Computational Statistics and Data Analysis 157 (2021) 107157

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Regression analysis of censored data with nonignorable missing covariates and application to Alzheimer Disease

Mingyue Du [a], Huiqiong Li [b,*], Jianguo Sun [c]

[a] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China*
[b] *Department of Statistics, Yunnan University, Kunming, 650091, China*
[c] *Department of Statistics, University of Missouri, Columbia, MO, 65211, USA*

ARTICLE INFO

Article history:
Received 14 February 2020
Received in revised form 15 October 2020
Accepted 9 December 2020
Available online 27 December 2020

Keywords:
EM algorithm
Interval-censored data
Missing covariate
Semiparametric transformation models

ABSTRACT

In this paper, we discuss regression analysis of censored failure time data when there exist missing covariates and more specifically, we will consider interval-censored data, a general form of censored data, and the nonignorable missing. Although many methods have been proposed in the literature for censored data with missing covariates, they only apply to limited situations and it does not seem to exist an established procedure for the situation discussed here. For the analysis, we employ the semiparametric linear transformation model and develop a two-step estimation procedure. In addition, the asymptotic properties of the resulting estimators are established and a Poisson variable-based EM algorithm is provided for the implementation of the proposed estimation procedure. Finally the proposed approach is applied to an Alzheimer Disease study that motivated this investigation.

1. Introduction

In this paper, we discuss regression analysis of censored failure time data when there exist missing covariates. It is well-known that in general, such analysis highly depends on the censoring mechanism and it is relatively easy if one faces missing completely at random or missing at random (Little and Rubin, 2002). However, sometimes one may face or has to deal with nonignorable missing, meaning that the missing may depend on both the observed and missing values (Lipsitz et al., 1999), and the analysis of such data is challenging or difficult since it usually requires some assumptions that are hard to verify. In the following, we will discuss regression analysis of interval-censored data under the semiparametric linear transformation model with nonignorable missing covariates.

This work was motivated by an Alzheimer's Disease study, the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is a longitudinal multi-centre study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of the Alzheimer's disease (AD). It begun in 2004 and has been recruited the participants across North America who agreed to complete a variety of imaging and clinical assessments. Among others, one main objective of the study is to detect AD at the earliest possible stage and identify ways to track the diseases progression with biomarkers. During the study, as most of longitudinal or follow-up studies, some participants miss the scheduled visits, drop out the study and/or fail to provide covariate information. Also as most of follow-up studies, only interval-censored observations are available for the occurrence times of the events of interest such as the AD conversion. More details will be given below.

* Corresponding author.
E-mail address: lihuiqiong@ynu.edu.cn (H. Li).

https://doi.org/10.1016/j.csda.2020.107157

In addition to the study described above, missing data occur in many other areas such as longitudinal follow-up studies and sample survey and also in many forms in terms of missing parts and missing mechanism (Graham, 2012; Lipsitz et al., 1999; Little and Rubin, 2002; Molenberghs et al., 2015). In longitudinal studies, for example, it is common that study subjects miss some scheduled visits and/or drop out the study before the end of the study and thus give some missing values. It is apparent that such missing can easily be nonignorable if the missing visits or drop-out is related to the event or variable under investigation such that the event may be some status of a breast cancer patient while the drop-out is due to death. In sample survey, it is also common to have missing values on both response variables and covariates with the missingness being to depend on missing values and thus nonignorable, and among others, Little and Rubin (2002) described many such nonignorable examples. In the case of interval-censored data, a general example of nonignorable missing covariates occurs when there exist some internal covariates or correlated longitudinal covariates (Kalbfleisch and Prentice, 2002; Wullfsohn and Tsiatis, 1997).

As mentioned above, many methods have been proposed in the literature for regression analysis of censored failure time data with missing covariates (Hu et al., 2015; Qi et al., 2005; Luo et al., 2009; Ning et al., 2018; Wang and Chen, 2001; Xu et al., 2009). However, most of them apply only to right-censored data, the censored data with missing completely at random or missing at random, or the data arising from a specific regression model such as the proportional hazards model. It is well-known that when the missing is nonignorable, the application of the methods developed under the simpler missing mechanism could yield seriously biased estimation, and the proportional hazards model may be too restrictive or not fit censored data well sometimes. Limited research also exists for right-censored data with nonignorable missing covariates. For example, Cook et al. (2011) gave a maximum likelihood estimation approach by utilizing supplementary information under the proportional hazards model. In the following, we will consider interval-censored data, a more general form of censored data that includes right-censored data as a special case (Finkelstein, 1986; Sun, 2006), under the semiparametric linear transformation model, a class of flexible models that includes many commonly used models as special cases (Chen et al., 2002; Zeng et al., 2016; Zhang and Zhao, 2013).

The remainder of this paper is organized as follows. In Section 2, we will first describe some notation and the models and some assumptions that will be used throughout the paper. A two-step estimation procedure is then proposed in Section 3. In the first step, an estimating equation approach is given to estimate the parameter involved in the missing mechanism and then an approximated maximum likelihood estimation procedure is developed for estimation of regression parameters along with others. Furthermore the asymptotic properties of the proposed estimators are established. In Section 4, for the implementation of the proposed estimation procedure, a novel EM algorithm is developed with the use of Poisson variables in the data augmentation part. Section 5 presents some results obtained from a simulation study conducted for the assessment of the finite sample properties of the proposed method, and they suggest that the approach works well for practical situations. In Section 6, we apply the proposed approach to the AD study described above and Section 7 contains some discussion and concluding remarks.

2. Models, assumptions and the likelihood function

Consider a failure time study that involves n independent subjects. For subject i, let T_i denote the failure time of interest and suppose that there exists a p-dimensional vector of covariates denoted by $(X_i, Z_i)'$, where X_i denotes the covariates that may suffer missing and Z_i the covariates that can always be observed, $i = 1, \ldots, n$. In the following, we will assume that given the covariates X_i and Z_i, the cumulative hazard function of T_i has the form

$$\Lambda_T(t|X_i, Z_i) = G\left\{\Lambda(t)\exp({X_i}^T\beta_1 + {Z_i}^T\beta_2)\right\}. \tag{1}$$

In the above, G is a known increasing function, $\Lambda(t)$ denotes an unknown increasing baseline cumulative hazard function, and $\beta = (\beta_1, \beta_2)'$ is a p-dimensional vector of regression parameters. Suppose that the main goal is to estimate β.

It is easy to see that the class of transformation models described above is flexible and includes many commonly used models as special cases (Chen et al., 2002; Zhang and Zhao, 2013). For example, one can obtain the proportional hazards model by letting $G(x) = x$ and it gives the proportional odds model when $G(x) = \log(1+x)$. Also it is not difficult to show that model (1) can be rewritten as a linear transformation model

$$\log \Lambda(t) = -{X_i}^T\beta_1 - {Z_i}^T\beta_2 + \epsilon\,,$$

and given X_i and Z_i, the survival function of T_i has the form

$$S(t|X_i, Z_i) = \exp\left\{-G\left[\Lambda(t)\exp({X_i}^T\beta_1 + {Z_i}^T\beta_2)\right]\right\}\,,$$

where ϵ is an error term with the distribution function $1 - \exp[-G\{\exp(x)\}]$.

To describe the observed interval-censored data, suppose that for subject i, there exist two observation times denoted by U_i and V_i with $U_i < V_i$, and define the indicator functions $\delta_{1i} = I(T_i \leqslant U_i)$, $\delta_{2i} = I(U_i < T_i \leqslant V_i)$, and $\delta_{3i} = 1 - \delta_{1i} - \delta_{2i}$. That is, one only observes whether the failure event of interest for subject i occurs before U_i, within $(U_i, V_i]$, or after V_i. Also define the indicator function $r_i = 1$ if X_i is observed and $r_i = 0$ otherwise and assume that the missing probability $\mathrm{P}(r_i = 1|X_i, Z_i) = \pi(X_i, Z_i) = \pi(\alpha; X_i, Z_i)$ is known up a vector of parameters α. That is, the missing mechanism is nonignorable or the missing not at random (MNAR). In the following, it will be also assumed that given the X_i and Z_i, T_i is independent of U_i and V_i, meaning that we have independent or non-informative censoring, and that given the covariates, the covariate missingness is independent of the observation on T_i.

Define $\mathcal{C} = \{i : r_i = 1\}$ and $\bar{\mathcal{C}} = \{i : r_i = 0\}$. Note that for the subject in $\mathcal{C}$, the density function of a general observation $\{U, V, \delta_1, \delta_2, \delta_3, Z, rX, r\}$ has the form

$$\begin{aligned} f(U, V, \delta_1, \delta_2, \delta_3, Z, rX, r) &= f(U, V, \delta_1, \delta_2, \delta_3, Z, X, r = 1) \\ &= f(X, Z) f(r = 1|X, Z) f(U, V, \delta_1, \delta_2, \delta_3|X, Z)\,, \end{aligned}$$

while for the subject in $\bar{\mathcal{C}}$, the corresponding density function takes the form

$$\begin{aligned} f(U, V, \delta_1, \delta_2, \delta_3, Z, rX, r) &= f(U, V, \delta_1, \delta_2, \delta_3, Z, r = 0) \\ &= \int f(X, Z) f(r = 0|X, Z) f(U, V, \delta_1, \delta_2, \delta_3|X, Z)\, dX\,. \end{aligned}$$

It follows that based on the observed data $\mathbf{O} = \{\, \mathbf{O_i} = (U_i, V_i, \delta_{i1}, \delta_{i2}, \delta_{i3}, Z_i, r_i X_i, r_i)\,, i = 1, \ldots, n \,\}$, the likelihood function can be written as

$$\begin{aligned} L(\beta_1, \beta_2, \Lambda, \alpha|\mathbf{O}) = \prod_{i\in\mathcal{C}} \Bigg\{ & \left(1 - \exp\left\{-G\left[\Lambda(U_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\}\right)^{\delta_{i1}} \\ & \left(\exp\left\{-G\left[\Lambda(U_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\} - \exp\left\{-G\left[\Lambda(V_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\}\right)^{\delta_{i2}} \\ & \left(\exp\left\{-G\left[\Lambda(V_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\}\right)^{\delta_{i3}} f(X_i|Z_i) f(Z_i) f(r_i = 1|X_i, Z_i) \Bigg\} \\ \times \prod_{i\in\bar{\mathcal{C}}} \Bigg\{ \int & \left(1 - \exp\left\{-G\left[\Lambda(U_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\}\right)^{\delta_{i1}} \\ & \left(\exp\left\{-G\left[\Lambda(U_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\} - \exp\left\{-G\left[\Lambda(V_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\}\right)^{\delta_{i2}} \\ & \left(\exp\left\{-G\left[\Lambda(V_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\}\right)^{\delta_{i3}} f(X_i|Z_i) f(Z_i) f(r_i = 0|X_i, Z_i) dX_i \Bigg\}\,. \end{aligned}$$

Let $(L_i, R_i]$ denote the smallest interval that brackets T_i, or define $L_i = 0$ and $R_i = U_i$ if $\delta_{i1} = 1$, $L_i = U_i$ and $R_i = V_i$ if $\delta_{i2} = 1$, and $L_i = V_i$, $R_i = \infty$ if $\delta_{i3} = 1$. Then the likelihood function above can be rewritten as

$$\begin{aligned} L(\beta_1, \beta_2, \Lambda, \alpha|\mathbf{O}) = \prod_{i\in\mathcal{C}} \Bigg\{ \Bigg(& \exp\left\{-G\left[\Lambda(L_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\} \\ & - \exp\left\{-G\left[\Lambda(R_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\} \Bigg) f(X_i|Z_i) f(Z_i) f(r_i = 1|X_i, Z_i) \Bigg\} \\ \times \prod_{i\in\bar{\mathcal{C}}} \Bigg\{ \int \Bigg(& \exp\left\{-G\left[\Lambda(L_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\} \\ & - \exp\left\{-G\left[\Lambda(R_i)\exp(X_i^T\beta_1 + Z_i^T\beta_2)\right]\right\} \Bigg) f(X_i|Z_i) f(Z_i) f(r_i = 0|X_i, Z_i) dx_i \Bigg\}\,. \end{aligned}$$

In the next section, we will discuss estimation of regression parameters as well as others.

3. Estimation procedures

Now we discuss estimation of regression parameters β_1 and β_2 as well as Λ and α. For this, we will first consider estimation of the parameter α in the missing probability $\pi(\alpha; X_i, Z_i)$ and present an estimating equation approach. Then an approximated maximum likelihood estimator is proposed for the others.

As above, let $f(X|Z)$ denote the probability density function of the covariate X given the covariate Z. For estimation of α, note that if $f(X|Z)$ was known, it would be natural to use the maximum likelihood estimator (Kim and Shao, 2013). Of course, this is not true and to deal with this, we assume that the r_i's are independently generated from a Bernoulli distribution with probability $\pi_i(\alpha) = \pi(\alpha; X_i, Z_i)$. If the X_i's were completely observed, the likelihood function of α would be

$$L(\alpha) = \prod_{i=1}^{n} \{\pi(\alpha; X_i, Z_i)\}^{\gamma_i} \{1 - \pi(\alpha; X_i, Z_i)\}^{1-r_i},$$

and one could obtain the maximum likelihood estimator of α by solving the score equation

$$S(\alpha) = \frac{\partial \log L(\alpha)}{\partial \alpha} = \sum_{i=1}^{n} s(\alpha; r_i, X_i, Z_i) = \sum_{i=1}^{n} \frac{r_i - \pi(\alpha; X, Z)}{\pi(\alpha; X, Z)\{1 - \pi(\alpha; X, Z)\}} \frac{\partial \pi(\alpha; X, Z)}{\partial \alpha} = 0\,.$$

However, because some of X_i are missing and thus the score equation above is not applicable. Corresponding to this, we can consider maximizing the observed likelihood function

$$L_{obs}(\alpha) = \prod_{i=1}^{n} \{\pi(\alpha; X_i, Z_i)\}^{\gamma_i} [\int \{1-\pi(\alpha; X_i, Z_i)\} f(X_i|Z_i) dX_i]^{1-r_i} ,$$

and the resulting maximum likelihood estimator of α can be obtained by solving the observed score equation

$$S_{obs}(\alpha) = \log L_{obs}(\alpha)/\partial\alpha = 0.$$

On the other hand, finding the solution to the observed score equation above is computationally challenging because it involves the integration with unknown parameters. Thus we propose to find the maximum likelihood estimator of α by solving the following mean score equation

$$\begin{aligned} \bar{S}(\alpha) &= n^{-1}\sum_{i=1}^{n} E\{s(\alpha; r_i, X_i, Z_i)|r_i, X_{obs,i}, Z_i\} \\ &= n^{-1}\sum_{i=1}^{n} [r_i s(\alpha; r_i, X_i, Z_i) + (1-r_i)E_0\{s(\alpha; r_i, X_i, Z_i)|Z_i = z_i, r_i = 0\}] = 0 \end{aligned} \tag{2}$$

which is equivalent to $S_{obs}(\alpha) = 0$ (Louis, 1982).

To solve the mean score equation $\bar{S}(\alpha) = 0$, we need to determine the conditional distribution $f(X|Z, r = 0)$ since we need to compute the conditional expectation of the score function $E_0(\cdot)$ in (2). For this, note that

$$f(X|Z, r=0) = f(X|Z, r=1)\frac{O(\alpha; X, Z)}{E\{O(\alpha; X, Z)|Z, r=1\}}$$

based on the Bayes formula, where

$$O(\alpha; X, Z) = \frac{f(r=0|X, Z)}{f(r=1|X, Z)} = \pi^{-1}(\alpha; X, Z) - 1 .$$

It thus follows that

$$\begin{aligned} E_0\{s(\alpha; r_i, X_i, Z_i)|Z_i = z_i, r_i = 0\} &= \frac{\int s(\alpha; r, X_i, Z)O(\alpha; X_i, Z)f(r=1|X_i, Z)dX_i}{\int O(\alpha; X_i, Z)f(r=1|X_i, Z)dX_i} \\ &\triangleq \frac{F_s(\alpha; Z_i, r_i)}{D(\alpha; Z_i)} . \end{aligned}$$

By using the kernel smoothing, we can estimate $F_s(\alpha; Z_i, r_i)$ and $D(\alpha; Z_i)$ by

$$\hat{F}_s(\alpha; Z, r) = (nh^d)^{-1}\sum_{j=1}^{n} r_j K_h(Z_j - Z)O(\alpha; X_j, Z)s(\alpha; r, X_j, Z) ,$$

and

$$\hat{D}(\alpha; Z) = (nh^d)^{-1}\sum_{j=1}^{n} r_j K_h(Z_j - Z)O(\alpha; X_j, Z) ,$$

respectively, where d denotes the dimension of Z, $K : \mathbb{R}^d \to \mathbb{R}$ is a kernel function, $K_h(Z) = K(Z/h)$, and h is an appropriate bandwidth that satisfies certain regularity conditions. Thus the mean score equation (2) can be approximated by (Morikawa et al., 2017; Zhao et al., 2017)

$$\hat{S}(\alpha) = n^{-1}\sum_{i=1}^{n}\left\{ r_i s(\alpha; r_i, X_i, Z_i) + (1-r_i)\frac{\hat{F}_s(\alpha; Z_i, r_i)}{\hat{D}(\alpha; Z_i)} \right\} = 0 . \tag{3}$$

It is worth noting that the estimation procedure for α above is valid for any parametric model and without the need of specifying a parametric distribution on the variable X.

Let $\hat{\alpha}$ denote the estimator of α given by the solution to Eq. (3). Then it is natural to estimate β_1, β_2 and Λ by maximizing the estimated or approximate likelihood function $L(\beta_1, \beta_2, \Lambda, \hat{\alpha}|\mathbf{O})$. Before this, note that to simplify the maximization, one can convert the class of transformation model given in (1) into the proportional hazards frailty model by using the Laplace transformation. Specifically, let ξ be a random variable whose density $f(\xi)$ is the inverse Laplace transformation of $\exp\{-G(t)\}$ or given by

$$\exp\{-G(t)\} = \int_0^\infty \exp(-t\xi)f(\xi)d\xi$$

with the support $[0, \infty)$. By letting $f(\xi_i)$ be the gamma density function with mean 1 and variance γ, we have $G(x) = \log(1 + \gamma x)/\gamma$, the logarithmic transformation function family. Then by omitting $f(Z)$ and $f(r = 1|X, Z)$, the likelihood function can be rewritten as

$$
\begin{aligned}
L(\beta_1, \beta_2, \Lambda, \hat{\alpha}|\mathbf{O}) = \prod_{i \in \mathcal{C}} \Bigg\{ \Bigg(\int_{\xi_i} \Bigg[& \exp\left\{-\xi_i \Lambda(L_i) \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2)\right\} \\
& - \exp\left\{-\xi_i \Lambda(R_i) \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2)\right\} \Bigg] f(\xi_i) d\xi_i \Bigg) f(X_i|Z_i) \Bigg\} \\
& \times \prod_{i \in \bar{\mathcal{C}}} \Bigg\{ \int \Bigg[\int_{\xi_i} \Bigg(\exp\left\{-\xi_i \Lambda(L_i) \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2)\right\} \\
& - \exp\left\{-\xi_i \Lambda(R_i) \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2)\right\} \Bigg) f(\xi_i) d\xi_i \Bigg] f(X_i|Z_i) f(r_i = 0|X_i, Z_i) dX_i \Bigg\} .
\end{aligned}
$$

For estimation of $\theta = (\beta_1, \beta_2, \Lambda)$, we will take the nonparametric approach that treats Λ as a step function at all different observation times and estimate it by the value, denoted by $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$, that maximizes the approximated likelihood function $L(\beta_1, \beta_2, \Lambda, \hat{\alpha}|\mathbf{O})$. Let $\theta_0 = (\beta_0, \Lambda_0)$ denote the true value of θ and define the distance between $\theta^1 = (\beta_1{}^1, \beta_2{}^1, \Lambda^1)$ and $\theta^2 = (\beta_1{}^2, \beta_2{}^2, \Lambda^2)$ as

$$
d(\theta^1, \theta^2) = \{ \|\beta_1{}^1 - \beta_1{}^2\|^2 + \|\beta_2{}^1 - \beta_2{}^2\|^2 + \|\Lambda^1 - \Lambda^2\|_2^2 \}^{1/2} ,
$$

where $\|v\|$ denotes the Euclidean norm of a vector v and $\|\Lambda^1 - \Lambda^2\|_2^2 = \int [\{\Lambda^1(u) - \Lambda^2(u)\}^2 + \{\Lambda^1(v) - \Lambda^2(v)\}^2] df(u, v)$. Then in the Appendix A, we will show that under some regularity conditions and as $n \to \infty$, $d(\hat{\theta}_n, \theta_0) \to 0$ almost surely and

$$
\sqrt{n}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma)
$$

in distribution with Σ given in Appendix A. That is, $\hat{\theta}_n$ is consistent and $\hat{\beta}_n$ asymptotically follows the normal distribution.

For inference about β, it is apparent that one needs to estimate Σ and one common approach would be to employ the Louis' Formula, proposed by Louis (1982) for covariance estimation when an EM algorithm is used, or the profile likelihood approach. However, both methods would require the determination of some quantities that are analytically complicated. Corresponding to this and by following Wen and Chen (2011) and others, we propose to employ the nonparametric bootstrap method (Efron, 1981; Su and Wang, 2016). Specifically, let Q be an integer and for each $1 \le q \le Q$, draw a new data set, denoted by $O^{(q)}$, of the sample size n with replacement from the original observed data $\{ O_i;\ i = 1, \ldots, n \}$. Let $\hat{\beta}_n^q$ denote the estimator of β defined above based on the bootstrap samples $O^{(q)}$, $q = 1, \ldots, Q$. respectively. Then one can estimate the covariance matrix of $\hat{\beta}_n$ by using the sample covariance matrix of the $\hat{\beta}_n^{(q)}$'s and the numerical results below suggest that it seems to work well.

4. A Poisson variable-based EM algorithm

In this section, we present an EM algorithm for the determination of the proposed estimator $\hat{\theta}_n$. Let $0 = t_0 < t_1 < \cdots < t_m$ denote all different, ordered time points of the L_i's and finite R_i's and λ_k the jump of Λ at t_k with $\lambda_0 = 0$. Then the approximated likelihood function $L(\beta_1, \beta_2, \Lambda, \hat{\alpha}|\mathbf{O})$ can be rewritten as

$$
\begin{aligned}
L(\beta_1, \beta_2, \Lambda, \hat{\alpha}|\mathbf{O}) = \prod_{i \in \mathcal{C}} \Bigg\{ \Bigg(\int_{\xi_i} & \exp\left\{ -\xi_i \sum_{t_k \leqslant L_i} \lambda_k \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2) \right\} \\
& \times \left[1 - \exp\{-\xi_i \sum_{L_i < t_k \leqslant R_i} \lambda_k \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2)\} \right]^{I(R_i < \infty)} f(\xi_i) d\xi_i \Bigg) f(X_i|Z_i) \Bigg\} \\
& \times \prod_{i \in \bar{\mathcal{C}}} \Bigg\{ \int \Bigg[\int_{\xi_i} \exp\left\{ -\xi_i \sum_{t_k \leqslant L_i} \lambda_k \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2) \right\} \\
& \times \left[1 - \exp\{-\xi_i \sum_{L_i < t_k \leqslant R_i} \lambda_k \exp(X_i{}^T \beta_1 + Z_i{}^T \beta_2)\} \right]^{I(R_i < \infty)} f(\xi_i) d\xi_i \Bigg] \\
& \times f(X_i|Z_i) f(r_i = 0|X_i, Z_i) dX_i \Bigg\} .
\end{aligned}
$$

To augment the observed data, by following Wang et al. (2016) and Zeng et al. (2016), given ξ_i, let $\{W_{ik};\ i=1,\ldots,n;\ k=1,\ldots,m\}$ be independent Poisson random variables with the means $\xi_i\lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)$. Then $L(\beta_1,\beta_2,\Lambda,\hat{\alpha}|\mathbf{O})$ can be written as

$$L(\beta_1,\beta_2,\Lambda,\hat{\alpha}|\mathbf{O})=\prod_{i\in\mathcal{C}}\left[\int_{\xi_i}\left\{\prod_{t_k\leqslant L_i}P(W_{ik}=0|\xi_i)\right\}\left\{1-P(\sum_{L_i<t_k\leqslant R_i}W_{ik}=0|\xi_i)\right\}^{I(R_i<\infty)}f(\xi_i)d\xi_i\right]$$
$$\times f(X_i|Z_i)\prod_{i\in\bar{\mathcal{C}}}\left\{\int\left[\int_{\xi_i}\left\{\prod_{t_k\leqslant L_i}P(W_{ik}=0|\xi_i)\right\}\left\{1-P(\sum_{L_i<t_k\leqslant R_i}W_{ik}=0|\xi_i)\right\}^{I(R_i<\infty)}f(\xi_i)d\xi_i\right]\right.$$
$$\left.f(X_i|Z_i)f(r_i=0|X_i,Z_i)dX_i\right\}.$$

Furthermore, for the subjects in $\bar{C}$, define the latent variable $\tilde{Z}$ which takes the values on the observed $Z_1,\ldots,Z_n$ and satisfies the equations $P(\tilde{Z}=Z_j|Z=Z_i)=w_{ji}$, $P(X=x_s|Z=Z_i,\tilde{Z}=Z_j)=P(X=x_s|\tilde{Z}=Z_j)=p_{sj}$, and $P(U,V,\delta_1,\delta_2,\delta_3|X,Z,\tilde{Z})=P(U,V,\delta_1,\delta_2,\delta_3|X,Z)$. Then we have that $P(X=x_s|Z=Z_i)=\sum_{j=1}^n w_{ji}p_{sj}$ for the subjects in $\bar{C}$.

By treating the ξ_i's, W_{ik}'s, X_i and $\tilde{Z}_i$'s as the missing data in the EM algorithm, we have the pseudo complete data log likelihood as

$$l_c(\theta)=\sum_{i\in\mathcal{C}}\left\{\sum_{k=1}^m I(t_k\leqslant R_i{}^*)\left[W_{ik}\log\{\xi_i\lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)\}-\xi_i\lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)\right.\right.$$
$$\left.\left.-\log W_{ik}!\right]+\log f(\xi_i)+\log f(X_i|Z_i)\right\}+\sum_{i\in\bar{\mathcal{C}}}\left\{\sum_{k=1}^m I(t_k\leqslant R_i{}^*)\left[W_{ik}\log\{\xi_i\lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)\}\right.\right.$$
$$\left.-\xi_i\lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)-\log W_{ik}!\right]+\log f(\xi_i)+\log f(X_i|\tilde{Z}_i)$$
$$\left.+\log f(\tilde{Z}_i|Z_i)+log(f(r_i=0|X_i,Z_i))\right\},$$

where $R_i{}^*=L_iI(R_i=\infty)+R_iI(R_i<\infty)$. Here we approximate $\log f(X_i|Z_i)$ and $\log f(X_i|\tilde{Z}_i)$ by

$$\sum_{j=1}^n w_{ji}\log P(X_i|Z=Z_j)=\sum_{s=1}^M I(X_i=x_s)\sum_{j=1}^n w_{ji}\log p_{sj}\,,$$

and

$$\sum_{j=1}^n\sum_{s=1}^M I(X_i=x_s,\tilde{Z}_i=Z_j)\log p_{sj}\,,$$

respectively,

$$w_{ji}=\frac{K(Z_j-Z_i)/a}{\sum_{j=1}^n K(Z_j-Z_i)/a}\,,$$

where $K(.)$ is a symmetric kernel function, a is a constant, and p_{sj} denotes the point mass of $P(X|Z_j)$ at x_s with $x_1,\ldots,x_M$ denoting the distinct observed values of X and the constraints $\sum_{s=1}^M p_{sj}=1$ and $p_{sj}\geq 0, j=1,\ldots,n$. When Z is discrete, we can choose a small enough such that $w_{ji}=I(j=i)$.

At the $(d+1)$th iteration of the E-step and given $\theta^{(d)}=(\beta_1^{(d)},\beta_2^{(d)},\lambda^{(d)},p^{(d)})'$, we need to determine $Q(\theta|\theta^{(d)})=E[l_c(\theta)|\mathbf{O},\theta^{(d)}]$, which has the form

$$Q(\theta|\mathbf{O},\theta^{(d)})=\sum_{i\in\mathcal{C}}\left\{\sum_{k=1}^m I(t_k\leqslant R_i{}^*)[E(W_{ik}\log\{\xi_i\lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)\})\right.$$
$$-\ \lambda_k\exp(X_i{}^T\beta_1+Z_i{}^T\beta_2)E(\xi_i)-E(\log W_{ik}!)]$$

$$
\begin{aligned}
&+\ E(\log f(\xi_i)) + \sum_{s=1}^{M} I(X_i = x_s) \sum_{j=1}^{n} w_{ji} \log p_{sj} \Bigg\} \\
&+\ \sum_{i\in\bar{C}} \Bigg\{ \sum_{k=1}^{m} I(t_k \leqslant R_i{}^*)[E(W_{ik} \log\{\xi_i \lambda_k \exp({X_i}^T \beta_1 + {Z_i}^T \beta_2)\}) \\
&-\ \lambda_k \exp({Z_i}^T\beta_2) E(\xi_i \exp({X_i}^T\beta_1)) - E(\log W_{ik}!)] + E(\log f(\xi_i)) \\
&+\ E\left(\sum_{j=1}^{n}\sum_{s=1}^{M} I(X_i = x_s, \tilde{Z}_i = Z_j) \log p_{sj}\right) + E\left(\sum_{j=1}^{n} I(\tilde{Z}_i = Z_j)\log w_{ji}\right) \\
&+\ E(\sum_{s=1}^{M} I(X_i = x_s) log(f(r_i = 0|X_i = x_s, Z_i))) \Bigg\}\ .
\end{aligned}
$$

For this, given the observed data for $i \in \bar{C}$, we can calculate the conditional expectations of $I(X_i = x_s)$ and $I(X_i = x_s, \tilde{Z}_i = Z_j)$ as

$$
\hat{q}_{is} = \frac{f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|X_i = x_s, Z_i)(\sum_{j=1}^{n} w_{ji}p_{sj})f(r_i = 0|X_i = x_s, Z_i)}{\sum_{s=1}^{M} f(U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}|X_i = x_s, Z_i)(\sum_{j=1}^{n} w_{ji}p_{sj})f(r_i = 0|X_i = x_s, Z_i)}\ , \tag{4}
$$

and

$$
\hat{\psi}_{sji} = \frac{w_{ji}p_{sj}}{\sum_{j=1}^{n} w_{ji}p_{sj}} \hat{q}_{is}\ , \tag{5}
$$

respectively, $s = 1, \ldots, M$ and $j = 1, \ldots, n$.

Also for the subjects belonging to C, we need to determine

$$
\begin{aligned}
E_X\{E(W_{ik})\} &= I(R_i < \infty)\lambda_k \exp({X_i}^T\beta_1 + {Z_i}^T\beta_2) \\
&\times \frac{\int_{\xi_i} \xi_i\{\exp(-\xi_i S_{i1}) - \exp(-\xi_i S_{i2})\}[1 - \exp(-\xi_i(S_{i2} - S_{i1}))]^{-1} f(\xi_i) d\xi_i}{\exp\{-G(S_{i1})\} - \exp\{-G(S_{i2})\}} I(L_i < t_k \le R_i)\,, \\
E_X\{E(W_{ik})X_i\} &= I(R_i < \infty)\lambda_k \exp({X_i}^T\beta_1 + {Z_i}^T\beta_2)X_i \\
&\times \frac{\int_{\xi_i} \xi_i\{\exp(-\xi_i S_{i1}) - \exp(-\xi_i S_{i2})\}[1 - \exp(-\xi_i(S_{i2} - S_{i1}))]^{-1} f(\xi_i) d\xi_i}{\exp\{-G(S_{i1})\} - \exp\{-G(S_{i2})\}} I(L_i < t_k \le R_i)\,, \\
E(\xi_i) &= I(R_i < \infty)\frac{\exp\{-G(S_{i1})\}G'(S_{i1}) - \exp\{-G(S_{i2})\}G'(S_{i2})}{\exp\{-G(S_{i1})\} - \exp\{-G(S_{i2})\}} + I(R_i = \infty)G'(S_{i1})\,, \\
E_X\{E(\xi_i)\exp(X_i^T\beta_1)\} &= \Bigg[\ I(R_i < \infty)\frac{\exp\{-G(S_{i1})\}G'(S_{i1}) - \exp\{-G(S_{i2})\}G'(S_{i2})}{\exp\{-G(S_{i1})\} - \exp\{-G(S_{i2})\}} + \\
&\qquad I(R_i = \infty)G'(S_{i1})\ \Bigg] \exp(X_i^T\beta_1)\,, \\
E_X\{E(\xi_i)\exp(X_i^T\beta_1)X_i\} &= \Bigg[\ I(R_i < \infty)\frac{\exp\{-G(S_{i1})\}G'(S_{i1}) - \exp\{-G(S_{i2})\}G'(S_{i2})}{\exp\{-G(S_{i1})\} - \exp\{-G(S_{i2})\}} + \\
&\qquad I(R_i = \infty)G'(S_{i1})\ \Bigg] \exp(X_i^T\beta_1)X_i\,,
\end{aligned}
$$

and

$$
\begin{aligned}
E_X\{E(\xi_i)\exp(X_i^T\beta_1)X_i^TX_i\} &= \Bigg[\ I(R_i < \infty)\frac{\exp\{-G(S_{i1})\}G'(S_{i1}) - \exp\{-G(S_{i2})\}G'(S_{i2})}{\exp\{-G(S_{i1})\} - \exp\{-G(S_{i2})\}} + \\
&\qquad I(R_i = \infty)G'(S_{i1})\ \Bigg] \exp(X_i^T\beta_1)X_i^TX_i\,.
\end{aligned}
$$

Correspondingly for the subjects belonging to $\bar{C}$, we need to determine

$$
\begin{aligned}
E_X\{E(W_{ik})\} &= \sum_{s=1}^{M} \hat{q}_{is} I(R_i < \infty)\lambda_k \exp({x_s}^T\beta_1 + {Z_i}^T\beta_2) \\
&\quad \times \frac{\int_{\xi_i} \xi_i\{\exp(-\xi_i S_{is1}) - \exp(-\xi_i S_{is2})\}[1 - \exp(-\xi_i(S_{is2} - S_{is1}))]^{-1} f(\xi_i) d\xi_i}{\exp\{-G(S_{is1})\} - \exp\{-G(S_{is2})\}} I(L_i < t_k \le R_i)\,,
\end{aligned}
$$

$$E_X\{E(W_{ik})X_i\} = \sum_{s=1}^{M} \hat{q}_{is} I(R_i < \infty)\lambda_k \exp({x_s}^T\beta_1 + {Z_i}^T\beta_2)x_s \times \frac{\int_{\xi_i} \xi_i\{\exp(-\xi_i S_{is1}) - \exp(-\xi_i S_{is2})\}[1-\exp(-\xi_i(S_{is2}-S_{is1}))]^{-1} f(\xi_i)d\xi_i}{\exp\{-G(S_{is1})\} - \exp\{-G(S_{is2})\}} I(L_i < t_k \le R_i)\,,$$

$$E_X\{E(\xi_i)\exp(X_i^T\beta_1)\} = \sum_{s=1}^{M} \hat{q}_{is} \left\{ I(R_i < \infty)\frac{\exp\{-G(S_{is1})\}G'(S_{is1}) - \exp\{-G(S_{is2})\}G'(S_{is2})}{\exp\{-G(S_{is1})\} - \exp\{-G(S_{is2})\}} + I(R_i = \infty)G'(S_{is1}) \right\} \exp(x_s^T\beta_1)\,,$$

$$E_X\{E(\xi_i)\exp(X_i^T\beta_1)X_i\} = \sum_{s=1}^{M} \hat{q}_{is} \left\{ I(R_i < \infty)\frac{\exp\{-G(S_{is1})\}G'(S_{is1}) - \exp\{-G(S_{is2})\}G'(S_{is2})}{\exp\{-G(S_{is1})\} - \exp\{-G(S_{is2})\}} + I(R_i = \infty)G'(S_{is1}) \right\} \exp(x_s^T\beta_1)x_s\,,$$

and

$$E_X\{E(\xi_i)\exp(X_i^T\beta_1)X_i^TX_i\} = \sum_{s=1}^{M} \hat{q}_{is} \left\{ I(R_i < \infty)\frac{\exp\{-G(S_{is1})\}G'(S_{is1}) - \exp\{-G(S_{is2})\}G'(S_{is2})}{\exp\{-G(S_{is1})\} - \exp\{-G(S_{is2})\}} + I(R_i = \infty)G'(S_{is1}) \right\} \exp(x_s^T\beta_1)x_s^Tx_s\,.$$

In the above, $S_{i1} = \sum_{t_k \leqslant L_i} \lambda_k \exp\{{X_i}^T\beta_1 + {Z_i}^T\beta_2\}$, $S_{i2} = \sum_{t_k \leqslant R_i} \lambda_k \exp\{{X_i}^T\beta_1 + {Z_i}^T\beta_2\}$, $S_{is1} = \sum_{t_k \leqslant L_i} \lambda_k \exp\{{x_s}^T\beta_1 + {Z_i}^T\beta_2\}$, $S_{is2} = \sum_{t_k \leqslant R_i} \lambda_k \exp\{{x_s}^T\beta_1 + {Z_i}^T\beta_2\}$, and $f'(x) = df(x)/dx$ for any function f. In particular, if $f(\xi_i)$ is the gamma density function with the parameter γ, we have

$$G'(x) = \frac{\int_{\xi_i} \xi_i \exp(-x\xi_i)f(\xi_i)d\xi_i}{\exp\{-G(x)\}} = \frac{(\gamma x + 1)^{-\gamma^{-1}-1}}{\exp\{-G(x)\}}\,.$$

Also to calculate the following type integration

$$\int_{\xi_i} \xi_i\{\exp(-\xi_i S_{i1}) - \exp(-\xi_i S_{i2})\}[1-\exp(-\xi_i(S_{i2}-S_{i1}))]^{-1} f(\xi_i)d\xi_i$$

that has no closed-form, we suggest to employ the Gauss–Laguerre quadrature technique.

In the M-step, we need to maximize $Q(\theta|\mathbf{O}_i, \theta^{(d)})$ with respect to β_1, β_2, the λ_k's and the p_{sj}'s. To update the $\{p_{sj};\ j = 1, \ldots, n;\ s = 1, \ldots, M\}$, one requires to maximize $\sum_{i\in C}\sum_{s=1}^{M} I(X_i = x_s)w_{ji}\log p_{sj} + \sum_{i\in\bar{C}}\sum_{s=1}^{M}\hat{\psi}_{sji}\log p_{sj}$ under the constraint $\sum_{s=1}^{M} p_{sj} = 1$, which gives

$$p_{sj} = \frac{\sum_{i\in C} I(X_i = x_s)w_{ji} + \sum_{i\in\bar{C}}\hat{\psi}_{sji}}{\sum_{s=1}^{M}\{\sum_{i\in C} I(X_i = x_s)w_{ji} + \sum_{i\in\bar{C}}\hat{\psi}_{sji}\}}\,. \tag{6}$$

Also for updating the λ_k's, one can easily obtain the following closed-form expression

$$\lambda_k = \frac{\sum_{i=1}^{n} I(t_k \leqslant {R_i}^*)E_X\{E(W_{ik})\}}{\sum_{i=1}^{n} I(t_k \leqslant {R_i}^*)E_X\{E(\xi_i)\exp({X_i}^T\beta_1)\}\exp({Z_i}^T\beta_2)}\,. \tag{7}$$

To obtain the updated estimators of β_1 and β_2, the one-step Newton–Raphson method can be used based on the following equations

$$\frac{\partial Q}{\partial\beta_1} = \sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \leqslant {R_i}^*)\left\{ E_X\{E(W_{ik})X_i\} - E_X\{E(W_{ik})\}\frac{\sum_{i=1}^{n} I(t_k \leqslant {R_i}^*)E_X\{E(\xi_i)\exp({X_i}^T\beta_1)X_i\}\exp({Z_i}^T\beta_2)}{\sum_{i=1}^{n} I(t_k \leqslant {R_i}^*)E_X\{E(\xi_i)\exp({X_i}^T\beta_1)\}\exp({Z_i}^T\beta_2)} \right\} = 0\,, \tag{8}$$

and

$$\frac{\partial Q}{\partial \beta_2}=\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \leqslant R_i{}^*)E_X\{E(W_{ik})\} \times \left\{ Z_i - \frac{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)Z_i}{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)} \right\}=0\,. \tag{9}$$

For the implementation of the Newton–Raphson method, we need the following functions

$$\frac{\partial^2 Q}{\partial \beta_1^2}=\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \leqslant R_i{}^*)E_X\{E(W_{ik})\} \left\{ \left(\frac{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)X_i\}\exp(Z_i{}^T\beta_2)}{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)}\right)^2 - \frac{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)X_i^TX_i\}\exp(Z_i{}^T\beta_2)}{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)} \right\},$$

$$\frac{\partial^2 Q}{\partial \beta_2^2}=\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \leqslant R_i{}^*)E_X\{E(W_{ik})\} \left\{ \left(\frac{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)Z_i}{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)}\right)^2 - \frac{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)Z_i^TZ_i}{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)} \right\},$$

and

$$\frac{\partial^2 Q}{\partial \beta_1 \partial \beta_2}=\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k \leqslant R_i{}^*)E_X\{E(W_{ik})\} \left\{ \frac{(\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)X_i\}\exp(Z_i{}^T\beta_2))}{(\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2))} \times \frac{(\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)Z_i)}{(\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2))} - \frac{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)X_i\}\exp(Z_i{}^T\beta_2)Z_i}{\sum_{i=1}^n I(t_k \leqslant R_i{}^*)E_X\{E(\xi_i)\exp(X_i{}^T\beta_1)\}\exp(Z_i{}^T\beta_2)} \right\}.$$

In summary, the EM algorithm described above can be summarized as follows.

Step 1. Choose an initial estimate $\theta^{(0)}$.

Step 2. At the $(d+1)$th iteration, first calculate the quantities $\hat{q}_{is}$ and $\hat{\psi}_{sji}$ given in (4) and (5), respectively, and then the conditional expectations $E_X\{E(W_{ik})X_i\}$, $E_X\{E(W_{ik})\}$, $E(\xi_i)$, $E_X\{E(\xi_i)\exp(X_i^T\beta_1)\}$, $E_X\{E(\xi_i)\exp(X_i^T\beta_1)X_i\}$ and $E_X\{E(\xi_i)\exp(X_i^T\beta_1)X_i^T X_i\}$.

Step 3. Obtain the updated estimates $\beta_1^{(d+1)}$ and $\beta_2^{(d+1)}$ by solving the score equations (8) and (9) with the use of the one-step Newton–Raphson method.

Step 4. Obtain the updated estimate $\lambda_k^{(d+1)}$ from (7).

Step 5. Obtain the updated estimate $p_{sj}^{(d+1)}$ from (6).

Step 6. Repeat Steps 2–5 until the convergence is achieved.

For the implementation of the EM algorithm above, it is apparent that one needs to choose the initial estimates and determine the convergence criterion. For the former, we suggest to take some reasonable values as we tried different values in the numerical study below and the algorithm does not seem to be sensitive to the initial values. For the latter, it is clear that many criteria can be used and in the numerical study below, we use the summation of the absolute differences of the consecutive estimates being less than or equal to a constant, taken to be 0.001. Also for the implementation, one has to choose the bootstrap sample size for variance estimation and based on the experience from the numerical study below, the size 100 seems to be large enough to give stable results. For a specific example, of course, one could also try different sizes and compare the results. Also although the EM algorithm may not be fast partly due to the use of the bootstrap procedure and the need of estimation of α, the numerical study did not seem to have any convergence or other issues.

5. A simulation study

Now we present some results obtained from a simulation study conducted to assess the performance of the estimation procedure proposed in the previous sections with the focus on estimation of β. In the study, for covariates, we considered two situations and in both cases, we assumed that the covariates Z_i's follow the Bernoulli distribution with the success probability 0.5. In the first situation, we generated the covariates X_i's from the Bernoulli distribution with the success probability $\exp(1-Z_i)/(1+\exp(1-Z_i))$, while in the second situation, the X_i's were generated from the normal distribution with the mean Z_i and variance 1. Given the covariates, the failure times T_i's were generated from model (1) with $\Lambda(t) = \log(1 + 0.5t)$ and $G(x) = \log(1 + \gamma x)/\gamma$ for different values of γ. Note that as mentioned above, it gives the proportional hazards model with $\gamma = 0$ and $\gamma = 1$ corresponds to the proportional odds model.

Table 1
Simulation results with discrete missing covariates.

Model	Method	Bias	ESE	SSE	CP
$\gamma = 0$	*Proposed method*	0.0383 0.0704	0.3370 0.2598	0.3228 0.2432	0.9620 0.9690
	Complete case	0.0495 0.1139	0.3567 0.3665	0.3400 0.3276	0.9720 0.9780
	Full data	0.0184 0.0550	0.2126 0.2135	0.2182 0.2104	0.9540 0.9470
$\gamma = 0.5$	*Proposed method*	0.0194 0.0416	0.3604 0.2516	0.3842 0.2861	0.9440 0.9140
	Complete case	0.0225 0.0613	0.4325 0.4072	0.3980 0.3951	0.9700 0.9650
	Full data	0.0116 0.0311	0.2679 0.2633	0.2593 0.2626	0.9550 0.9570
$\gamma = 1$	*Proposed method*	0.0322 0.0472	0.4329 0.3011	0.4383 0.3260	0.9460 0.9310
	Complete case	0.0519 0.0723	0.4995 0.4843	0.4556 0.4507	0.9750 0.9680
	Full data	0.0189 0.0370	0.3208 0.3136	0.3132 0.3040	0.9550 0.9570

For the generation of the censoring intervals defined by the U_i's and V_i's, we assumed that the U_i's follow the uniform distribution over $(0, 3\tau/4)$ and took $V_i = \min\{\, 0.1 + U_i + \tau a_i/2, \tau \,\}$, where the a_i's were generated from the exponential distribution with mean 1. Also here the constant τ was chosen based on the required censoring percentages or proportions of δ_{1i}, δ_{2i} and δ_{3i} being one. We used the Gaussian kernel with $h = 1.5\hat{\sigma}_{w_1}\, n^{-1/3}$, where $\hat{\sigma}_{w_1}$ denotes the sample standard deviation of the continuous part of the variable W_1. Finally we generated the covariate missing indicators r_i's from the Bernoulli distributions with the probability

$$\pi_i(\alpha) = \frac{1}{1 + \exp(\alpha_0 + \alpha_1 Z_i + \alpha_2 X_i)} .$$

The results given below are based on $n = 200$ or 300 with 1000 replications.

Tables 1 and 2 present the simulation results obtained on estimation of β_1 and β_2 of under the two covariate situations, respectively, with $\gamma = 0$, 0.5 and 1, $\beta_1 = 0.5$, $\beta_2 = 1$ and $(\alpha_0, \alpha_1, \alpha_2) = (0.3, -0.3, -0.2)$. Here the percentages of left-, interval- and right-censored observations are about 50%, 20% and 30%, respectively. The results include the estimated bias (Bias) given by the average of the estimates minus the true value, the sample standard error of the estimates (SSE), the average of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP). Here in addition to the proposed estimation procedure, we also applied the naive complete case approach that performed the analysis based only on the subjects with complete covariate information, and the full data approach that assumed no missing covariates for comparison. One can see that the proposed estimation procedure seems to work well overall with the performance between the complete data method and the full data method in all aspects.

Note that under the set-ups considered above, the missing rate (MR) for covariates is 50%. To see its possible effects on the estimation, we repeated the study giving the results in Table 1 with $\gamma = 0$ except $(\alpha_0, \alpha_1, \alpha_2) = (-1, -0.1, 1)$ or $(0.4, -0.1, 0.8)$ and present the simulation results in Table 3. They correspond to the MR being 40% and 70%, respectively, and it seems that they gave the same conclusions as above. To see the convergence and the effect of sample sizes on the estimation of regression parameters, Table 4 gives the estimation results obtained by repeating the study giving the results in Table 1 with $\gamma = 0$ except $n = 300$. As expected, the results became better and again indicate that the proposed estimation procedure works well. We also considered other set-ups, including different observation processes and more and different covariates as well as the use of different kernel functions, and obtained similar results. In particular, the algorithm and obtained results became stable if $n \geq 200$.

6. Analysis of Alzheimer's disease neuroimaging initiative study

In this section, we apply the approach proposed in the previous sections to the Alzheimer's Disease Neuroimaging Initiative (ADNI) study described above. As mentioned before, it is a longitudinal study designed to collect the information on various clinical, imaging and genetic factors as well as biochemical biomarkers that may affect and help the early detection and tracking of the Alzheimer's disease (AD). As most of longitudinal studies, the information collection happens from time to time and depends on the participant's follow-ups. Thus it is natural to exist some missing covariates and also the observations on most of time-to-events suffer interval censoring. In the study, the participants are classified at the baseline into three groups based on their cognitive conditions, cognitively normal, mild cognitive impairment and

Table 2
Simulation results with continuous missing covariates.

Model	Method	Bias	ESE	SSE	CP
$\gamma = 0$	*Proposed method*	0.0462 0.0272	0.1730 0.2597	0.1658 0.2410	0.9440 0.9630
	Complete case	0.0599 0.0782	0.1920 0.3876	0.1809 0.3479	0.9580 0.9680
	Full data	0.0277 0.0288	0.1181 0.2274	0.1178 0.2166	0.9430 0.9630
$\gamma = 0.5$	*Proposed method*	0.0225 0.0271	0.2033 0.3180	0.1977 0.3168	0.9567 0.9478
	Complete case	0.0313 0.0471	0.2207 0.4478	0.2072 0.4223	0.9544 0.9589
	Full data	0.0143 0.0206	0.1409 0.2833	0.1317 0.2868	0.9567 0.9489
$\gamma = 1$	*Proposed method*	0.0216 0.0442	0.2394 0.3801	0.2436 0.3738	0.9477 0.9537
	Complete case	0.0342 0.0772	0.2828 0.5750	0.2575 0.5143	0.9517 0.9658
	Full data	0.0228 0.0251	0.1637 0.3361	0.1533 0.3224	0.9477 0.9618

Table 3
Simulation results with different missing rates.

Model	Method	Bias	SEE	SSE	CP
MR = 0.4	*Proposed method*	0.0129 0.0437	0.2749 0.2324	0.2628 0.2240	0.9670 0.9620
	Complete case	0.0379 0.0630	0.2913 0.3006	0.2782 0.2871	0.9630 0.9660
	Full data	0.0219 0.0417	0.2121 0.2114	0.2052 0.2064	0.9600 0.9570
MR = 0.7	*Proposed method*	0.0066 0.0605	0.4321 0.2954	0.3806 0.2614	0.9680 0.9750
	Complete case	0.0666 0.1742	0.5036 0.5421	0.4330 0.4526	0.9750 0.9790
	Full data	0.0186 0.0499	0.2109 0.2103	0.1978 0.2083	0.9680 0.9500

Table 4
Simulation results with $n = 300$.

Method	Bias	SEE	SSE	CP
Proposed method	0.0259 0.0252	0.2490 0.1898	0.2407 0.1820	0.9540 0.9600
Complete case	0.0323 0.0573	0.2606 0.2609	0.2513 0.2445	0.9570 0.9640
Full data	0.0193 0.0217	0.1664 0.1648	0.1667 0.1637	0.9540 0.9490

Alzheimer's disease, and one failure event of interest is the AD conversion. As described above, due to the nature of the study and instead of being known exactly, the occurrence time of the conversion is only known to be between the last observation time when AD had not occurred and the first observation time when the AD had already occurred.

In the analysis below, by following Li et al. (2017), we will consider 371 participants in the mild cognitive impairment group with the focus on the time from the baseline visit date to the AD conversion and the association between the time and eight covariates. They are the two AD assessment scale test results (ADAS11, ADAS13), middle temporal gyrus volume (Midtemp), Rey auditoryverbal learning test score of immediate recall (RAVLT.i), and the functional assessment questionnaire score (FAQ) along with three baseline covariates Age, Gender and years of education (EDU). Note that Li et al. (2017) identified the covariates ADAS11, ADAS13, Midtemp, RAVLT.i and FAQ as the most important clinical and demographic factors associated with the AD conversion based on the individual variable analysis. Among them, about 20% values of the Midtemp are missing and there are no missing for other covariates. Also there seems to exist some correlations between all eight covariates.

Table 5
The estimated covariate effects on the AD conversion with $\gamma = 0$.

Covariate	Proposed			CC		
	Estimate	ESE	p-value	Esimate	ESE	p-value
Midtemp	−0.5882	0.1546	0.0001	−0.6248	0.1185	0.0000
ADAS11	0.0032	0.2920	0.9914	0.3389	0.3182	0.2868
ADAS13	0.5982	0.2982	0.0448	0.3213	0.3268	0.3255
RAVLT.i	−0.8406	0.2227	0.0002	−0.7378	0.2403	0.0021
FAQ	0.6382	0.1924	0.0009	0.5812	0.1949	0.0029
Gender	−0.1011	0.2360	0.6682	−0.0132	0.2366	0.9556
Age	−0.2468	0.2006	0.2185	−0.3914	0.2016	0.0522
EDU	−0.0625	0.1889	0.7408	−0.1119	0.1936	0.5633

Table 6
The estimated covariate effects on the AD conversion with $\gamma = 1$.

Covariate	Proposed			CC		
	Estimate	SEE	pvalue	Esimate	SEE	p-value
Midtemp	−0.6771	0.2210	0.0022	−0.7832	0.1597	0.0000
ADAS11	0.1192	0.3755	0.7509	0.5212	0.3728	0.1621
ADAS13	0.7090	0.3698	0.0552	0.4521	0.3767	0.2301
RAVLT.i	−1.1810	0.3086	0.0001	−1.0916	0.3253	0.0008
FAQ	0.9296	0.2739	0.0007	0.8604	0.2873	0.0027
Gender	−0.1276	0.3228	0.6927	0.0384	0.3374	0.9094
AGE	−0.4190	0.2780	0.1318	−0.7153	0.2957	0.0156
EDU	−0.0248	0.2745	0.9280	−0.0730	0.2785	0.7931

Tables 5 and 6 show the estimated covariate effects given by the proposed estimation procedure with the use of the same G function employed in the previous section and $\gamma = 0$ and 1, respectively, and they include the proposed estimates $\hat{\beta}_n$, the estimated standard errors and the p-values for testing each regression parameter being zero. Note that the results with $\gamma = 1$ were chosen since it gave the smallest AIC value, while the results with $\gamma = 0$ are provided for comparison and the fact that $\gamma = 0$ corresponds to the Cox model. Here as in the simulation study, we used the Gaussian kernel with the bandwidth $h = 1.5\hat{\sigma}_{w_1} n^{-1/3}$. In addition, we also applied the complete case approach investigated in the previous section and include the obtained estimation results in the tables for comparison.

One can see from the two tables that both proposed and complete data methods gave similar results on the six covariates Gender, EDU, ADS11, Midtemp, RAVLT.i and FAQ and suggested that among them, only Midtemp, RAVLT.i and FAQ were significant predictors for the AD conversion. Note that Midtemp, RAVLT.i and FAQ represent some measurements on the neuroimaging, neuripsychological, and functional and behavioural assessments, respectively, and the results suggest that the AD conversion was negatively related to Midtemp and RAVLT.i but positively related to FAQ. On the covariate Age, the proposed method indicates that it did not seem to have any effect on the AD conversion, while the complete case approach suggests some correlation and may overestimate the effect. On the covariate ADAS13, another measurement on the neuropsychological assessment, the proposed approach suggests that it seems to be positively related to or a significant predictor for the AD conversion but the complete case approach indicates otherwise.

Note that in contrast to the conclusions above, the individual analyses indicated that both ADAS11, also a measurement on the neuropsychological assessment, and ADAS13 were strong predictors for the AD conversion (Li et al., 2017). A possible reason for this is the strong correlation between ADAS11 and ADAS13. To provide a graphical view about the analysis results, Fig. 1 displays the estimated survival function given by the proposed estimation procedure corresponding to the subject with RAVLT.i less (the bottom curve) or greater (the top curve) than the observed mean value of the covariate and other covariates being zero, respectively. Here we took the same G function with $\gamma = 0$ as above and the figure again suggests that RAVLT.i was significantly negatively correlated with the AD conversion.

7. Discussion and concluding remarks

This paper discussed regression analysis of interval-censored failure time data, a general type of censored data, arising from a class of semiparametric linear transformation models when there exist nonignorable missing covariates. For inference, a two-step estimating procedure was proposed and the proposed estimators of regression parameters were shown to be consistent and asymptotically follow the normal distribution. As mentioned above, although there exists some literature on the type of problems discussed here, there was no established estimation procedure for the situation considered here and the proposed approach fills a hole in the missing data literature. Furthermore, the numerical studies suggested that the proposed method seems to work well for practical situations and it should be used in the presence of nonignorable missing covariates.

Note that although the focus above has been on estimation of regression parameters β_1 and β_2, the proposed estimation procedure also gives an estimator of the baseline cumulative hazard function $\Lambda(t)$. Although we cannot establish the

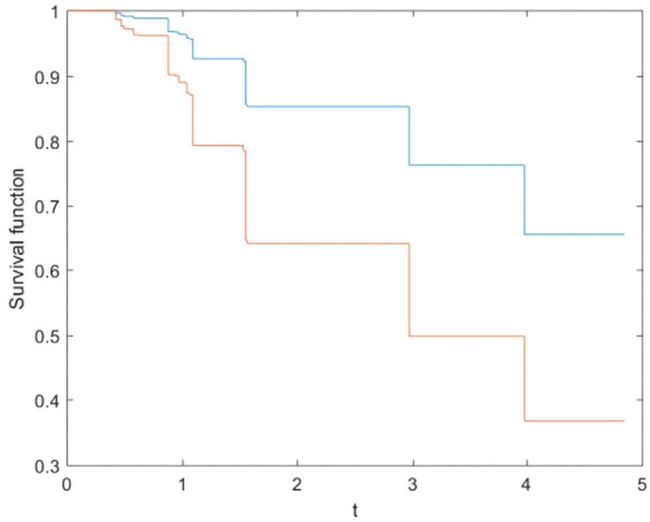

Fig. 1. The estimated survival functions for the subjects with the covariate RAVLT.i less (the bottom curve) or greater (the top curve) than the observed mean value.

asymptotic properties of the resulting estimator, the numerical study indicated that the estimator seems to be consistent and perform well. Also note that for the estimation procedure proposed above, we have assumed that the missing mechanism for covariates can be described by a known distribution up to a vector of unknown parameters and this applies to many situations. However, this would not be valid if missing covariates depend on the failure time of interest and one such example is the case-cohort study.

It is worth to note that the estimation procedure proposed in the previous sections relies on several assumptions and made use of some techniques or tools. One important assumption is the independent or non-informative interval censoring, which means that one can perform the analysis or estimation conditional on the observation process. Sometimes this may not be true, or the follow-up schedule or the observation process may be related to the failure variable of interest even given covariates (Sun, 2006). That is, we have informative interval censoring, and as many authors pointed out, there is no method available to check this in general unless there exists some extra information. Also as pointed out in the literature, in the presence of informative censoring, the use of the methods that assume the independent censoring would yield biased or misleading results or conclusions. In other words, one needs to develop a different or more general estimation procedure. In the development above, we have focused on model (1) and although it is quite general and flexible, sometimes one may still prefer to check its appropriateness. However, it seems that even for relatively simpler right-censored data, it does not seem to exist an established procedure specifically developed for it.

Of course, another assumption made in the proposed method is about the missing mechanism discussed above and it is apparent that it would be useful to relax this assumption. On the other hand, as discussed by many authors, this is a difficult question as the observed information may not be enough to allow the estimation or checking of a general assumption or model on the nonignorable missing mechanism. In the development of the proposed method, we introduced the latent variable ξ with the density function $f(\xi)$ and the Poisson random variable W. For the former, although we have focused on the case of $f(\xi)$ being the gamma density function, the proposed method is still valid for other density function, while for the latter, an EM algorithm can still be developed for the determination of the proposed estimators but may be different.

Acknowledgements

The authors wish to thank the Co-Editor, Dr. Park, the Associate Editor and two reviewers for their many helpful and thoughtful comments and suggestions that greatly improved the paper. The research was partially supported by grants from the Natural Science Foundation of China [Grant Number 11731011], a grant from key project of the Yunnan Province Foundation, China [Grant Number 202001BB050049] and a grant from reserve talents for academic and technical leaders of middle-aged and young people in Yunnan Province, China [Grant Number 2017HB004]. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Appendix A. Proofs of the asymptotic properties

In this Appendix, we will sketch the proof for the consistency and asymptotic normality of the proposed estimators described above by employing the empirical process theory and nonparametric techniques. Define $Pf = \int f(x)dP(x)$, and $P_nf = n^{-1}\sum_{i=1}^{n} f(X_i)$ for a function f, a probability function P and a sample $X_1, \ldots, X_n$. For the proof, we need the following regularity conditions.

(A1) Assume that $\Lambda(\tau_1) < \infty$, $\Lambda(\tau_2) < \infty$, and there exists a positive constant a such that $P(V - U > a) > 0$. Also the union of the supports of U and V is contained in the interval $[r_1, r_2]$ with $0 < r_1 < r_2 < +\infty$.

(A2) The transformation function G is twice continuously differentiable on $[0, \infty)$ with $G(0) = 0, G'(x) > 0$ and $G(\infty) = \infty$.

(A3) The set of covariates (X, Z) has bounded support.

(A4) The bandwidth h satisfies that $h \to 0$, $nh^d \to \infty$ and $nh^{2m} \to 0$.

(A5) The function $\Lambda_0 \in M$ is continuously differentiable up to order r in $[r_1, r_2]$, with the first derivative being strictly positive, and satisfies $\alpha^{-1} < \Lambda_0(r_1) < \Lambda_0(r_2) < \alpha$ for some positive constant α.

(A6) If $g(t) + X_i^T\beta_1 + Z_i^T\beta_2 = 0$ for all $t \in [r_1, r_2]$ with probability 1, then $g(t) = 0$ for $t \in [r_1, r_2]$, $\beta_1 = 0$ and $\beta_2 = 0$.

First we will prove the consistency and for this, we will verify the conditions of Theorem 5.7 of Van Der Vaart (1998). Let $BV_\omega[r_1, r_2]$ denote the functions whose total variation in $[r_1, r_2]$ are bounded by a given constant. Then the class of functions

$$F_\omega = \left\{\int_0^{U_k} \exp\{X_i^T\beta_1 + Z_i^T\beta_2\} d\Lambda(s) : \Lambda \in BV_\omega[r_1, r_2]\right\}$$

is a convex hull of functions $\{I(U_k \geqslant s)\exp\{X_i^T\beta_1 + Z_i^T\beta_2\}$, so it is a Donsker class. Furthermore,

$$\exp\left(-G\left[\int_0^{U_k} \exp\{X_i^T\beta_1 + Z_i^T\} d\Lambda(s)\right]\right) - \exp\left(-G\left[\int_0^{U_{k+1}} \exp\{X_i^T\beta_1 + Z_i^T\} d\Lambda(s)\right]\right)$$

is bounded away from zero. Therefore, $l(\theta, \hat{\alpha}|\mathbf{O}) = \log L(\theta, \hat{\alpha}|\mathbf{O})$ belongs to some Donsker class due to the preservation property of the Donsker class under Lipschitz-continuous transformations. Then we can conclude that $\sup_{\theta\in\Theta_n}|P_nl(\theta, \hat{\alpha}|\mathbf{O}) - P_nl(\theta_0, \hat{\alpha}|\mathbf{O})|$ converges in probability to 0 as $n \to 0$.

Now we verify that another condition of Theorem 5.7 of Van Der Vaart (1998) also holds. That is, for any $\varepsilon > 0$, we have

$$\sup_{d(\theta,\theta_0)>\varepsilon} Pl(\theta, \hat{\alpha}|\mathbf{O}) < Pl(\theta_0, \hat{\alpha}|\mathbf{O}).$$

Note that this condition is satisfied if we can prove the model is identifiable. According to condition (A6) and similar arguments to the proof of Theorem 2.1 of Chang et al. (2007), we can show the identifiability of the model parameters. Now, by Theorem 5.7 of Van Der Vaart (1998), we have $d(\hat{\theta}_n, \theta_0) = o_p(1)$, which completes the proof of consistency.

Before proving the asymptotic normality, we will need to establish the convergence rate. For this, we will first define the covering number of the class $\mathcal{L} = \{l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta\}$ and establish a needed lemma.

Lemma 1. *Assume that Conditions (A1), (A3)–(A6) hold. Then the covering number of the class* $\mathcal{L} = \{l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta\}$ *satisfies*

$$N(\epsilon, \mathcal{L}, L_2(P)) = O(\epsilon^{-1}).$$

Proof of Lemma 1. The proof is similar to that of Zeng et al. (2016) and Hu et al. (2017) and thus omitted.

To establish the convergence rate, for any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) : \theta \in \Theta, d(\theta, \theta_{n0}) \leqslant \eta\}$ with $\theta_{n0} = (\beta_0, \Lambda_{n0})$. Following the calculation of Shen and Wong (1994, P.597), we can establish that $\log N_{[]}(\epsilon, \mathcal{F}_\eta, \|\,.\,\|_2) \leqslant CN \log(\eta/\epsilon)$ with $N = m + 1$, where $N_{[]}(\epsilon, \mathcal{F}_\eta, d)$ denotes the bracketing number (see the Definition 2.1.6 in Van der Vaart and Wellner, 1996 with respect to the metric or semi-metric d of a function class $\mathcal{F}$. Moreover, some algebraic calculations lead to $\| l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) \|_2^2 \leqslant C\eta^2$ for any $l(\theta_{n0}, \hat{\alpha}|\mathbf{O}) - l(\theta, \hat{\alpha}|\mathbf{O}) \in \mathcal{F}_\eta$. Therefore, by Lemma 3.4.2 of Van der Vaart and Wellner (1996), we obtain

$$E_p \| n^{1/2}(P_n - P) \|_{\mathcal{F}_\eta} \leqslant CJ_\eta(\epsilon, \mathcal{F}_\eta, \|\,.\,\|_2)\{1 + \frac{J_\eta(\epsilon, \mathcal{F}_\eta, \|\,.\,\|_2)}{\eta^2 n^{1/2}}\}, \tag{S}$$

where $J_{[]}(\eta, \mathcal{F}_\eta, \|\,.\,\|_2) = \int_0^\eta \{logN_{[]}(\epsilon, \mathcal{F}_\eta, \|\,.\,\|_2)\}^{1/2} d\epsilon$. The right-hand side of (S) yields $\phi_n(\eta) = C\eta^{1/2}(1 + \frac{\eta^{1/2}}{\eta^2 n^{1/2}} M_1)$, where M_1 is a positive constant. Then $\phi_n(\eta)/\eta$ is a decreasing function, and $n^{2/3}\phi_n(-1/3) = O(n^{1/2})$. According the theorem 3.4.1 of Van der Vaart and Wellner (1996), we can conclude that $d(\hat{\theta}, \theta_0) = O_p(n^{-1/3})$.

Now we prove the asymptotic normality of $\hat{\beta}_n$. Following the proof of Theorem 2 in Zeng et al. (2016), one can obtain that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = (E\{l_\beta - l_\Lambda(s^*)\}\{l_\beta - l_\Lambda(s^*)\}^T)^{-1} G_n\{l_\beta - l_\Lambda(s^*)\} + o_p(1),$$

where l_β is the score function for β, $l_\Lambda(s^*)$ is the score function along this submodel $d\Lambda_{\epsilon,s^*} = (1+\epsilon s^*)d\Lambda$. This implies that the influence function for $\hat{\beta}_n$ is exactly the efficient influence function, so that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound (Bickel et al., 1993, p. 65).

References

Bickel, P., Klassen, C., Ritov, Y., Wellner, J., 1993. Efficient and Adaptive Estimation for Semiparametric Models. John Hopkins University Press, Baltimore.

Chang, I.S., Wen, C.C., Wu, Y.J., 2007. A profile likelihood theory for the correlated gamma-frailty model with current status family data. Statist. Sinica 17, 1023–1046.

Chen, K., Jin, Z., Ying, Z.L., 2002. Semiparametric analysis of transformation model with censored data. Biometrika 82, 659–668.

Cook, V.J., Hu, x.J., Swartz, T.B., 2011. Cox regression with covariates missing not at random. Stat. Biosci. 3, 208–222.

Efron, B., 1981. Censored data and the bootstrap. J. Amer. Statist. Assoc. 76, 312–319.

Finkelstein, D.M., 1986. A proportional hazards model for interval-censored failure time data. Biometrics 42, 845–854.

Graham, J.W., 2012. Missing Data. Springer, New York.

Hu, N., Chen, X.R., Sun, J.G., 2015. Regression analysis of length-biased and right-censored failure time data with missing covariates. Scand. J. Stat. 42, 438–452.

Hu, T., Zhou, Q., Sun, J., 2017. Regression analysis of bivariate current status data under the proportional hazards model. Canad. J. Statist. 45, 410–424.

Kalbfleisch, J.D., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data, second ed. Wiley, New York.

Kim, J.K., Shao, J., 2013. Statistical Methods for Handling Incomplete Data. CRC Press, Boco Raton.

Li, K., Chan, W., Doody, R.S., Quinn, J., Luo, S., 2017. Prediction of conversion to alzheimer's disease with longitudinal measures and time-to-event data. J. Alzheimers Dis. 58, 361–371.

Lipsitz, S.R., Ibrahim, J.G., Chen, M.H., Peterson, H., 1999. Nonignorable missing covariates in generalized linear models. Stat. Med. 18, 2435–2448.

Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data, second ed. Wiley, New York.

Louis, T.A., 1982. Finding the observed information matrix when using the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 44, 226–233.

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

Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Tsiatis, A., Verbeke, G. (Eds.), 2015. Handbook of Missing Data Methodology. Chapman & Hall.

Morikawa, K., Kim, J.K., Knao, Y., 2017. Semiparametric maximum likelihood estimation with data missing not at random. Canad. J. Statist. 45, 393–409.

Ning, Y., Yi, G., Reid, N., 2018. A class of weighted estimating equations for semiparametric transformation models with missing covariates. Scand. J. Stat. 45, 87–109.

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

Shen, X., Wong, W.H., 1994. Convergence rate of sieve estimates. Ann. Statist. 22, 580–615.

Su, Y.R., Wang, J.L., 2016. Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data. Ann. Statist. 44, 1298–1331.

Sun, J., 2006. The Statistical Analysis of Interval-Censored Failure Time Data. Springer, New York.

Van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes: With Applications to Statistics. Springer-Verlag, New York.

Van Der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press.

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

Wang, L., McMahan, C.S., Hudgens, M.G., Qureshi, Z.P., 2016. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. Biometrics 72, 222–231.

Wen, C.C., Chen, Y.H., 2011. Nonparametric maximum likelihood analysis of clustered current status data with the gamma frailty Cox model. Comput. Statist. Data Anal. 55, 1053–1060.

Wullfsohn, M.S., Tsiatis, A.A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics 53, 330–339.

Xu, Q., Paik, M.C., Luo, X., Tsai, W.Y., 2009. Reweighting estimators for Cox regression with missing covariates. J. Amer. Statist. Assoc. 104, 1155–1167.

Zeng, D., Mao, L., Lin, D.Y., 2016. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. Biometrika 103, 253–271.

Zhang, Z., Zhao, Y., 2013. Empirical likelihood for linear transformation models with interval-censored failure time data. J. Multivariate Anal. 116, 398–409.

Zhao, P.Y., Tang, N.S., Qu, A., Jiang, D.P., 2017. Semiparametric Estimating Equations Inference with Nonignorable Missing Data. Statist. Sinica 27, 89–113.