2]). Under the relation (2.3), $\Lambda_0(t)$ can be naturally estimated using the Breslow's type estimator,

$$\widehat{\Lambda}_0^{(L)}(\widehat{\boldsymbol{\beta}}^{(L)}, t) = \int_0^t \frac{\sum_{i=1}^n [dN_i(u) - Y_i(u)\widehat{\boldsymbol{\beta}}^{(L)^{\mathrm{T}}}\boldsymbol{Z}_i(u)du]}{\sum_{i=1}^n Y_i(u)},$$

where $\widehat{\boldsymbol{\beta}}^{(L)}$ is a consistent estimator of $\boldsymbol{\beta}_0$. Lin and Ying [26] further proposed the estimator $\widehat{\boldsymbol{\beta}}^{(L)}$ by solving the following conditional estimating function:

$$U^{(L)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)[dN_i(t) - Y_i(t)d\widehat{\Lambda}_0^{(L)}(\beta, t) - Y_i(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt],$$

which is equivalent to

$$U^{(L)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(L)}(t)][dN_i(t) - Y_i(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt], \tag{2.4}$$

where τ is the largest follow-up time, and $\overline{\boldsymbol{Z}}^{(L)}(t) = \sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)/\sum_{i=1}^n Y_i(t)$. Solving the conditional estimating equation $U^{(L)}(\boldsymbol{\beta}) = 0$, $\boldsymbol{\beta}_0$ can be estimated as

$$\widehat{\boldsymbol{\beta}}^{(L)} = \left[\sum_{i=1}^n \int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(L)}(t)]^{\otimes 2}dt\right]^{-1}\left[\sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(L)}(t)]dN_i(t)\right],$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$ is the Kronecker product for any vector $\boldsymbol{a}$. Using the empirical process theory, Lin and Ying [26] derived that $n^{1/2}U^{(L)}(\boldsymbol{\beta})$ is the sum of i.i.d. random vectors; therefore it converges in distribution to a zero-mean multivariate normal distribution with a covariance matrix $\Sigma^{(L)}$, where $\Sigma^{(L)}$ can be consistently estimated by $\widehat{\Sigma}^{(L)} = n^{-1}\sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(L)}(t)]^{\otimes 2}dN_i(t)$. By Taylor's series expansion and some mathematical arguments, they also showed that $\sqrt{n}(\widehat{\boldsymbol{\beta}}^{(L)} - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean multivariate normal distribution with a variance matrix $D^{(L)^{-1}}\Sigma^{(L)}D^{(L)^{-1}}$, which can be consistently estimated by $\widehat{D}^{(L)^{-1}}\widehat{\Sigma}^{(L)}\widehat{D}^{(L)^{-1}}$, where

$$\widehat{D}^{(L)} = n^{-1}\sum_{i=1}^n \int_0^\tau Y_i(t)[\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(L)}(t)]^{\otimes 2}dt.$$

Clearly, the estimator $\widehat{\beta}^{(L)}$ for β can be very inefficient in this situation, because the distribution of truncation time is apparently not considered. To overcome this difficulty, we propose a more efficient estimator in the following sections. Our estimator embeds the data into the left-truncated data. Therefore, we treat (2.4) as the working estimating equation for the complete data, and then derive an embedded estimating equation for biased-sampling data based on (2.4).

2.3 Embedded estimating equation: Biased-sampling data

In this section, we consider the data only from biased sampling without censoring. We assume that the latent truncation A^* has the conditional distribution $W(t \mid \boldsymbol{Z}(t)) = \mathrm{P}(A^* \leqslant t \mid \boldsymbol{Z}(t))$, and the lifetime T^* has the survival distribution $S(t \mid \boldsymbol{Z}(t)) = 1 - \mathrm{P}(T^* \leqslant t \mid \boldsymbol{Z}(t))$. We additionally assume that A^* and T^* are independently conditional on $\boldsymbol{Z}(\cdot)$; however, we observe $(A, T^0, \boldsymbol{Z}(\cdot))$ only if $T^0 \geqslant A$. Therefore, given $\boldsymbol{Z}(\cdot) = \boldsymbol{z}$, the conditional density of observing (A, T^0) at (a, t) is $w(a \mid \boldsymbol{z})f(t \mid \boldsymbol{z})I(t \geqslant a > 0)/\mu$, where $w(a \mid \boldsymbol{z}) = \partial W(a \mid \boldsymbol{z})/\partial a$ and $\mu = \mathrm{E}[W(T^* \mid \boldsymbol{z})]$. The density function of the observed T^0 can be then given by

$$\frac{\int w(a \mid \boldsymbol{z}) f(t \mid \boldsymbol{z}) I(t \geqslant a > 0) da}{\mu} = \frac{W(t \mid \boldsymbol{z}) f(t \mid \boldsymbol{z})}{\mu},$$

which takes the same form as the conditional density function in (2.1).

To derive a new class of weighted estimating functions, we follow Tsai [44], who obtained a partial likelihood by embedding the data into a left-truncated and right-censored model. However, we emphasize that our proposed estimator is based on the conditional estimating equation, which is different from the pseudo-partial likelihood method of [44]. Specifically, we treat $(A, T^0, \boldsymbol{Z}(\cdot))$ as the complete data vector and $(T^0, \boldsymbol{Z}(\cdot))$ as the incomplete data with the truncation time A completely missing. We further consider

$$\begin{aligned} M_i(t) &\equiv \mathrm{E}[M_i^{(L)} \mid \text{data}] \\ &= \mathrm{E}\left[N_i(t) - \int_0^t Y_i(u)[d\Lambda_0(u \mid \boldsymbol{Z}_i(u)) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(u) du] \,\middle|\, \text{data}\right] \\ &= N_i(t) - \int_0^t \pi_i(u)[d\Lambda_0(u \mid \boldsymbol{Z}_i(u)) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(u) du], \end{aligned} \tag{2.5}$$

where $\pi_i(t) \equiv \mathrm{E}[Y_i(t) \mid \text{data}] = \frac{W(t \mid \boldsymbol{Z}_i(t))}{W(T_i^0 \mid \boldsymbol{Z}_i(T_i))}$.

Obviously, by the double expectation formula, $M_i(t)$ is a zero-mean process. In the view of (2.5), we can estimate $\Lambda_0(t)$ by a Breslow's type estimator,

$$\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, t) = \int_0^t \frac{\sum_{i=1}^n [dN_i(u) - \pi_i(u) \widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}_i(u) du]}{\sum_{i=1}^n \pi_i(u)}.$$

As $M_i(t)$ is a zero-mean process, we propose the following estimating function:

$$U(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t) dM_i(t).$$

Replacing $M_i(t)$ in the aforementioned function by its empirical counterpart, we can estimate $\boldsymbol{\beta}_0$ using the following estimating function:

$$U(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)[dN_i(t) - \pi_i(t) d\widehat{\Lambda}_0(\boldsymbol{\beta}, t) - \pi_i(t) \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt],$$

which is equivalent to

$$U(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)][dN_i(t) - \pi_i(t) \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt], \tag{2.6}$$

where $\overline{\boldsymbol{Z}}(t) = \sum_{i=1}^n \pi_i(t) \boldsymbol{Z}_i(t) / \sum_{i=1}^n \pi_i(t)$.

Here, $W(t \mid \boldsymbol{Z}(t))$ is the cumulative distribution function of the left-truncation time, which is assumed to be either completely known or be estimated using other methods. For example, if $W(t \mid \boldsymbol{Z}(t))$ is unknown in practice, it can be replaced by its consistent estimator $\widehat{W}(t \mid \boldsymbol{Z}(t))$. The estimating equation can be further derived as

$$\widehat{U}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\widehat{\boldsymbol{Z}}}(t)][dN_i(t) - \widehat{\pi}_i(t) \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt], \tag{2.7}$$

where $\widehat{\pi}_i(t) = \frac{\widehat{W}(t \mid \boldsymbol{Z}_i(t))}{\widehat{W}(T_i^0 \mid \boldsymbol{Z}_i(t))} I(T_i^0 \geqslant t)$ and

$$\overline{\widehat{\boldsymbol{Z}}}(t) = \sum_{i=1}^n \widehat{\pi}_i(t) \boldsymbol{Z}_i(t) \Big/ \sum_{i=1}^n \widehat{\pi}_i(t).$$

Hence, the resulting estimator takes the closed form of

$$\widehat{\boldsymbol{\beta}} = \Big[\sum_{i=1}^{n}\int_{0}^{\tau}\widehat{\pi}_i(t)[\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}(t)]^{\otimes 2}dt\Big]^{-1}\Big[\sum_{i=1}^{n}\int_{0}^{\tau}[\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}(t)]dN_i(t)\Big].$$

2.4 Embedded estimating equation: Right censored biased-sampling data

We now consider right-censored biased-sampling data, which are considerably challenging than uncensored biased-sampling data. Inspired by (2.6), it requires to calculate the weights $\pi_i(t) = \mathrm{E}[Y_i(t)\,|\,\text{data}]$. These weights depend on the censoring mechanisms. As discussed in [44], the section only considers two types of censoring mechanisms.

2.4.1 *Censoring mechanism* I: *Censoring before biased sampling*

In the first type of censoring, the censoring mechanism is applied before sampling the data with bias. Given covariates $\boldsymbol{Z}_i(\cdot)$, T^*, A^* and C^* are assumed to be mutually independent. The observed data (A_i, T_i, δ_i) are conditional on $T_i \geqslant A_i$, where $T_i = \min(T_i^*, C_i^*)$ and $\delta_i = I(T_i^* \leqslant C_i^*)$. Hence, in this type of censoring, given $\boldsymbol{Z}_i(\cdot) = \boldsymbol{z}$, the joint density of (A_i^*, C_i^*) at (a, c) is given by

$$g(a, c\,|\,\boldsymbol{z}) = w(a\,|\,\boldsymbol{z})h(c\,|\,\boldsymbol{z})I(a \leqslant c)/\mu_1,$$

where $\mu_1 = \int_0^\infty\int_0^c w(a\,|\,\boldsymbol{z})h(c\,|\,\boldsymbol{z})dadc$, and $w(a\,|\,\boldsymbol{z}), h(c\,|\,\boldsymbol{z})$ are the probability density functions of A_i^* and C_i^*, respectively. Given $\boldsymbol{Z}_i(\cdot) = \boldsymbol{z}$, the conditional density function of the observed data (A_i, T_i, δ_i) at (a, t, δ) can be further obtained as

$$\phi_1(a, t, \delta\,|\,\boldsymbol{z}) = \frac{w(a\,|\,\boldsymbol{z})\{f(t\,|\,\boldsymbol{z})\overline{H}(t)\}^{\delta}\{S(t\,|\,\boldsymbol{z})h(t)\}^{(1-\delta)}}{\int_0^\infty w(a\,|\,\boldsymbol{z})\overline{H}(a)S(a\,|\,\boldsymbol{z})da},$$

where $\overline{H}(t) = \int_t^\infty h(s)ds$ is the survival function of the censoring time C^*. It follows that

$$E[Y_i(t)\,|\,\text{observed data}] = I(T_i \geqslant t)W(t\,|\,\boldsymbol{Z}_i(T_i))/W(T_i\,|\,\boldsymbol{Z}_i(T_i)),$$

which indicates that the estimation procedures proposed in Subsection 2.2 remain valid. Furthermore, under this type of censoring, the estimating equation (2.7) and the resulting estimator exhibit the same formulation when T_i^0 is replaced by T_i and $N_i(t)$ is defined by $N_i(t) = I(T_i \leqslant t, \delta_i = 1)$. Specifically,

$$\widehat{U}^{(1)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}[\boldsymbol{Z}_i(t) - \overline{\widehat{\boldsymbol{Z}}}^{(1)}(t)][dN_i(t) - \widehat{\pi}_i^{(1)}(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt], \tag{2.8}$$

where $\widehat{\pi}_i^{(1)}(t) = \frac{\widehat{W}(t\,|\,\boldsymbol{Z}_i(t))}{\widehat{W}(T_i\,|\,\boldsymbol{Z}_i(T_i))}I(T_i \geqslant t)$ and $\overline{\widehat{\boldsymbol{Z}}}^{(1)}(t) = \sum_{i=1}^n \widehat{\pi}_i^{(1)}(t)\boldsymbol{Z}_i(t)/\sum_{i=1}^n\widehat{\pi}_i^{(1)}(t)$. The resulting estimator in closed form can be easily derived as presented above,

$$\widehat{\boldsymbol{\beta}}^{(1)} = \Big[\sum_{i=1}^{n}\int_{0}^{\tau}\pi_i^{(1)}(t)\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}^{(1)}(t)\}^{\otimes 2}dt\Big]^{-1}\Big[\sum_{i=1}^{n}\int_{0}^{\tau}\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}^{(1)}(t)\}dN_i(t)\Big].$$

2.4.2 *Censoring mechanism* II: *Censoring after biased sampling*

The second type of censoring mechanism is applying the censoring of residual lifetime after the data are sampled with bias. We define $R_i^* = T_i^* - A_i^*$ and $R_{ci} = C_i^* - A_i^*$ as the residual lifetime and the residual censoring time of subject i, respectively, and apply right censoring to the residual lifetime R_i^* rather than to the lifetime T_i^*. As assumed in [44], we assume that R_{ci} is independent of (R_i^*, A_i^*) given covariates $\boldsymbol{Z}_i$ and $C_i^* \geqslant A_i^*$. The observed data (A_i, T_i, δ_i) satisfy $A_i \leqslant T_i$, where $T_i = A_i + R_i$, $R_i = \min(R_i^*, R_{ci})$ and $\delta_i = I(R_i^* \leqslant R_{ci})$. However, C_i^* and A_i^* are not independent, because $\mathrm{Cov}(C_i^*, A_i^*) = \mathrm{Cov}(A_i^* + R_i^*, A_i^*)$ $= \mathrm{Var}(A_i^*) + \mathrm{Cov}(R_{ci}, A_i^*) \neq 0$. In this type of censoring mechanism, given $\boldsymbol{Z}(\cdot) = \boldsymbol{z}$, the joint density of (A_i^*, C_i^*) at (a, c) can be given by

$$g(a, c\,|\,\boldsymbol{z}) = w(a\,|\,\boldsymbol{z})g_{rc}(c - a\,|\,\boldsymbol{z})I(c \geqslant a),$$

where $g_{rc}(t\,|\,\boldsymbol{z})$ is the density function of the residual censoring time R_{ci}. Then, given $\boldsymbol{Z}_i = \boldsymbol{z}$, the conditional density function of the observed data (A_i, T_i, δ_i) at (a, t, δ) can be further given by

$$\phi_2(a,t,\delta\,|\,\boldsymbol{z}) = \frac{w(a\,|\,\boldsymbol{z})\{f(t\,|\,\boldsymbol{z})\overline{G}_{rc}(t-a\,|\,\boldsymbol{z})\}^{\delta}\{S(t\,|\,\boldsymbol{z})g_{rc}(t-a\,|\,\boldsymbol{z})\}^{1-\delta}}{\int_0^\infty w(a\,|\,\boldsymbol{z})S(a\,|\,\boldsymbol{z})da},$$

where $\overline{G}_{rc}(t\,|\,\boldsymbol{z}) = \int_t^\infty g_{rc}(s\,|\,\boldsymbol{z})ds$ is the survival function of R_{ci}. We additionally obtain

$$\begin{aligned}
&\mathrm{E}[Y_i(t)\,|\,\text{observed data}]\\
&\quad= \frac{\int_0^t \phi_2(a,T_i,\delta_i\,|\,\boldsymbol{z}_i)da}{\int_0^{T_i} \phi_2(a,T_i,\delta_i\,|\,\boldsymbol{z}_i)da} I(T_i \geqslant t)\\
&\quad= \left\{\delta_i \frac{\int_0^t w(a\,|\,\boldsymbol{z}_i)\overline{G}_{rc}(T_i-a\,|\,\boldsymbol{z}_i)da}{\int_0^{T_i} w(a\,|\,\boldsymbol{z}_i)\overline{G}_{rc}(T_i-a\,|\,\boldsymbol{z}_i)da} + (1-\delta_i)\frac{\int_0^t w(a\,|\,\boldsymbol{z}_i)g_{rc}(T_i-a\,|\,\boldsymbol{z}_i)da}{\int_0^{T_i} w(a\,|\,\boldsymbol{z}_i)g_{rc}(T_i-a\,|\,\boldsymbol{z}_i)da}\right\} I(T_i \geqslant t)\\
&\quad\equiv \left\{\delta_i \frac{W_1(T_i,t)}{W_1(T_i,T_i)} + (1-\delta_i)\frac{W_0(T_i,t)}{W_0(T_i,T_i)}\right\} I(T_i \geqslant t).
\end{aligned}$$

Note that, in this type of censoring, the conditional expectation $E[Y_i\,|\,\text{observed data}]$ is a function of the censoring distribution. Hence, using the aforementioned procedure, the estimating function can be obtained as,

$$U^{(2)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(2)}(t)][dN_i(t) - \pi_i^{(2)}(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt], \tag{2.9}$$

where

$$\pi_i^{(2)}(t) = I(T_i \geqslant t)\left\{\delta_i \frac{W_1(T_i,t)}{W_1(T_i,T_i)} + (1-\delta_i)\frac{W_0(T_i,t)}{W_0(T_i,T_i)}\right\},$$

and

$$\overline{\boldsymbol{Z}}^{(2)}(t) = \sum_{i=1}^n \pi_i^{(2)}(t)\boldsymbol{Z}_i(t) \Big/ \sum_{i=1}^n \pi_i^{(2)}(t).$$

When $\overline{G}_{rc}(t)$ is unknown in practice, $\overline{G}_{rc}(t)$ can be replaced by the Kaplan-Meier estimator $\widehat{G}_{rc}(t)$ using the residual observations $\{R_i, \delta_i : i = 1, \ldots, n\}$:

$$\widehat{G}_{rc}(t) = 1 - \prod_{R_i \leqslant t}\left\{1 - \frac{\Delta N^G(R_i)}{C^G(R_i)}\right\},$$

where $N^G(t) = n^{-1}\sum_{i=1}^n I(R_i \leqslant t, \delta_i = 0)$, $C^G(t) = n^{-1}\sum_{i=1}^n I(R_i \geqslant t)$, and $\Delta N^G(t) = N^G(t+) - N^G(t-)$. The conditional at-risk process can be further estimated as follows:

$$\widehat{\pi}_i^{(2)}(t) = \left\{\delta_i \frac{\widehat{W}_1(T_i,t)}{\widehat{W}_1(T_i,T_i)} + (1-\delta_i)\frac{\widehat{W}_0(T_i,t)}{\widehat{W}_0(T_i,T_i)}\right\} I(T_i \geqslant t).$$

By replacing $\pi_i^{(2)}(t)$ with $\widehat{\pi}_i^{(2)}(t)$ in (2.9), the estimating equation can be obtained as

$$\widehat{U}^{(2)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\widehat{\boldsymbol{Z}}}^{(2)}(t)][dN_i(t) - \widehat{\pi}_i^{(2)}(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt], \tag{2.10}$$

where $\overline{\widehat{\boldsymbol{Z}}}^{(2)}(t) = \sum_{i=1}^n \widehat{\pi}_i^{(2)}(t)\boldsymbol{Z}_i(t)/\sum_{i=1}^n \widehat{\pi}_i^{(2)}(t)$. The resulting closed-form estimator can be given by

$$\widehat{\boldsymbol{\beta}}^{(2)} = \left[\sum_{i=1}^n \int_0^\tau \widehat{\pi}_i^{(2)}(t)\{\boldsymbol{Z}_i(t) - \overline{\widehat{\boldsymbol{Z}}}^{(2)}(t)\}^{\otimes 2}dt\right]^{-1}\left[\sum_{i=1}^n \int_0^\tau \{\boldsymbol{Z}_i(t) - \overline{\widehat{\boldsymbol{Z}}}^{(2)}(t)\}dN_i(t)\right].$$

2.5 Extension to the missing covariates

Let T^* and C be the latent failure and censoring times, respectively. Let $T = \min(T^*, C)$ be the observed time, and $\delta = I(T^* \leqslant C)$ be the failure indicator. It is assumed that T^* is conditional independent of C given covariates $\boldsymbol{Z}(\cdot)$. Suppose that $\boldsymbol{Z}(\cdot)$ can be partitioned as $\boldsymbol{Z}(\cdot) = (\boldsymbol{Z}^{m^{\mathrm{T}}}(\cdot), \boldsymbol{Z}^{c^{\mathrm{T}}}(\cdot)^{\mathrm{T}})$, where $\boldsymbol{Z}^m(\cdot)$ is the covariates with possibly missing values, and $\boldsymbol{Z}^c(\cdot)$ is the complete covariates. $(T, \delta, \boldsymbol{Z}^c(\cdot))$ are observed for all subjects in the study; however, the missing covariates $\boldsymbol{Z}^m(\cdot)$ are observed only for a subset. By introducing the selection indicator, $V = I(\boldsymbol{Z}^m(\cdot) \text{ is observed})$, and by assuming the missing-at-random (see [41]), the distribution of V given $(T, \delta, \boldsymbol{Z}(\cdot))$ is

$$W(T \,|\, \delta, \boldsymbol{Z}^c(\cdot)) \equiv \mathrm{P}(V = 1 \,|\, T, \delta, \boldsymbol{Z}^c(\cdot)) = \mathrm{P}(V = 1 \,|\, T, \delta, \boldsymbol{Z}(\cdot)).$$

Let $(T_i, \delta_i, \boldsymbol{Z}_i(\cdot), V_i)$ be the independent and identically distributed copies of $(T, \delta, \boldsymbol{Z}(\cdot), V)$ for $i = 1, \ldots, n$, where $\boldsymbol{Z}_i(\cdot) = (\boldsymbol{Z}_i^{m^{\mathrm{T}}}(\cdot), \boldsymbol{Z}_i^{c^{\mathrm{T}}}(\cdot))^{\mathrm{T}}$. The observed data are $\{(T_i, \delta_i, \boldsymbol{Z}_i^c, V_i\boldsymbol{Z}_i^m(\cdot), V_i) : i = 1, \ldots, n\}$.

To validate the embedded estimating estimation method for the missing covariates, we introduce a latent non-negative random variable U with conditional distribution $\mathrm{P}(U \leqslant t | \boldsymbol{Z}^c(\cdot) = \boldsymbol{z}^c, \boldsymbol{Z}^m(\cdot) = \boldsymbol{z}^m)$ $= W(t \,|\, \boldsymbol{z}^c)$. We assume that U and (T, δ) are conditionally independent given $\boldsymbol{Z}(\cdot)$, and that $(U, T, \boldsymbol{Z}(\cdot), \delta)$ are observable only if $U \leqslant T$. Consequently, the indicator $I(U_i \leqslant T_i)$ can be considered as a selection indicator V_i, which indicates whether or not the i-th subject contains missing covariates. To show that the data with missing covariates can be viewed as a biased sample from a population, we derive the density of the observed time T without censoring. Specifically, in the absence of right censoring, the joint density of (U, T) conditional on $(\boldsymbol{Z}^c(\cdot), \boldsymbol{Z}^m(\cdot)) = (\boldsymbol{z}^c, \boldsymbol{z}^m)$ is given by

$$\begin{aligned}
&\mathrm{P}(T = t, U = u \,|\, \boldsymbol{Z}^c(\cdot) = \boldsymbol{z}^c, \boldsymbol{Z}^m(\cdot) = \boldsymbol{z}^m) \\
&\quad = \mathrm{P}(T^* = t, U = u \,|\, U \leqslant T^*, \boldsymbol{Z}^c(\cdot) = \boldsymbol{z}^c, \boldsymbol{Z}^m(\cdot) = \boldsymbol{z}^m) \\
&\quad = \frac{f(t \,|\, \boldsymbol{z}^c) w(u \,|\, \boldsymbol{z}^c) I(t > u)}{\mu},
\end{aligned}$$

where $w(u \,|\, \boldsymbol{z}^c)$ is the density of U given $\boldsymbol{z}^c$, and $\mu = \mathrm{E}W(T^* \,|\, \boldsymbol{z}^c)$. Because the density of the observed T, $W(t \,|\, \boldsymbol{z}^c) f(t \,|\, \boldsymbol{z}^c)/\mu$, takes the same form as that of (2.1) for biased-sampling data, the missing data can be viewed as a biased sample from the population. The probability of selecting subjects with $U \leqslant T$ from this population is proportional to $W(t \,|\, \boldsymbol{z}^c)$. As suggested by [31], this sampling scheme can be termed as left-truncation, where U plays the role of the left-truncation time.

For the complete data $(U_i, T_i, \boldsymbol{Z}_i(\cdot), \delta_i)$, the local square-integrable martingale is given by

$$M_i^U(t) = N_i^U(t) - \int_0^t Y_i^U(t)[d\Lambda_0(u) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(u) du],$$

where $N_i^U(t) = I(U_i \leqslant T_i) N_i(t)$, $Y_i^U(t) = I(U_i \leqslant t \leqslant T_i)$.

Note that the latent left truncation time U is not observable. Using the similar idea in the Cox model with missing covariates in [31], we can treat $(T_i, \boldsymbol{Z}_i, \delta_i)$ as the missing data with the left-truncation time U_i completely missing. Furthermore, by the assumption $W(T \,|\, \boldsymbol{Z}^c, \delta) = W(T \,|\, \boldsymbol{Z}^c)$ and the relation $I(U_i \leqslant T_i) = V_i$, we obtain

$$\begin{aligned}
M_i^{(m)}(t) &\equiv \mathrm{E}[M_i^U(t) \,|\, \text{data}] \\
&= V_i N_i(t) - \int_0^t V_i \pi_i^{(m)}(u)[d\Lambda_0(u \,|\, \boldsymbol{Z}_i) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(u) du], \qquad (2.11)
\end{aligned}$$

where $\pi_i^{(m)}(t) = \frac{W(t \,|\, \boldsymbol{Z}_i^c)}{W(T_i \,|\, \boldsymbol{Z}_i^c)} I(T_i \geqslant t)$. Clearly, $M_i(t)$ is a mean-zero process. Similar to the aforementioned procedure, the estimating equation can be derived as

$$U^{(m)}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}^{(m)}(t)] V_i [dN_i(t) - \pi_i^{(m)}(t) \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt], \qquad (2.12)$$

where $\overline{Z}^{(m)}(t)=\sum_{i=1}^n V_i\pi_i^{(m)}(t)\boldsymbol{Z}_i/\sum_{i=1}^n V_i\pi_i^{(m)}(t)$.

However, $W(T,\delta,\boldsymbol{Z}^c(\cdot))$ is unknown in practice, and $W(T,\delta,\boldsymbol{Z}^c(\cdot))$ depends on the failure indicator δ. By the similar arguments to that of [31], we can generalize the weight function as $\widehat{\pi}_i^{(m)}(t)=\widehat{W}(t\,|\,\boldsymbol{Z}_i^c(\cdot),1)I(T_i\geqslant t)/\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c(\cdot),\delta_i)$, by replacing $W(t\,|\,\delta,\boldsymbol{Z}^c(\cdot))$ with its consistent estimator $\widehat{W}(t\,|\,\delta,\boldsymbol{Z}^c(\cdot))$. The embedded estimating equation can be proposed as

$$\widehat{U}^{(m)}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^n V_i\int_0^\tau[\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}^{(m)}(t)][dN_i(t)-\widehat{\pi}_i^{(m)}(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt], \tag{2.13}$$

where

$$\overline{\widehat{\boldsymbol{Z}}}^{(2)}(t)=\frac{\sum_{i=1}^n V_i\widehat{\pi}_i^{(m)}(t)\boldsymbol{Z}_i(t)}{\sum_{i=1}^n V_i\widehat{\pi}_i^{(m)}(t)}.$$

Thus, the resulting closed-form estimator is given by

$$\widehat{\boldsymbol{\beta}}^{(m)}=\bigg[\sum_{i=1}^n V_i\int_0^\tau\widehat{\pi}_i^{(m)}(t)\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}^{(m)}(t)\}^{\otimes2}dt\bigg]^{-1}\bigg[\sum_{i=1}^n V_i\int_0^\tau\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}^{(m)}(t)\}dN_i(t)\bigg].$$

Remark 2.1. (i) For the additive model with missing covariates, Lin [27] proposed the following simple weighted estimator:

$$\widehat{\boldsymbol{\beta}}_{SW}=\bigg[\sum_{i=1}^n\frac{V_i}{\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}\int_0^\tau\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}_{SW}(t)\}^{\otimes2}dt\bigg]^{-1}\sum_{i=1}^n\frac{V_i}{\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}\int_0^\tau\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}_{SW}(t)\}dN_i(t),$$

where

$$\overline{\widehat{\boldsymbol{Z}}}_{SW}(t)=\frac{\sum_{i=1}^n V_iI(T_i\geqslant t)\boldsymbol{Z}_i(t)/\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}{\sum_{i=1}^n V_iI(T_i\geqslant t)/\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}.$$

To reduce the overweighting problem using the inverse probability weighted method, Hao [12] proposed the reweighting estimator

$$\widehat{\boldsymbol{\beta}}_{SR}=\bigg[\sum_{i=1}^n\int_0^\tau\frac{V_i\widehat{W}(t\,|\,\boldsymbol{Z}_i^c,1)}{\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}_{SR}(t)\}^{\otimes2}dt\bigg]^{-1}\sum_{i=1}^n\int_0^\tau\frac{V_i\widehat{W}(t\,|\,\boldsymbol{Z}_i^c,1)}{\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}\{\boldsymbol{Z}_i(t)-\overline{\widehat{\boldsymbol{Z}}}_{SR}(t)\}dN_i(t),$$

where

$$\overline{\widehat{\boldsymbol{Z}}}_{SR}(t)=\frac{\sum_{i=1}^n V_iI(T_i\geqslant t)\boldsymbol{Z}_i(t)\widehat{W}(t\,|\,\boldsymbol{Z}_i^c,1)/\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}{\sum_{i=1}^n V_iI(T_i\geqslant t)\widehat{W}(t\,|\,\boldsymbol{Z}_i^c,1)/\widehat{W}(T_i\,|\,\boldsymbol{Z}_i^c,\delta_i)}.$$

Note that our proposed estimator $\widehat{\boldsymbol{\beta}}^{(m)}$ is exactly the estimator $\widehat{\boldsymbol{\beta}}_{SR}$ proposed by [12] data with missing covariates.

(ii) As pointed out by a referee, to improve robustness and efficiency of $\widehat{\boldsymbol{\beta}}^{(m)}$, one may add an augmented term on the estimating functions based on the double robust technique in [40]. This resulting estimator possesses the double-robustness property, i.e., given the observed data, the estimator is consistent whether the selection probability or the conditional distribution of the missing covariates is correctly specified (see [51]). Thus, Hao [12] constructed an augmented reweighting estimator $\widehat{\boldsymbol{\beta}}_{AR}$, which is more efficient than the simple weighted estimator $\widehat{\boldsymbol{\beta}}_{SR}$. Furthermore, we should recall our proposed estimator $\widehat{\boldsymbol{\beta}}^{(m)}$ is identical to the estimator $\widehat{\boldsymbol{\beta}}_{SR}$. Therefore, The augmented reweighted version of $\widehat{\boldsymbol{\beta}}^{(m)}$ can be constructed in a similar manner to $\widehat{\boldsymbol{\beta}}_{SR}$ in [12]. The details have been omitted to save space.

2.6 Asymptotic properties

In this subsection, we derive the asymptotic properties of the proposed estimator. Note that the estimating equations (2.8), (2.9) and (2.13) takes the form of (2.6). Therefore, without loss of ambiguity, we write the estimating equation $U(\boldsymbol{\beta})$ as (2.6), and express $\widehat{U}(\boldsymbol{\beta})$ by replacing the weight function $\pi_i(t)$ with $\widehat{\pi}_i(t)$.

Theorem 2.2. *Under the regularity conditions in Appendix* A, $n^{1/2}\widehat{U}(\boldsymbol{\beta})$ *converges in distribution to a zero-mean multivariate normal distribution with a covariance matrix* Σ, *where* Σ *can be consistently estimated by*

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^{\tau}\{\boldsymbol{Z}_i(t) - \widehat{\boldsymbol{Z}}(t)\}\{dN_i(t) - \widehat{\pi}_i(t)\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt\}\right]^{\otimes 2},$$

with $\widehat{\boldsymbol{Z}}(t) = \sum_i^n \widehat{\pi}_i(t)\boldsymbol{Z}_i(t)/\sum_i^n \widehat{\pi}_i(t)$. *Furthermore,* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *converges in distribution to a zero-mean multivariate normal distribution with variance matrix* $D^{-1}\Sigma D^{-1}$, *which can be consistently estimated by* $\widehat{D}^{-1}\widehat{\Sigma}\widehat{D}^{-1}$, *where*

$$\widehat{D} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}[\boldsymbol{Z}_i(t) - \overline{\widehat{\boldsymbol{Z}}}_i(t)]^{\otimes 2}\widehat{\pi}_i(t)dt.$$

The subsequent theorem exhibits the weak convergence of the process $\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, t)$.

Theorem 2.3. *Under the regularity conditions in Appendix* A, $\sqrt{n}[\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}}, t) - \Lambda_0(t)]$ *converges weakly to a zero-mean Gaussian process with covariance function* $\eta(s,t) \equiv \mathrm{E}\{\Psi_1(s)\Psi_1(t)^{\mathrm{T}}\}$, *where*

$$\Psi_i(t) = \int_0^t \frac{dM_i(u)}{s^{(0)}(u)} - \boldsymbol{h}^{\mathrm{T}}(\boldsymbol{\beta}_0, t)D^{-1}\int_0^{\tau}[\boldsymbol{Z}_i(u) - \overline{\boldsymbol{z}}(u)]dM_i(u),$$

and $\boldsymbol{h}(\boldsymbol{\beta}_0, t) = \int_0^t \frac{\mathrm{E}[\pi_1(u)\boldsymbol{Z}_1]}{s^{(0)}(u)}du$.

3 Model-checking techniques

The i-th residual in model (2.2) can be defined as

$$\widehat{M}_i(t) = N_i(t) - \int_0^t \widehat{\pi}_i(u)\{d\widehat{\Lambda}_0(u) + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_i(u)du\}.$$

Note that $\widehat{M}_i(t)$ can be viewed as the difference between the observed and model-predicted number of failures for the i-th subject by time t. Thus, using these residuals, one can verify the adequacy of (2.2) for biased-sampling data. However, as the $\widehat{M}_i(t)$ are not martingales, the standard tests of fit for additive risk models (see [16, 56]) are not directly applicable. Following the basic ideas of [25], we develop a class of graphical and numerical methods based on the cumulative sums of the residuals $\widehat{M}_i(t)$.

The following multi-parameter stochastic process involves various forms of cumulative sums of $\widehat{M}_i(t)$,

$$Q(t, \boldsymbol{z}) = n^{-1/2}\sum_{i=1}^{n}\int_0^t q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u) \leqslant \boldsymbol{z})d\widehat{M}_i(u),$$

where $q(\cdot)$ is a known vector-valued bounded function, and $I(\boldsymbol{Z}_i(\cdot) \leqslant \boldsymbol{z}) = I(Z_{i1}(\cdot) \leqslant z_1, \ldots, Z_{ip}(\cdot) \leqslant z_p)$.

Intuitively, if (2.2) holds, these processes will randomly fluctuate around zero. However, we need to establish the asymptotic properties of $Q(t, \boldsymbol{z})$. To proceed, we define some notation. Let

$$S_q(t, \boldsymbol{z}) = \frac{\sum_{i=1}^n q(\boldsymbol{Z}_i(t))I(\boldsymbol{Z}_i(t) \leqslant \boldsymbol{z})\widehat{\pi}_i(t)}{\sum_{i=1}^n \widehat{\pi}_i(t)},$$

and

$$B_q(t, \boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u) \leqslant \boldsymbol{z})\widehat{\pi}_i(u)[\boldsymbol{Z}_i(u) - \overline{\widehat{\boldsymbol{Z}}}(u)]^{\mathrm{T}}du,$$

where $\overline{\widehat{\boldsymbol{Z}}}(u) = \sum_{i=1}^n \widehat{\pi}_i(u)\boldsymbol{Z}_i(u)/\sum_{i=1}^n \widehat{\pi}_i(u)$. Furthermore, we denote $s_q(t, z)$, $b_q(t, \boldsymbol{z})$ and $\overline{\boldsymbol{z}}(t)$ as the limit of $S_q(t, \boldsymbol{z})$, $B_q(t, \boldsymbol{z})$ and $\overline{\widehat{\boldsymbol{Z}}}(t)$, respectively. Subsequently, we define

$$\gamma_i(t, z) = \int_0^t [q(\boldsymbol{Z}_i)I(\boldsymbol{Z}_i(u) \leqslant \boldsymbol{z}) - s_q(u, \boldsymbol{z})]dM_i(u)$$

$$
- b_q(t,\boldsymbol{z})D^{-1}\int_0^{\tau}[\boldsymbol{Z}_i(u)-\overline{z}(u)]dM_i(u),
$$

$$
\begin{aligned}
\widehat{\gamma}_i(t,\boldsymbol{z}) = &\int_0^t [q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u)\leqslant \boldsymbol{z}) - S_q(u,\boldsymbol{z})]d\widehat{M}_i(u)\\
&- B_q(t,\boldsymbol{z})\widehat{D}^{-1}\int_0^{\tau}[\boldsymbol{Z}_i(u)-\overline{\widehat{\boldsymbol{Z}}}(u)]d\widehat{M}_i(u).
\end{aligned}
$$

The following theorem provides the asymptotic property of $Q(t,z)$.

Theorem 3.1. *Under the regularity conditions in Appendix* A*,* $Q(t,z)$ *converges weakly to a zero-mean Gaussian process with a covariance function* $\mathrm{E}[\gamma_1(t,\boldsymbol{z})\gamma_1(t^*,\boldsymbol{z}^*)^{\mathrm{T}}]$ *at* $(t,\boldsymbol{z})$ *and* $(t^*,\boldsymbol{z}^*)$*.*

In practice, the limiting distribution of $Q(t,z)$ can be approximating through a Monte Carlo simulation technique. First, we can independently generate a simple random sample $\{\xi_1,\ldots,\xi_n\}$ from the standard normal distribution $N(0,1)$, which are independent of $\{N_i(t),Y_i(t),\boldsymbol{Z}_i(t):i=1,\ldots,n\}$ based on the observed data. Then, we can obtain the perturbed version of the stochastic process $\widehat{Q}(t,\boldsymbol{z}) = n^{-1/2}\sum_{i=1}^n\widehat{\gamma}_i(t,\boldsymbol{z})\xi_i$. The perturbation procedure can be theoretically justified by the following theorem.

Theorem 3.2. *Given* $\{N_i(t),Y_i(t),\boldsymbol{Z}_i(t):i=1,\ldots,n\}$*,* $\widehat{Q}(t,\boldsymbol{z})$ *converges weakly to the same zero-mean Gaussian process as that of* $Q(t,\boldsymbol{z})$*.*

Subsequently, we illustrate the application of $Q(t,\boldsymbol{z})$ to model-fitting evaluations with different purposes. First, we consider the problem of checking the functional forms of the covariates. For the j-th component of $\boldsymbol{Z}(\cdot)$, we take $q(\boldsymbol{z})=1$, $t=\tau$, and $z_k=\infty$ for all $k\neq j$, obtaining

$$
Q_j(z) = n^{-1/2}\sum_{i=1}^n\int_0^{\tau} I(Z_{ij}(u)\leqslant z)d\widehat{M}_i(u).
$$

The null distribution of $Q(\boldsymbol{z})$ can be approximated by the corresponding zero-mean Gaussian process $\widehat{Q}_j(z)$,

$$
\begin{aligned}
\widehat{Q}_j(z) = &\, n^{-1/2}\sum_{i=1}^n\int_0^{\tau}[I(Z_{ij}(u)\leqslant z) - S_q^j(u,\boldsymbol{z})]d\widehat{M}_i(u)\xi_i\\
&- B_q^j(\tau,\boldsymbol{z})\widehat{D}^{-1}n^{-1/2}\sum_{i=1}^n\int_0^{\tau}[\boldsymbol{Z}_i(u)-\overline{\widehat{\boldsymbol{Z}}}(u)]d\widehat{M}_i(u)\xi_i, \qquad (3.1)
\end{aligned}
$$

where S_q^j and $B_q^j(\tau,z)$ are the j-th component of S_q and $B_q(\tau,\boldsymbol{z})$, respectively. To assess unusual the observed residual pattern, we simultaneously plot several (e.g., 50) realizations of $\widehat{Q}_j(z)$ along with the observed $Q_j(z)$. Furthermore, to obtain the p-value of the supremum test $\sup_{\boldsymbol{z}}|Q(\boldsymbol{z})|$, we can generate a large number of, say 1,000, realizations of $\sup_{\boldsymbol{z}}|\widehat{Q}(\boldsymbol{z})|$, and then calculate the percentage of those greater than the observed value of $\sup_{\boldsymbol{z}}|Q(\boldsymbol{z})|$.

To check the additive risk assumption under (2.2), we consider the standardized score-type process

$$
U_j^*(t) = (\widehat{\Sigma}_{jj}^{-1})^{1/2}n^{1/2}U_j(\widehat{\boldsymbol{\beta}},t),
$$

where $U_j(\widehat{\boldsymbol{\beta}},t)$ is the j-th component of $U(\widehat{\beta},t)$ and $\widehat{\Sigma}_{jj}^{-1}$ is the j-th diagonal element of $\widehat{\Sigma}^{-1}$. Clearly,

$$
n^{1/2}U(\widehat{\beta},t) = n^{-1/2}\sum_{i=1}^n\int_0^t[\boldsymbol{Z}_i(u)-\overline{\widehat{\boldsymbol{Z}}}(u)]d\widehat{M}_i(u)
$$

is a special case of $Q(t,\boldsymbol{z})$ with $q(\boldsymbol{Z}_i(u)) = \boldsymbol{Z}_i(u)-\overline{\widehat{\boldsymbol{Z}}}(u)$ and $\boldsymbol{z}=\infty$. In this case, $S_q(u,\boldsymbol{z})=0$. As shown in the proof of Theorem 3.1, the null distribution of $U_j^*(t)$ can be approximated by that of the zero-mean Gaussian process,

$$
\widehat{U}_j^*(t) = (\widehat{\Sigma}_{jj}^{-1})^{1/2}\bigg[n^{-1/2}\sum_{i=1}^n\int_0^t[Z_{ij}(u)-\overline{\widehat{\boldsymbol{Z}}}_j(u)]d\widehat{M}_i(u)\xi_i
$$

$$- B_q^j(t,\boldsymbol{z})\widehat{D}^{-1}n^{-1/2}\sum_{i=1}^n\int_0^\tau[\boldsymbol{Z}_i(u)-\widehat{\overline{\boldsymbol{Z}}}(u)]d\widehat{M}_i(u)\xi_i\bigg],$$

where $B_q^j(t,z)$ is the j-th component of $B_q(t,\boldsymbol{z})$. Graphical and numerical inspections can be performed by simulation in the same fashion as for $\widehat{Q}_j(z)$. The test statistic for checking the additive hazards structure of the j-th covariate $(j=1,\ldots,p)$ is given by

$$S_j=\sup_{t\in[0,\tau]}|U_j^*(t)|.$$

The p-value ($=\mathrm{P}(S_j>s_j)$) can be approximated by $\mathrm{P}(\widehat{S}_j^*>s_j)$, where s_j is the observed value of S_j, and

$$\widehat{S}_j=\sup_{t\in[0,\tau]}|\widehat{U}_j^*(t)|.$$

The p-value can be empirically estimated by the percentage of $(\widehat{S}_j>s_j)$ through generating many realizations of $\widehat{S}_j$. The overall test statistic for the joint additivity of all covariates is given by

$$S_a=\sup_{t\in[0,\tau]}\sum_{j=1}^p|U_j^*(t)|.$$

4 Simulation studies

4.1 Accuracy of estimation

A series of simulation studies were conducted for evaluating the inference procedure proposed in the previous sections. As we pointed out in Subsection 2.5, our embedded estimating equation estimator for data with missing covariatesis is exactly the reweighted estimator proposed by [12]. Thus, the results of missing covariates are omitted for saving space.

The data were generated as follows. The survival time T^* was generated from the additive risk model,

$$\lambda(t\,|\,Z)=\lambda_0(t)+\beta_1Z_1+\beta_2Z_2,$$

where $\lambda_0(t)=0.5t^{-1/2}$, $Z_1\sim\mathrm{Binom}(1,0.5)$ and $Z_2\sim\mathrm{Unif}(0.5,1.5)$, and the true value $(\beta_1,\beta_2)=(-0.5,1)$. The truncation time A was generated from two different types of biased-sampling data:

(1) Length-biased sampling: A^* was generated from a uniform distribution $\mathrm{Unif}(0,100)$;

(2) General biased sampling: A^* was generated from a Weibull distribution with a scale parameter 1.2 and a shape parameter 0.8.

The data from biased sampling were censored in two mechanisms:

(I) *Censoring mechanism* I: *censoring before biased sampling.*

The censoring time C^* was independently generated from an exponential distribution with parameter c, which controls the censoring percentages: $10\%, 30\%, 50\%$. In the censored data (T_i,δ_i), the observed data (A,T,δ) were obtained only when $T\geqslant A$, where $T=\min(T^*,C^*)$ and $\delta=I(T^*\leqslant C^*)$.

(II) *Censoring mechanism* II: *censoring after biased sampling.*

Let $R^*=T^*-A$ be the residual lifetime. We then independently generated the residual censoring time R_c from an exponential distribution with parameter c, where c controls the censoring percentages: $10\%, 30\%, 50\%$. The observations (A,T,δ) were retained when $T\geqslant A$, where $T=A^*+\min(R^*,R_c)$ and $\delta=I(R^*\leqslant R_c)$. The observed data were (A,T,δ), where $T=A+R$, $R=\min(R^*,R_c)$ and $\delta=I(R^*\leqslant R_c)$.

In each scenario, 1,000 repetitions were conducted with the sample size $n=150$ and $n=300$. For comparison, both the proposed estimator $\widehat{\boldsymbol{\beta}}$ and the naive estimator $\widehat{\boldsymbol{\beta}}^{(L)}$ were used. The estimate results under censoring mechanisms I and II are summarized in Tables 1 and 2, respectively. In each scenario, all the estimators appear to be unbiased over the range of right censoring rates from low (10%) to heavy (50%), and the estimated standard errors ("ESE") are close to the empirical standard errors ("SD"),

Table 1 Performance comparison between the proposed estimator and the naive estimator based on 1,000 simulated samples under censoring mechanism I (censoring before biased sampling)

		$n=150$				$n=300$			
		Naive		Proposed		Naive		Proposed	
$C\%$		$\widehat{\beta}_1^{(L)}$	$\widehat{\beta}_2^{(L)}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1^{(L)}$	$\widehat{\beta}_2^{(L)}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
		Length-biased sampling							
10%	Bias	−0.007	−0.014	0.002	−0.016	−0.004	−0.006	0.003	−0.010
	SD	0.223	0.340	0.154	0.250	0.151	0.245	0.107	0.180
	ESE	0.219	0.357	0.153	0.255	0.154	0.250	0.107	0.176
	CP	0.948	0.964	0.955	0.976	0.949	0.955	0.950	0.955
	MSE	0.050	0.115	0.024	0.063	0.023	0.060	0.012	0.032
	RMSE	1.000	1.000	0.474	0.543	1.000	1.000	0.507	0.536
30%	Bias	−0.005	−0.026	0.004	−0.029	−0.004	0.001	0.006	−0.006
	SD	0.259	0.429	0.199	0.329	0.183	0.304	0.142	0.233
	ESE	0.257	0.430	0.200	0.342	0.180	0.303	0.139	0.236
	CP	0.953	0.953	0.957	0.981	0.947	0.948	0.940	0.963
	MSE	0.067	0.184	0.039	0.109	0.033	0.092	0.020	0.054
	RMSE	1.000	1.000	0.588	0.590	1.000	1.000	0.604	0.589
50%	Bias	0.002	−0.019	0.020	−0.028	−0.006	−0.004	0.001	−0.013
	SD	0.335	0.573	0.279	0.461	0.235	0.392	0.196	0.316
	ESE	0.333	0.573	0.287	0.495	0.234	0.399	0.199	0.341
	CP	0.951	0.954	0.957	0.985	0.957	0.948	0.960	0.984
	MSE	0.112	0.328	0.078	0.213	0.055	0.154	0.038	0.100
	RMSE	1.000	1.000	0.695	0.649	1.000	1.000	0.694	0.651
		General biased sampling							
10%	Bias	−0.007	−0.001	−0.008	0.000	−0.003	−0.007	−0.002	−0.007
	SD	0.241	0.394	0.205	0.330	0.174	0.288	0.146	0.236
	ESE	0.239	0.402	0.190	0.327	0.167	0.280	0.131	0.223
	CP	0.953	0.958	0.933	0.978	0.944	0.949	0.923	0.959
	MSE	0.058	0.155	0.042	0.109	0.030	0.083	0.021	0.056
	RMSE	1.000	1.000	0.726	0.700	1.000	1.000	0.707	0.677
30%	Bias	−0.003	0.006	0.003	0.011	0.002	−0.003	0.001	−0.006
	SD	0.289	0.508	0.254	0.434	0.213	0.350	0.179	0.299
	ESE	0.287	0.493	0.246	0.426	0.201	0.344	0.169	0.291
	CP	0.957	0.948	0.945	0.978	0.938	0.947	0.938	0.962
	MSE	0.083	0.258	0.064	0.189	0.045	0.122	0.032	0.089
	RMSE	1.000	1.000	0.774	0.732	1.000	1.000	0.706	0.730
50%	Bias	−0.006	−0.003	0.002	0.008	0.001	0.002	0.005	0.008
	SD	0.392	0.654	0.347	0.562	0.273	0.456	0.240	0.396
	ESE	0.372	0.640	0.338	0.585	0.260	0.448	0.232	0.402
	CP	0.946	0.948	0.948	0.984	0.940	0.953	0.939	0.978
	MSE	0.153	0.427	0.120	0.316	0.074	0.207	0.058	0.157
	RMSE	1.000	1.000	0.784	0.740	1.000	1.000	0.776	0.756

Bias: the empirical bias; SD: empirical standard error; ESE: average estimated standard error; CP: 95% coverage probability; MSE: mean squared error; RMSE: MSE ratio of the proposed estimator to the naive estimator.

indicating good performance of the variance estimator. In addition, the empirical coverage probabilities ("CP") are close to the nominal level 95%. As expected, when the sample size increases from $n=150$ to $n=300$ or when censoring rate varies from heavy (50%) to light (10%), the mean squared errors ("MSE") trend to be smaller. To assess the efficiency of the proposed estimator, we compute the relative mean squared error ("RMSE"), which is defined by the MSE ratio of the proposed estimator to the naive estimator. As expected, all the RMSEs are less than 0.8, which indicates that the proposed estimator is more efficient than the naive estimator.

Table 2 Performance comparison between the proposed estimator and the naive estimator based on 1,000 simulated samples under censoring mechanism II (censoring after biased sampling)

		$n = 150$				$n = 300$			
		Naive		Proposed		Naive		Proposed	
$C\%$		$\widehat{\beta}_1^{(L)}$	$\widehat{\beta}_2^{(L)}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1^{(L)}$	$\widehat{\beta}_2^{(L)}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
				Length-biased sampling					
10%	Bias	0.000	−0.016	0.010	−0.018	0.004	−0.022	0.001	−0.004
	SD	0.212	0.354	0.144	0.238	0.160	0.251	0.105	0.172
	ESE	0.221	0.354	0.150	0.247	0.155	0.247	0.105	0.172
	CP	0.965	0.944	0.960	0.974	0.946	0.941	0.958	0.962
	MSE	0.045	0.125	0.021	0.057	0.026	0.063	0.011	0.029
	RMSE	1.000	1.000	0.465	0.454	1.000	1.000	0.431	0.466
30%	Bias	0.006	−0.027	0.022	−0.036	0.008	−0.033	0.007	−0.018
	SD	0.240	0.409	0.167	0.280	0.181	0.291	0.122	0.205
	ESE	0.249	0.407	0.181	0.300	0.175	0.285	0.127	0.210
	CP	0.961	0.948	0.957	0.987	0.948	0.940	0.960	0.967
	MSE	0.058	0.168	0.028	0.080	0.033	0.086	0.015	0.042
	RMSE	1.000	1.000	0.489	0.476	1.000	1.000	0.454	0.492
50%	Bias	0.015	−0.047	0.031	−0.061	0.012	−0.038	0.017	−0.029
	SD	0.288	0.513	0.213	0.374	0.219	0.354	0.158	0.261
	ESE	0.299	0.499	0.236	0.395	0.212	0.350	0.167	0.278
	CP	0.962	0.945	0.960	0.984	0.948	0.950	0.960	0.973
	MSE	0.083	0.265	0.046	0.144	0.048	0.127	0.025	0.069
	RMSE	1.000	1.000	0.557	0.541	1.000	1.000	0.525	0.542
				General biased sampling					
10%	Bias	0.002	0.016	0.000	0.023	−0.005	−0.015	−0.004	−0.003
	SD	0.235	0.417	0.196	0.361	0.164	0.276	0.137	0.233
	ESE	0.237	0.402	0.188	0.324	0.166	0.278	0.129	0.218
	CP	0.951	0.957	0.945	0.956	0.949	0.948	0.942	0.946
	MSE	0.055	0.174	0.038	0.131	0.027	0.076	0.019	0.054
	RMSE	1.000	1.000	0.697	0.750	1.000	1.000	0.698	0.711
30%	Bias	0.008	0.007	0.010	0.010	−0.004	−0.011	0.000	−0.006
	SD	0.274	0.487	0.228	0.407	0.192	0.333	0.157	0.273
	ESE	0.276	0.473	0.228	0.395	0.194	0.328	0.157	0.268
	CP	0.960	0.950	0.960	0.974	0.961	0.946	0.951	0.961
	MSE	0.075	0.237	0.052	0.165	0.037	0.111	0.024	0.074
	RMSE	1.000	1.000	0.695	0.697	1.000	1.000	0.665	0.673
50%	Bias	0.017	−0.009	0.021	−0.019	0.001	−0.012	0.007	−0.012
	SD	0.361	0.623	0.298	0.508	0.249	0.440	0.205	0.360
	ESE	0.353	0.610	0.308	0.530	0.248	0.424	0.213	0.364
	CP	0.949	0.949	0.961	0.982	0.949	0.939	0.953	0.959
	MSE	0.130	0.388	0.089	0.258	0.062	0.194	0.042	0.130
	RMSE	1.000	1.000	0.683	0.666	1.000	1.000	0.681	0.670

Bias: the empirical bias; SD: empirical standard error; ESE: average estimated standard error; CP: 95% coverage probability; MSE: mean squared error; RMSE: MSE ratio of the proposed estimator to the naive estimator.

4.2 Performance of model checking

The performance of the model checking test with finite sample sizes was also investigated using simulation studies. We generated the failure time T^* from the null model H_0 and two alternative models H_{a1} and H_{a2} as follows:

- $H_0 : \lambda(t \mid Z) = 1 + Z$, where $Z \sim \text{Unif}(0, 1)$;

Table 3 Test for the additive risk assumption based on the score process. Empirical sizes and powers of the proposed test were calculated from 1,000 simulated samples at the significance level of 0.05

Censoring mechanism	$C\%$	$n=150$ Size H_0	$n=150$ Power H_{a1}	$n=150$ Power H_{a2}	$n=300$ Size H_0	$n=300$ Power H_{a1}	$n=300$ Power H_{a2}
			Length-biased sampling				
I	10%	0.075	0.355	0.815	0.052	0.580	0.905
	30%	0.062	0.225	0.765	0.060	0.512	0.875
	50%	0.071	0.175	0.595	0.085	0.405	0.795
II	10%	0.051	0.295	0.765	0.060	0.570	0.912
	30%	0.045	0.275	0.682	0.075	0.504	0.875
	50%	0.050	0.235	0.565	0.032	0.383	0.820
			General biased sampling				
I	10%	0.095	0.382	0.863	0.055	0.625	0.960
	30%	0.064	0.345	0.764	0.065	0.502	0.953
	50%	0.061	0.201	0.625	0.071	0.380	0.845
II	10%	0.025	0.385	0.805	0.065	0.612	0.925
	30%	0.075	0.335	0.770	0.075	0.545	0.920
	50%	0.034	0.230	0.585	0.055	0.375	0.801

Null model, $H_0 : \lambda(t\,|\,Z) = 1 + Z$; alternative models, $H_{a1} : \lambda(t\,|\,Z) = t\exp(Z)$ and $H_{a2} : \lambda(t\,|\,Z) = tZ$.

Table 4 Test for the functional forms of the covariate based on the cumulative residual process. Empirical sizes and powers of the proposed test are calculated from 1,000 simulated samples at the significance level of 0.05

Censoring mechanism	$C\%$	$n=150$ Size H_0	$n=150$ Power H_{a2}	$n=300$ Size H_0	$n=300$ Power H_{a2}
		Length-biased sampling			
I	10%	0.024	0.745	0.040	0.965
	30%	0.020	0.551	0.030	0.835
	50%	0.022	0.265	0.025	0.574
II	10%	0.051	0.725	0.050	0.961
	30%	0.033	0.553	0.038	0.900
	50%	0.015	0.420	0.045	0.712
		General biased sampling			
I	10%	0.015	0.622	0.026	0.945
	30%	0.014	0.425	0.036	0.794
	50%	0.028	0.247	0.027	0.442
II	10%	0.018	0.653	0.038	0.945
	30%	0.034	0.429	0.032	0.785
	50%	0.027	0.256	0.025	0.520

Null model, $H_0 : \lambda(t\,|\,Z) = 1 + Z$; and alternative model, $H_{a2} : \lambda(t\,|\,Z) = tZ$.

- $H_{a1} : \lambda(t\,|\,Z) = t\exp(Z)$, where $Z \sim \mathrm{Unif}(0, 1)$;
- $H_{a2} : \lambda(t\,|\,Z) = tZ$, where Z is a log-normal variable with a shape parameter 0 and a scale parameter 1.

The other simulation settings were as described in the previous subsection, i.e., two different types of biased-sampling (length-biased sampling and general biased sampling), and two censoring mechanisms (I and II). In each configuration, 1,000 simulations were conducted at two sample sizes, $n = 150$ and $n = 300$. The p-value of the proposed test was approximated using 1,000 bootstrap realizations of the approximating Gaussian process for each simulated sample.

Table 3 summarizes the results under the null hypothesis, providing the probabilities of rejecting

$H_0 : \lambda(t \mid Z) = 1 + Z$ at the nominal level of 0.05. In each scenario, the empirical test sizes are close to the nominal level. To detect violation of the additive risk assumption, Table 3 also presents the empirical test powers of the tests against two alternatives,

$$H_{a1} : \lambda(t \mid Z) = t \exp(Z) \quad \text{and} \quad H_{a2} : \lambda(t \mid Z) = tZ.$$

In each scenario, the proposed test based on the score process has reasonable power to reject the null hypothesis. As expected, the power tends to increase with increasing the sample size or decreasing the censoring rate.

To examine the functional form of the covariate, Table 4 summarizes the empirical test size under the null hypothesis $H_0 : \lambda(t \mid Z) = 1 + Z$, and the empirical power against the alternative hypothesis $H_{a2} : \lambda(t \mid Z) = tZ$. The proposed test has type I error close to the nominal level and reasonable power. Thus, the proposed test based on the cumulative residual process over a covariate may be a powerful test of the covariate functional form.

5 Applications

In this section, we analyze two types of biased-sampling datasets.

Example 5.1. Length-biased data.

The original shrub data, which can be found in [34], present the widths of 46 shrubs. We consider the lifetime proxy outcome of the shrub width T^*, given two covariates: $Z_1 = I(T^*$ belongs to transect I), and $Z_2 = I(T^*$ belongs to transect II). The additive risk model for the population width T^* is given by

$$\lambda(t \mid Z_1, Z_2) = \lambda_0 + \beta_1 Z_1 + \beta_2 Z_2.$$

As argued by [52], the probability of observing a shrub is proportional to its width. Thus, this data set is length-biased, and the weight function can be chosen as $W(t) = t$.

The estimated coefficients are $\widehat{\beta}_1 = 1.426$ (SE $= 0.542$) and $\widehat{\beta}_2 = 0.117$ (SE $= 0.580$), with p-values of 0.004 and 0.420, respectively. The estimates are similar to the results of the proportional hazards models reported by [44, 52]. However, the covariate Z_2 is not statistically significant as its p-value exceeds 0.05. We also applied the proposed model-checking techniques to these data. From 1,000 simulated realizations, the p-values for testing the functional forms of the covariates Z_1 and Z_2 are 1. This is expected because the covariates are dichotomous. Furthermore, the p-values for testing the additive risk assumption of Z_1 and Z_2 are 0.912 and 0.715, respectively, which suggests that the additive assumption is appropriate. As depicted in Figure 1, the observed score processes appear to be within the ranges of the initial 50 simulated score processes. It graphically supports that there is no evidence against the assumed model.

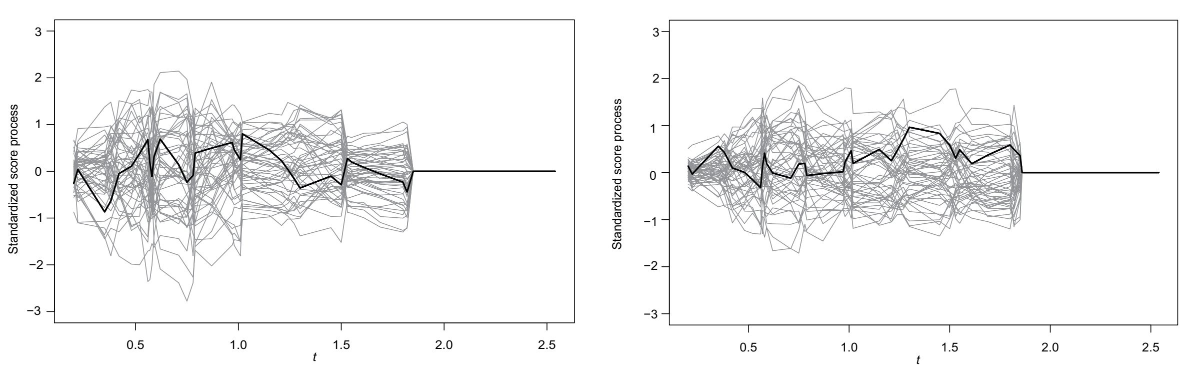

Figure 1 Standardized score processes for assessing the additive risk assumption in the shrub data. The standardized score processes is plotted against (a) Z_1 (p-value 0.912) and (b) Z_2 (p-value 0.715). Bold line = observed process; gray lines = simulated processes

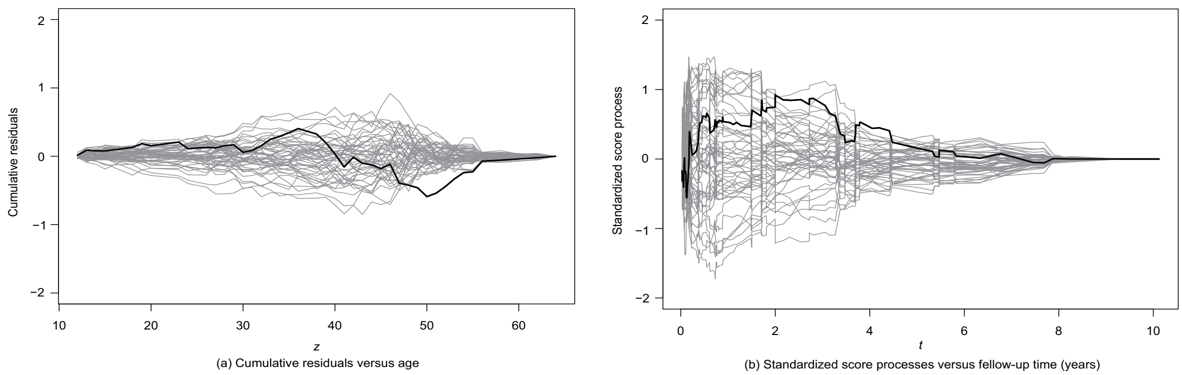

Figure 2 Plots of residual processes for the Stanford heart transplant data: (a) cumulative residuals versus age (p-value 0.167); (b) standardized score processes versus follow-up time (p-value 0.420). Bold line = observed process; gray lines = simulated processes

Example 5.2. Biased-sampling and right-censored data.

Miller and Halpern [33] reported the survival or censoring times and the ages of 184 patients, enrolled in the Stanford heart transplant program from October 1967 to February 1980. The event time T^* (in days) of interest is the survival time after entry. The censoring time is the duration between the calendar entry date of the patients and February 1980. Assuming no loss of follow-up, we let E be the calendar entry date of the patient, and calculate two quantities: the censoring time $C =$ February 1980 $- E$, and the truncation time (or the transplant waiting time) $A =$ transplant calendar date $- E$.

As pointed out by [44], the data can be viewed as biased-sampling data with right censoring mechanism I. Furthermore, the survival times T can be treated as a biased sample with a weight function equal to the distribution of transplant waiting time A. By fitting the transplant waiting time provided by [9, 44] used the weight function

$$W(t) = 1 - \exp(-0.027t^{0.925}).$$

Here, we analyzed the data using the additive risk model, $\lambda(t) = \lambda_0(t) + \beta_1 \times \text{age}$. We converted the survival time to years (i.e., $T/365$). Following [33], we deleted 27 patients lacking the T5 mismatch score and five patients with survival times of lower than 10 days from the 184 patients in the original dataset. Based on the remaining 152 patients, the proposed estimate is $\widehat{\beta}_1 = 0.010$ (SE $= 0.006$) with the p-value 0.040. The results show that there is a strong relation between survival time and age, which coincides with the conclusion of [44].

The p-values for checking the adequacy of the assumed model are calculated based on 1,000 simulated realizations. The graphical and numerical results for checking the adequacy of the assumed model are summarized in Figure 2. Figure 2(a) pertains to the functional forms of age (p-values 0.167), whereas Figure 2(b) pertains to the additive hazards assumption (p-values 0.420). In all the plots, the observed residual processes appear to be completely covered by the first 50 simulated ones. This result graphically demonstrates that there is no evidence against the assumed model.

6 Conclusion

This paper proposed an inference procedure for the regression parameters in the additive risk models with biased-sampling data. General biased-sampling data include the length-biased data and data with missing covariates. This is a parallel work of the proportional hazards model in [44] extending to the additive risk model. However, the proposed approach is different from his pseudo-partial likelihood approach for proportional hazards model. We propose an embedded estimating equation by embedding the biased data into a left-truncated and right-censored model, provided that the weight function is completely known or estimated by other methods. The proposed estimator is more efficient than the conditional estimating equation estimator because it incorporates the information in the biased-sampling data.

Similar to the estimator for the proportional hazards model in [44], the proposed estimator depends on that the weight function $W(t\,|\,\boldsymbol{Z})$ in biased density function is either completely known or can be estimated from other methods. The weight function $W(t\,|\,\boldsymbol{Z})$ is usually assumed to be equal to or proportional to the cumulative distribution function of the left-truncation or the probability of selecting complete cases in presence of missing covariates. If $W(t\,|\,\boldsymbol{Z})$ is misspecified, the proposed estimator may be inconsistent. However, as suggested in [30], $W(t\,|\,\boldsymbol{Z})$ can be possibly modeled by a parametric form or selected by varying $W(t\,|\,\boldsymbol{Z})$ in a sensitivity analysis. Further investigation is therefore needed. Furthermore, because the proposed estimator is not a nonparametric maximum likelihood estimator, its optimality cannot be guaranteed. The efficiency of the proposed estimating equation could be improved by constructing augmented estimators, especially when $W(t\,|\,\boldsymbol{Z})$ is completely unspecified. A careful investigation on this direction is also warranted in the further study.

Acknowledgements Zhang's work was supported by National Natural Science Foundation of China (Grant Nos. 11771133 and 11401194) and the Natural Science Foundation of Hunan Province of China (Grant No. 2017JJ3021). Zhao's work was supported by National Natural Science Foundation of China (Grant No. 11771366). Zhou's work was supported by the State Key Program of National Natural Science Foundation of China (Grant No. 71331006) and the State Key Program in the Major Research Plan of National Natural Science Foundation of China (Grant No. 91546202).

References

1 Aalen O O. A linear regression model for the analysis of life times. Stat Med, 1989, 8: 907–925

2 Andersen P K, Borgan O, Gill R D, et al. Statistical Models Based on Counting Processes. New York: Springer-Verlag, 1993

3 Asgharian M, M'Lan C, Wolfson D. Length-biased sampling with right censoring. J Amer Statist Assoc, 2002, 97: 201–209

4 Bilias Y, Gu M, Ying Z, et al. Towards a general asymptotic theory for Cox model with staggered entry. Ann Statist, 1997, 25: 662–682

5 Breslow N, Day N. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. IARC Sci Publ, 1986, 82: 1–406

6 Chen H Y, Little R J. Proportional hazards regression with missing covariates. J Amer Statist Assoc, 1999, 94: 896–908

7 Chen Y Q. Semiparametric regression in size-biased sampling. Biometrics, 2010, 66: 149–158

8 Cox D R, Oakes D. Analysis of Survival Data. London: Chapman and Hall, 1984

9 Crowley J, Hu M. Covariance analysis of heart transplant survival data. J Amer Statist Assoc, 1977, 72: 27–36

10 Ghosh D. Proportional hazards regression for cancer studies. Biometrics, 2008, 64: 141–148

11 Gross S. Weighted estimation in linear regression for truncated survival data. Scand J Statist, 1996, 23: 179–193

12 Hao M, Song X, Sun, L. Reweighting estimators for the additive hazards model with missing covariates. Canad J Statist, 2014, 42: 285–307

13 Huang C Y, Qin J. Nonparametric estimation for length-biased and right-censored data. Biometrika, 2011, 98: 177–186

14 Huang C Y, Qin J. Semiparametric estimation for the additive hazards model with left-truncated and right-censored data. Biometrika, 2013, 100: 877–888

15 Huang C Y, Qin J, Follmann D A. A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. Biometrika, 2012, 99: 199–210

16 Kim J, Song M S, Lee S. Goodness-of-fit tests for the additive risk model with ($p > 2$)-dimensional time-invariant covariates. Lifetime Data Anal, 1998, 4: 405–416

17 Kim J P, Lu W, Sit T, et al. A unified approach to semiparametric transformation models under general biased sampling schemes. J Amer Statist Assoc, 2013, 108: 217–227

18 Kim J P, Sit T, Ying Z. Accelerated failure time model under general biased sampling scheme. Biostatistics, 2016, 17: 576–588

19 Kong L, Cai J. Case-cohort analysis with accelerated failure time model. Biometrics, 2009, 65: 135–142

20 Kong L, Cai J, Sen P K. Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. Biometrika, 2004, 91: 305–319

21 Lai T, Ying Z. Estimating a distribution function with truncated and censored data. Ann Statist, 1991, 19: 417–442

22 Lai T, Ying Z. Rank regression methods for left-truncated and right-censored data. Ann Statist, 1991, 19: 531–556

23 Li Y, Ma H, Wang D, et al. Analyzing the general biased data by additive risk model. Sci China Math, 2017, 60: 685–700

24 Lin D Y, Wei L J, Yang I, et al. Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc Ser B Stat Methodol, 2000, 62: 711–730

25 Lin D Y, Wei L J, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika, 1993, 80: 557–572

26 Lin D Y, Ying Z. Semiparametric analysis of the additive risk model. Biometrika, 1994, 81: 61–71

27 Lin W. Missing covariates and high-dimensional variable selection in additive hazards regression. PhD Thesis. Los Angeles: University of Southern California, 2011

28 Lu W, Tsiatis A A. Semiparametric transformation models for the case-cohort study. Biometrika, 2006, 93: 207–214

29 Lu X, Qi Y. Empirical likelihood for the additive risk model. Probab Math Statist, 2004, 24: 419–431

30 Luo X, Tsai W Y. Nonparametric estimation for right-censored length-biased data: A pseudo-partial likelihood approach. Biometrika, 2009, 96: 873–886

31 Luo X, Tsai W Y, Xu Q. Pseudo-partial likelihood estimators for the Cox regression model with missing covariates. Biometrika, 2009, 96: 617–633

32 Ma H, Zhang F, Zhou Y. Composite estimating equation approach for additive risk model with length-biased and right-censored data. Statist Probab Lett, 2015, 96: 45–53

33 Miller R, Halpern J. Regression with censored data. Biometrika, 1982, 69: 521–531

34 Muttlak H, McDonald L. Ranked set sampling with size-biased probability of selection. Biometrics, 1990, 46: 435–445

35 Paik M C, Tsai W-Y. On using the Cox proportional hazards model with missing covariates. Biometrika, 1997, 84: 579–593

36 Pollard D. Empirical Processes: Theory and Applications. Hayward: Institute of Mathematics Statistics, 1990

37 Qi L, Wang C, Prentice R L. Weighted estimators for proportional hazards regression with missing covariates. J Amer Statist Assoc, 2005, 100: 1250–1263

38 Qin J, Ning J, Liu H, et al. Maximum likelihood estimations and EM algorithms with length-biased data. J Amer Statist Assoc, 2011, 106: 1434–1449

39 Qin J, Shen Y. Statistical methods for analyzing right-censored length-biased data under Cox model. Biometrics, 2010, 66: 382–392

40 Robins J M, Rotnitzky A, Zhao L P. Estimation of regression coefficients when some regressors are not always observed. J Amer Statist Assoc, 1994, 89: 846–866

41 Rubin D B. Inference and missing data. Biometrika, 1976, 63: 581–592

42 Shen Y, Ning J, Qin J. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. J Amer Statist Assoc, 2009, 104: 1192–1202

43 Shen Y, Ning J, Qin J. Nonparametric and semiparametric regression estimation for length-biased survival data. Lifetime Data Anal, 2017, 23: 3–24

44 Tsai W Y. Pseudo-partial likelihood for proportional hazards models with biased-sampling data. Biometrika, 2009, 96: 601–615

45 Tsai W Y, Jewell N P, Wang M C. A note on the product-limit estimator under right censoring and left truncation. Biometrika, 1987, 74: 883–886

46 Tsui K-L, Jewell N P, Wu C. A nonparametric approach to the truncated regression problem. J Amer Statist Assoc, 1988, 83: 785–792

47 Turnbull B W. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc Ser B Stat Methodol, 1976, 38: 290–295

48 Van der Vaart A W. Asymptotic Statistics. Cambridge: Cambridge University Press, 1998

49 Vardi Y. Nonparametric estimation in the presence of length bias. Ann Statist, 1982, 10: 616–620

50 Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. Biometrika, 1989, 76: 751–761

51 Wang C, Chen H Y. Augmented inverse probability weighted estimator for Cox missing covariate regression. Biometrics, 2001, 57: 414–419

52 Wang M C. Hazards regression analysis for length-biased data. Biometrika, 1996, 83: 343–354

53 Wang M C, Brookmeyer R, Jewell N P. Statistical models for prevalent cohort data. Biometrics, 1993, 49: 1–11

54 Wang M C, Jewell N P, Tsai W Y. Asymptotic properties of the product limit estimate under random truncation. Ann Statist, 1986, 14: 1597–1605

55 Xu Q, Paik M C, Luo X, et al. Reweighting estimators for Cox regression with missing covariates. J Amer Statist Assoc, 2009, 104: 1155–1167

56 Yuen K, Burke M. A test of fit for a semiparametric additive risk model. Biometrika, 1997, 84: 631–639

57 Zhang F, Chen X, Zhou Y. Proportional hazards model with varying coefficients for length-biased data. Lifetime Data Anal, 2014, 20: 132–157

58 Zhao M, Lin C, Zhou Y. Analyzing right-censored length-biased data with additive hazards model. Acta Math Appl Sin Engl Ser, 2017, 33: 893–908

Appendix A Proofs of asymptotic results

Let $F(\cdot)$, $W(\cdot)$ and $G(\cdot)$ be the distributions of failure time T^*, truncation time A^* and residual censoring variable C, respectively. Let l_F and u_F be the lower and upper bounds of the support of any distribution F, respectively. To establish the asymptotic distributions of the proposed estimators, we require the following regulatory conditions:

(i) There exists a constant $\tau > 0$ such that $P(Y_i(t) = 1, t \in [0, \tau]) > 0$;

(ii) $W(t\,|\,\boldsymbol{z})$ has a density function $w(t\boldsymbol{z})$ in its support $[0, u_W]$ with $u_W \geqslant u_F$;

(iii) $l_F > 0$;

(iv) either $u_G < u_F$ with $G(u_G) > 0$ or $u_G \geqslant u_F$;

(v) $\boldsymbol{Z}_i(\cdot)$ is bounded with total variations.

Appendix A.1 Proof of Theorem 2.2

The proof is divided into three steps:

(1) We show that, for sufficiently large n, the estimating equation $\widehat{U}(\boldsymbol{\beta})$ has the similar local behavior to $U(\boldsymbol{\beta})$ in the compact neighborhood. Under the regularity conditions (i)–(iv), Luo and Tsai [30] proved that

$$\sup_{t\in[0,\tau]} |\widehat{\pi}_i(t) - \pi_i(t)| \to 0,$$

almost surely. Coupled with condition (v), this implies that

$$\sup_{t\in[0,\tau]} |\widehat{\boldsymbol{Z}}(t) - \overline{\boldsymbol{Z}}(t)| \to 0,$$

almost surely. For some positive δ_n with $\delta_n \to 0$ as $n \to \infty$, we define the class

$$\mathcal{F} = \bigg\{ f(t) = \int_0^\tau (g(t) - \overline{\boldsymbol{Z}}(t))\pi(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(t)dt : \boldsymbol{\beta} \in \mathcal{B},$$
$$g \text{ is bounded function and } \sup_{t\in[0,\tau]} |g(t) - \overline{\boldsymbol{Z}}(t)| \leqslant \delta_n \bigg\}.$$

Thus, by definition,

$$\sup_{f\in\mathcal{F}} |f| \leqslant \delta_n \sup_{\boldsymbol{\beta}\in\mathcal{B}} \bigg| \int_0^\tau \pi(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(t)dt \bigg|.$$

Furthermore, it follows from [48] that $\mathcal{F}$ is Glivenko-Cantelli. Hence, $\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - Pf| \to 0$ almost surely. For a sufficiently large n,

$$\bigg| n^{-1}\sum_{i=1}^n \int_0^\tau [\widehat{\boldsymbol{Z}}(t) - \overline{\boldsymbol{Z}}(t)]\pi_i(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt \bigg| \leqslant \sup_{f\in\mathcal{F}} |Pf| + \sup_{f\in\mathcal{F}} |\mathbb{P}_n f - Pf|.$$

Thus, we show that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \bigg| n^{-1}\sum_{i=1}^n \int_0^\tau [\widehat{\boldsymbol{Z}}(t) - \overline{\boldsymbol{Z}}(t)]\pi_i(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt \bigg| \to 0$$

almost surely. Similarly, we can show that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \bigg| n^{-1}\sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i(t)[\widehat{\pi}_i(t) - \pi_i(t)]\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt \bigg| \to 0,$$
$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \bigg| n^{-1}\sum_{i=1}^n \int_0^\tau \overline{\boldsymbol{Z}}(t)[\widehat{\pi}_i(t) - \pi_i(t)]\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt \bigg| \to 0$$

almost surely. Thus, $\sup_{\boldsymbol{\beta}\in\mathcal{B}} |\widehat{U}(\boldsymbol{\beta}) - U(\boldsymbol{\beta})| \to 0$ almost surely.

(2) We show the consistency of $\widehat{\boldsymbol{\beta}}$. In the view of regularity conditions, we can use the strong law of large numbers to show that $\widehat{U}(\boldsymbol{\beta})$ converges almost surely to

$$\mathcal{U}(\boldsymbol{\beta}) \equiv \mathrm{E}\bigg[\int_0^\tau \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{z}}(t)\}[dN_i(t) - \pi_i(t)\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt]\bigg],$$

for every $\boldsymbol{\beta}$, where $\overline{\boldsymbol{z}}(t) = s^{(1)}(t)/s^{(0)}(t)$, and $s^{(k)}(t)$ is the limitation of $S^{(k)}(t) = n^{-1}\sum_{i=1}^n \widehat{\pi}_i(t)\boldsymbol{Z}_i^{\otimes k}(t)$ for $k = 0, 1, 2$. Clearly, $\mathcal{U}(\boldsymbol{\beta}_0) = 0$ by double expectation. Note that $\partial\widehat{U}(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta} = -D_2(\boldsymbol{\beta}_0)$ is negative semidefinite for sufficiently large n, we have $\widehat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_0$ almost surely by a standard argument.

(3) To establish the asymptotic normal of $\sqrt{n}U(\boldsymbol{\beta}_0)$ and $\widehat{\boldsymbol{\beta}}$, we define

$$\widetilde{D} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau [\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}_i(t)]^{\otimes 2}\pi_i(t)dt,$$
$$\widetilde{\Sigma} = \frac{1}{n}\sum_{i=1}^n \bigg[\int_0^\tau \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\}\{dN_i(t) - \pi_i(t)\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt\}\bigg]^{\otimes 2},$$

with $\overline{\boldsymbol{Z}}(t) = \sum_i^n \pi_i(t)\boldsymbol{Z}_i(t)/\sum_i^n \pi_i(t)$. With the help of similar arguments to that of the first step, we observe that $\widetilde{D}$ and $\widetilde{\Sigma}$ are consistent for D and Σ, respectively.

In the view of equation (2.5), we have

$$U(\boldsymbol{\beta}_0, t) = \overline{M}_{\boldsymbol{Z}}(t) - \int_0^t \overline{\boldsymbol{Z}}(u)d\overline{M}(u),$$

where $\overline{M}(t) = n^{-1}\sum_{i=1}^n M_i(t)$ and $\overline{M}_{\boldsymbol{Z}}(t) = n^{-1}\sum_{i=1}^n \int_0^t \boldsymbol{Z}_i(u)dM_i(u)$. Clearly, $U(\boldsymbol{\beta}_0) = U(\boldsymbol{\beta}_0, \tau)$. Note that both of them are sums of i.i.d. zero-mean terms for fixed t. By the multivariate central limit theorem, $(n^{1/2}\overline{M}(t), n^{1/2}\overline{M}_{\boldsymbol{Z}}(t))$ converges in finite dimensional distributions to a zero-mean Gaussian process with continuous sample paths, say $(\mathcal{W}_M, \mathcal{W}_{M_{\boldsymbol{Z}}})$. Clearly, $M_i(\cdot)$ is the difference of two monotonic functions in t. Since the condition (iii) implies that $\boldsymbol{Z}_i(\cdot)$ is bounded, we may assume without loss of generality that $\boldsymbol{Z}_i(\cdot) \geqslant 0$. Then $\int_0^t \boldsymbol{Z}_i(u)M_i(u)$ is also the difference of two monotonic functions in t. By the functional central limit theorem (see [36, Theorem 10.6]), $(n^{1/2}\overline{M}(t), n^{1/2}\overline{M}_{\boldsymbol{Z}}(t))$ is tight and thus converges weakly to $(\mathcal{W}_M, \mathcal{W}_{M_{\boldsymbol{Z}}})$.

Since $\boldsymbol{Z}_i(\cdot) \geqslant 0$ for $i = 1, \ldots, n$, $S^{(1)}(t)$ is also a monotonic function in t. It then follows from [4, Lemama A.3] that

$$n^{1/2}\int_0^t \frac{d\overline{M}(u)}{S^{(0)}(u)} \to \int_0^t \frac{d\mathcal{W}_M(u)}{s^{(0)}(u)},$$

uniformly in t almost surely. In addition, we obtain

$$n^{1/2}\int_0^t \frac{S^{(1)}(u)}{S^{(0)}(u)}d\overline{M}(u) \to \int_0^t \frac{s^{(1)}(u)}{s^{(0)}(u)}d\mathcal{W}_M(u),$$

uniformly in t almost surely. This convergence, coupled with the convergence of $n^{1/2}\overline{M}_{\boldsymbol{Z}}$ to $\mathcal{W}_{M_{\boldsymbol{Z}}}$, yields the uniform convergence of $n^{1/2}U(\beta_0, t)$ to $\mathcal{W}_{M_{\boldsymbol{Z}}}(t) - \int_0^t \overline{z}(u)d\mathcal{W}_M(u)$ almost surely. The limit covariance function is

$$\Sigma(s,t) = \mathrm{E}\bigg[\int_0^s \{\boldsymbol{Z}_1(u) - \overline{\boldsymbol{z}}(u)\}dM_1(u)\int_0^t \{\boldsymbol{Z}_1(u) - \overline{\boldsymbol{z}}(u)\}^{\mathrm{T}}dM_1(u)\bigg],$$

which can be approximated by

$$\widetilde{\Sigma}(s,t) = n^{-1}\sum_{i=1}^n \int_0^s \{\boldsymbol{Z}_i(u) - \overline{\boldsymbol{Z}}(u)\}dM_i(u)\int_0^t \{\boldsymbol{Z}_i(u) - \overline{\boldsymbol{Z}}(u)\}^{\mathrm{T}}dM_i(u).$$

Furthermore, expanding $U(\widehat{\beta})$ around β_0, we have

$$-\sqrt{n}U(\boldsymbol{\beta}_0) = \sqrt{n}U(\widehat{\boldsymbol{\beta}}) - U(\boldsymbol{\beta}) = \frac{\partial U(\boldsymbol{\beta}^*)}{\partial\boldsymbol{\beta}}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \beta_0),$$

where $\boldsymbol{\beta}^*$ lies on the segment between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$. Therefore,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = -\left[\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_0}\right]^{-1}\sqrt{n}U(\boldsymbol{\beta}_0)+o_P(1) = \widetilde{D}^{-1}\sqrt{n}U(\boldsymbol{\beta}_0)+o_P(1),$$

which converges to a multivariate normal distribution with zero-mean and covariance matrix $D^{-1}\Sigma D^{-1}$, by Slutsky's theorem. Here, $\Sigma = \Sigma(\tau,\tau)$. For the future reference, we display the asymptotic approximation

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = D^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{z}}(u)]dM_i(u)+o_p(1).$$

Appendix A.2 Proof of Theorem 2.3

We make the simple decomposition

$$\begin{aligned}
&\sqrt{n}[\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t)-\Lambda_0(t)]\\
&\quad= \sqrt{n}[\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t)-\widehat{\Lambda}_0(\boldsymbol{\beta}_0,t)]+\sqrt{n}[\widehat{\Lambda}_0(\boldsymbol{\beta}_0,t)-\Lambda_0(t)]\\
&\quad= \sqrt{n}\int_0^t \frac{\sum_{i=1}^n \widehat{\pi}_i(u)[\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}_i(u)-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_i(u)]}{\sum_{i=1}^n \widehat{\pi}_i(u)}du+\sqrt{n}\int_0^t \frac{d\overline{M}_i(u)}{S^{(0)}(u)}.
\end{aligned}$$

By the arguments of Appendix A.1, the second term is tight and equals to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{dM_i(u)}{s^{(0)}(u)}+o_p(1).$$

By the Taylor series expansion, the first term equals to $-H^{\mathrm{T}}(\boldsymbol{\beta}^*,t)\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$, where

$$H(\boldsymbol{\beta},t) = \int_0^t \frac{n^{-1}\sum_{i=1}^n \widehat{\pi}_t(u)\boldsymbol{Z}_i(u)du}{n^{-1}\sum_{i=1}^n \widehat{\pi}_i(u)},$$

and $\boldsymbol{\beta}^*$ is on the line segment between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. By [24, Lemma 1] and the uniform strong law of large numbers, $H(\boldsymbol{\beta}_0,t)$ converges almost surely to

$$\boldsymbol{h}(\boldsymbol{\beta}_0,t) = \int_0^t \frac{\mathrm{E}[\pi_1(u)\boldsymbol{Z}_1]du}{s^{(0)}(u)}$$

uniformly in t. Hence, the first term on the right-hand side of equation is tight and equals to

$$-\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{\beta}_0,t)D^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{z}}(u)]dM_i(u)+o_p(1).$$

Then,

$$\sqrt{n}[\widehat{\Lambda}_0(\widehat{\boldsymbol{\beta}},t)-\Lambda_0(t)] = n^{-1/2}\sum_{i=1}^{n}\Psi_i(t)+o_p(1),$$

which converges weakly to a zero-mean Gaussian process with covariance function

$$\eta(s,t) \equiv \mathrm{E}\{\Psi_1(s)\Psi_1(t)^{\mathrm{T}}\},$$

where

$$\Psi_i(t) = \int_0^t \frac{dM_i(u)}{s^{(0)}(u)} - \boldsymbol{h}^{\mathrm{T}}(\boldsymbol{\beta}_0,t)D_2^{-1}\int_0^{\tau}[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{z}}(u)]dM_i(u).$$

As argued before, $\eta(s,t)$ can be approximated by $\widehat{\eta}(s,t) = n^{-1}\sum_{i=1}^n \widehat{\Psi}_i(s)\widehat{\Psi}_i(t)^{\mathrm{T}}$, where

$$\widehat{\Psi}_i(t) = \int_0^t \frac{d\widehat{M}_i(u)}{S^{(0)}(u)} - H^{\mathrm{T}}(\widehat{\boldsymbol{\beta}},t)\widehat{D}^{-1}\int_0^{\tau}\left[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{Z}}(u)\right]d\widehat{M}_i(u).$$

This completes the proof.

Appendix A.3 Proof of Theorem 3.1

We further establish the weak convergence of $Q(t,z)$ and $\widehat{Q}(t,z)$ under (2.2). By Taylor series expansion and some simple algebra, we have

$$Q(t,z)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t[q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u)\leqslant\boldsymbol{z})-S_q(u,\boldsymbol{z})]dM_i(u)-B_q(t,\boldsymbol{z})\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0), \tag{A.1}$$

where

$$S_q(t,\boldsymbol{z})=\frac{\sum_{i=1}^n q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u)\leqslant\boldsymbol{z})\pi_i(t)}{\sum_{i=1}^n\pi_i(t)},$$

and

$$B_q(t,\boldsymbol{z})=\frac{1}{n}\sum_{i=1}^{n}\int_0^t q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u)\leqslant\boldsymbol{z})\pi_i(u)[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{Z}}(u)]^{\mathrm{T}}du.$$

By the strong consistency of $\widehat{\boldsymbol{\beta}}$ and the uniform strong law of large numbers, $S_q(u,\boldsymbol{z})$ and $B_q(u,\boldsymbol{z})$ converge almost surely to $s_q(u,\boldsymbol{z})$ and $b_q(u,\boldsymbol{z})$. Note that the first term on the right-hand side of (A.1) takes a similar form to $U(\boldsymbol{\beta}_0,t)$, its tightness follows from the arguments given in Appendix A.1. On the other hand, the second term is tight since $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ converges weakly and $B_q(t,\boldsymbol{z})$ converges uniformly to $b_q(t,\boldsymbol{z})$. Thus, $Q(t,\boldsymbol{z})$ is tight.

Furthermore, by (A.1) and the convergence of $\widehat{\boldsymbol{\beta}},S_q(t,\boldsymbol{z})$ and $B_q(t,\boldsymbol{z})$, we can derive that

$$Q(t,\boldsymbol{z})=n^{-1/2}\sum_{i=1}^{n}\gamma_i(t,\boldsymbol{z})+o_p(1),$$

where

$$\gamma_i(t,\boldsymbol{z})=\int_0^t[q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u)\leqslant\boldsymbol{z})-s_q(u,\boldsymbol{z})]dM_i(u)-b_q(t,\boldsymbol{z})D^{-1}\int_0^\tau[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{Z}}(u)]dM_i(u).$$

By the multivariate central limit theorem and the tightness of $Q(t,\boldsymbol{z})$, $Q(t,\boldsymbol{z})$ converges weakly to a zero-mean Gaussian process with covariance function $\mathrm{E}[\gamma_1(t,\boldsymbol{z})\gamma_1(t^*,\boldsymbol{z}^*)^{\mathrm{T}}]$ at $(t,\boldsymbol{z})$ and $(t^*,\boldsymbol{z}^*)$. This covariance function can be consistently estimated by

$$n^{-1}\sum_{i=1}^{n}\widehat{\gamma}_i(t,\boldsymbol{z})\widehat{\gamma}_i(t^*,\boldsymbol{z}^*)^{\mathrm{T}},$$

where

$$\widehat{\gamma}_i(t,\boldsymbol{z})=\int_0^t[q(\boldsymbol{Z}_i(u))I(\boldsymbol{Z}_i(u)\leqslant\boldsymbol{z})-S_q(u,\boldsymbol{z})]d\widehat{M}_i(u)-B_q(t,\boldsymbol{z})\widehat{D}^{-1}\int_0^\tau[\boldsymbol{Z}_i(u)-\overline{\boldsymbol{Z}}(u)]d\widehat{M}_i(u).$$

Appendix A.4 Proof of Theorem 3.2

We establish the weak convergence of $\widehat{Q}(t,\boldsymbol{z})$ in this subsection. Conditional on the data

$$\{N_i(t),Y_i(t),\boldsymbol{Z}_i(t):i=1,\dots,n\},$$

the only random components in $\widehat{Q}(t,\boldsymbol{z})=n^{-1/2}\sum_{i=1}^n\widehat{\gamma}_i(t,\boldsymbol{z})\xi_i$ are $\{\xi_1,\dots,\xi_n\}$. A straightforward calculation shows $\mathrm{E}\widehat{Q}(t,z)=0$ and

$$\mathrm{Cov}(\widehat{Q}(t,\boldsymbol{z}),\widehat{Q}(t^*,\boldsymbol{z}^*))=n^{-1}\sum_{i=1}^{n}\widehat{\gamma}_i(t,\boldsymbol{z})\widehat{\gamma}_i(t^*,\boldsymbol{z}^*)^{\mathrm{T}}.$$

By the multivariate central limit theorem, conditional on the data, $\widehat{Q}(t,\boldsymbol{z})$ converges in finite dimensional distribution to a zero-mean Gaussian process with covariance function

$$n^{-1}\sum_{i=1}^{n}\widehat{\gamma}_i(t,\boldsymbol{z})\widehat{\gamma}_i(t^*,\boldsymbol{z}^*)^{\mathrm{T}}.$$

Furthermore, by the strong consistency of $\widehat{\boldsymbol{\beta}}$ and the strong law of large numbers,

$$n^{-1}\sum_{i=1}^{n}\widehat{\gamma}_i(t,\boldsymbol{z})\widehat{\gamma}_i(t^*,\boldsymbol{z}^*)^{\mathrm{T}}$$

converges to

$$n^{-1}\sum_{i=1}^{n}\widehat{\gamma}_i(t,\boldsymbol{z})\widehat{\gamma}_i(t^*,\boldsymbol{z}^*)^{\mathrm{T}}$$

almost surely uniformly in s and t. Then $\widehat{Q}(t,\boldsymbol{z})$ converges to the same limiting distribution as $Q(t,\boldsymbol{z})$ provided that $\widehat{Q}(t,\boldsymbol{z})$ is tight. Indeed, since $\widehat{Q}(t,\boldsymbol{z})$ comprises monotone functions in t, the tightness of $\widehat{Q}(t,\boldsymbol{z})$ follows from the functional central limit theorem (see [36, p. 53]).