# Regression analysis of arbitrarily censored data under the proportional odds models

Lu Wang and Lianming Wang[*]

*Department of Statistics, University of South Carolina, 1523 Greene Street, Columbia,*

*South Carolina, 29208, U.S.A.*

[*]*Email: wang99@mailbox.sc.edu*

**Abstract**: Arbitrarily censored data refer to the survival data that contain a mixture of exactly observed, left-censored, interval-censored, and right-censored observations. Existing research work on regression analysis on arbitrarily censored data is sparse and limited to the proportional hazards model only. In this article, a novel estimation approach based on an EM algorithm is proposed for analyzing such data under the proportional odds model. The proposed EM algorithm has many appealing properties such as being robust to initial values, easy to implement, converging fast, and providing the variance estimate of the regression parameter estimate in closed form. Our method has shown excellent performance in estimating the regression parameters as well as the the baseline survival function in an extensive simulation study. Several real-life data applications are provided for illustration.

**Keywords**: Arbitrarily censored data; EM algorithm; Proportional odds model; Data augmentation; Semiparametric regression; Monotone spline.

# 1 Introduction

The analysis of survival data plays an indispensable role in a lot of areas, such as epidemiology, biomedical science, engineering and sociology. Nowadays, there are more and more instances where complex observation schemes have to be treated. Interval-censored data occurs naturally in clinical trials and epidemiology studies in which patients visit the clinic periodically and the event of interest is assessed on repeated visits. It also appears in retrospective cohort study when some of the event time are exactly observed but the others are only known to lie in certain time intervals. Current status data can be encountered when study the onset of tumor on mice. Since researchers can only examine whether a tumor has developed after a rat is sacrificed so the time is either right- or left-censored. Also, prevalent cases of a disease can be viewed as left-censored observations, etc. The diverse situation leads to a demand for versatile approaches which can accommodate arbitrarily censored data produced by complex observation schemes.

A series of nonparametric maximum likelihood estimators (NPMLE) with no covariates have been developed for general censoring data. Peto (1973) proposed a nonparametric maximum likelihood estimator (NPMLE) of the survival function for interval-censored data. Turnbull (1974) extended it to arbitrarily censored data and further extended it to fit arbitrarily censored and truncated data in 1976. Pan & Chappell (1998) suggested an iterative Nelson estimator (INE) to estimate the survival function nonparametrically. Other methods includes Wang,Jewell, and Tsai(1986),and Wang(1992),among others. Meanwhile, since regression analysis of censored time-to-event data is of central interest in health sciences research, some widely used approaches have been developed based on semiparametric models.

For the most widely used proportional hazard regression model, Finkelstein (1986) developed a maximum likelihood estimator for arbitrarily censored data. Tu et al. (1993) described a general discrete-time proportional hazards model and fitted it with EM algorithm. Alioum & Commenges (1996) extended Turnbull (1976)'s method to the proportional hazards model for arbitrarily censored and truncated data in continuous time.

Compared to the large amount of methods established for the PH model, the methodology development with respect to the proportional odds (PO) model is very limited. PO model performs as an alternative to PH model on analyzing time to event data. It specifies the log ratio of odds of survival given covariates to the baseline odds as a parametric regression function of covariates. The associated baseline odds function is left unspecified. Regression parameters in PO model is more interpretable than PH model in terms of odds ratio. Different from proportional hazard model, PO model constrains the ratio of the hazards converges to unity as time increases, so Bennett (1983a) and Murphy et al. (1997) suggest that PO model is more appropriate for demonstrating an effective cure or the case that the morbidity rates converge with time. Despite its pleasing interpretation, the PO model is rarely used, likely due to difficulty of implementation. Efforts have been done on addressing simpler instances, say right censored data. For example, reasonable estimations for the regression coefficients have been proposed by Bennett (1983a), Murphy et al. (1997), Yang & Prentice (1999) and Royston & Parmar (2002), among others. Because the complexity of data structure adds more complexity to the PO model there are only a handful of studies on fitting PO model with interval-censored data. Rossini & Tsiatis (1996) adapted the semi-parametric framework for modeling current status data by approximating the infinite-dimensional nuisance parameter, the baseline log-odds of failure, with a step function, and

carried out a maximum likelihood procedure. Huang & Rossini (1997) proposed a sieve maximum likelihood estimator for proportional odds model with interval censored data. Shen (1998) use monotone splines of variable orders and knots for approximating the odds of failure time and proposed a sieve maximum likelihood estimator for right-censored and case 2 interval-censored data. Lin & Wang (2011) proposed a Bayesian approach for analyzing case 2 interval-censored data under the semiparametric proportional odds model. This situation requires the analyst to seek specialized, distinct techniques according to different censoring patterns.

The aim of this paper is to propose an easy to implemented approach that can fit the proportional odds model for arbitrarily censored data in continuous time. The expectation-maximization (EM) algorithm we propose is so flexible that it can handle any combination of incomplete data. Specifically, it can successfully fit randomly right-censored data, left-censored data, current status data, case 2 interval-censored data, or a mixture of them. To our knowledge, icenReg(Anderson-Bergman 2017) is the only approach available for analyzing arbitrarily censored data under the PO model by now. This method is a combination of conditional Newton-Raphson, ICM algorithm (Pan 1999) and constrained gradient ascent algorithm. It is an efficient algorithm but it does not offer closed-form of standard errors. Inference on the regression parameters needs to be done using bootstrap standard errors. Usually composing $100 \sim 1000$ bootstrap samples will significantly prolong the analytic time span. Our method gets rid of this problem by providing a closed-form expressions of the asymptotic variance estimates. We will compare our approach with **icenReg** later in both simulation study and real data analysis.

In Section 2, we provide the methodological details of the proposed method. These

details include the use of monotone splines for approximating the baseline odds function in the PO model, a four-stage data augmentation process that leads to the development of an EM algorithm that can be used to find the maximum likelihood estimates of all unknown parameters, and closed-form expressions of the asymptotic variance estimates. In Section 3, the performance of the proposed approach is evaluated in simulated data against competing package *icenReg*. In Section 4 the proposed approach is applied to three real datasets. Section 5 provides a summary discussion and future plans.

## 2 The Method

### 2.1 Proportional Odds Model

Let $T_i$ denote the survival time of interest and $\mathbf{x}_i$ a $p \times 1$ vector of potential covariates for subject $i$, for $i = 1, \cdots, n$. In this article, we take a general notation $[L_i, R_i]$ to the observed interval for the failure time $T_i$, with $0 \le L_i \le R_i \le \infty$. This general interval yields an exactly observed failure time when $0 < L_i = R_i < \infty$, a left-censored observation when $0 = L_i < R_i < \infty$, a strictly interval-censored observation when $0 < L_i < R_i < \infty$, and a right-censored observation when $0 < L_i < R_i = \infty$. It is assumed that the failure time is conditionally independent with the observational process (i.e., the set of examination times) given covariates. Under this non-informative censoring assumption, the observed likelihood takes the following form

$$L_{obs} = \prod_{i=1}^{n} f(R_i|\mathbf{x}_i)^{\delta_{i0}} \{1 - S(R_i|\mathbf{x}_i)\}^{\delta_{i1}} \{S(L_i|\mathbf{x}_i) - S(R_i|\mathbf{x}_i)\}^{\delta_{i2}} \{S(L_i|\mathbf{x}_i)\}^{\delta_{i3}}, \qquad (1)$$

where $f(t|\mathbf{x})$ and $S(t|\mathbf{x})$ are the density and survival functions respectively given covariate $\mathbf{x}$, and $\delta_{i0}$, $\delta_{i1}$, $\delta_{i2}$, and $\delta_{i3}$ are all binary censoring indicators for exactly observed, left-censored, interval-censored, and right-censored observations, respectively, with the constraint $\delta_{i0} + \delta_{i1} + \delta_{i2} + \delta_{i3} = 1$ for subject $i$.

Under the PO model, the survival function is specified as $S(t|\mathbf{x}) = \{1 + \Lambda_0(t) \exp(\mathbf{x}'\boldsymbol{\beta})\}^{-1}$ and the density function takes the form

$$f(t|\mathbf{x}) = \frac{\Lambda_0'(t) \exp(\mathbf{x}_i'\boldsymbol{\beta})}{\{\Lambda_0(t) \exp(\mathbf{x}_i'\boldsymbol{\beta}) + 1\}^2},$$

where $\Lambda_0(t) = F_0(t)/\{1 - F_0(t)\}$ is the baseline odds function, $F_0(t)$ is the baseline cumulative distribution function when $\mathbf{x} = 0$, and $\Lambda_0'(t)$ is the first derivative of $\Lambda_0(t)$. Note that the baseline odds function $\Lambda_0(t)$ is an usepecified non-negative and non-decreasing function, and $\Lambda_0'(t)$ is a non-negative function under the PO model. Thus, the unknown parameters in the observed likelihood (1) include the regression parameters $\boldsymbol{\beta}$, the baseline odds function $\Lambda_0(t)$, and its derivative $\Lambda_0'(t)$.

## 2.2 Monotone Splines

The infinite dimension in the baseline odds function $\Lambda_0(\textbf{.})$ and its derivative $\Lambda_0'(\textbf{.})$ causes great trouble from both theoretic and computational perspectives. To reduce the number of unknown parameters while still allowing adequate modeling flexibility, we adopt the monotone splines of Ramsay (1988) for modeling $\Lambda(\textbf{.})$ as follows,

$$\Lambda_0(t) = \sum_{l=1}^{K} \gamma_l b_l(t), \tag{2}$$

where $b_l(\textbf{.})$'s are integrated spline (or I-spline) basis functions and $\gamma_l$'s are non-negative spline coefficients to ensure the monotonicity of the $\Lambda_0$. Each of the I-spline basis function is a

piecewise polynomial of specified degree $d-1$ or order $d$, taking 0 in an initial flat region, increasing in a mid region, and remaining at at 1 in the third region(Wang & Dunson 2011). The same or similar strategy has been effectively used to model unknown non-decreasing functions such as the transformed baseline cumulative distribution function in the probit model (Lin & Wang 2010), the logarithm of the baseline odds in the PO model (Wang & Dunson 2011), and the cumulative hazard function in the PH model (McMahan et al. 2013) among others.

Another benefit of using the I-splines is that it allows us to model $\Lambda_0'(\cdot)$ directly without the need of introducing any additional parameters. That is,

$$\Lambda_0'(t) = \sum_{l=1}^{K} \gamma_l M_l(t),$$

where $M_l(.)$'s are the so-called M-splines, the derivatives of the I-splines. All these basis functions are determined once the degree and knots are specified and can be obtained using iterative algorithms in our R functions. Note that these functions are calculated just once and do not need to be recalculated during the estimation process.

In general the degree determines the smoothness of the monotone splines, and together with the degree the placement of the knots determines the shape of the splines. Setting degree as 2 or 3 typically provides adequate smoothness. As for the placement of knots, it is reported that using $10 \sim 30$ equally-spaced knots provides adequate modeling flexibility for data sets containing up to thousands of observations (Cai et al. 2011, Wang & Dunson 2011). Lin & Wang (2011) and Lin et al. (2015) showed that for general interval-censored data, adopting equally-spaced knots in their methods outperforms the strategy of using quantile-based knots in terms of two commonly used Bayesian model selection criteria: the deviance information

criteria (DIC) and the logarithm of psedu-marginal likelihood (LPML). It is worthy noting that those Bayesian methods employ shrinkage priors for the spline coefficients and thus allow to use a large number of knots without causing over-fitting problems (Cai et al. 2011, Wang & Dunson 2011, Lin & Wang 2011). From a frequentist perspective, we recommend to follow the idea of Rosenberg (1995), McMahan et al. (2013), and Wang et al. (2016) to determine the number of knots. That is, we will fit the PO model using our method with different values for the number of knots and then choose the number that leads to the smallest value of Akaike's information criterion (AIC) or Bayesian information criteria (BIC).

## 2.3 A four-stage data augmentation

Direct optimization of the observed likelihood (1) encountered many numerical problems such as non-convergence from our experiences even though the number of unknown parameters is finite with the use of monotone splines. The main reason is that the optimization is very sensitive to initial values of the spline coefficients in addition to the complexity of the observed likelihood. To overcome such difficulties, we seek to explore an EM algorithm to obtain the MLE. To this end, we first introduce a four-stage data augmentation that leads to a complete data likelihood that has a nice form for our EM algorithm. The details of the four-stage data augmentation are given below.

The first-stage augmentation takes advantage of the relationship between the proportional odds model and the frailty proportional hazards model (Shen (1998), Murphy et al. (1997), McMahan et al. (2013) ). Specifically, one can write the survival function of the proportional odds model as the marginal survival function in the frailty proportional hazards model with

the frailty following an exponential distribution with mean 1 in the following manner,

$$S(t|\mathbf{x}) = \{\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}) + 1\}^{-1} = \int_0^\infty \exp\{-\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\phi\}\exp(-\phi)d\phi. \qquad (3)$$

Based on this, we introduce independent latent variables $\phi_i \sim \mathcal{E}xp(1)$ for all subjects and $\psi_i \sim \mathcal{E}xp(1)$ only for those exactly observed subjects (i.e, $\delta_{i0} = 1$). Conditioning on these latent variables, the augmented data likelihood takes the following form

$$
\begin{aligned}
\mathcal{L}_1 = \prod_{i=1}^n & \{\Lambda_0'(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_{i0}}\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i+\psi_i)\}^{\delta_{i0}} \\
& \times [1 - \exp\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i1}} \\
& \times [\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\} - \exp\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i2}} \\
& \times [\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i3}} \\
& \times \exp(-\phi_i)\exp(-\psi_i\delta_{i0}).
\end{aligned}
\qquad (4)
$$

In the augmented likelihood $\mathcal{L}_1$, the middle three multiplicative terms essentially form the likelihood for interval-censored data under the PH model (Lin et al. (2015), Wang et al. (2016)). Our second-stage data augmentation generalizes the ideas in Lin et al. (2015) and Wang et al. (2016) with additional frailties. In order to make this part self-complete, we provide the following motivations and justifications.

Let $N_i(t)$ denote a latent non-homogeneous Poisson process with cumulative intensity function $\Lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i$ conditioning on unobserved frailty $\phi_i$ for subject $i$ with $\delta_{i0} = 0$ (i.e., not exactly observed). Define two time points $t_{i1}$ and $t_{i2}$ based on the observed data for subject $i$ as follows: $t_{i1} = R_iI(\delta_{i1} = 1) + L_iI(\delta_{i2} = 1)$ and $t_{i2} = R_iI(\delta_{i2} = 1) + L_iI(\delta_{i3} = 1)$. Then define $Z_i = N(t_{i1})$ and $W_i = N(t_{i2}) - N(t_{i1})$ for subject $i$ with $\delta_{i0} = 0$ (i.e., not exactly observed). Based on the independent increment property of the Poisson process,

$Z_i$ and $W_i$ are conditionally independent Poisson random variables given frailty $\phi_i$, $Z_i \sim$ Poisson$(\Lambda_0(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)$, and $W_i \sim$ Poisson$(\{\Lambda_0(t_{i2}) - \Lambda_0(t_{i1})\} \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)$ for subject $i$ with $\delta_{i0} = 0$. With the introduced latent variables $Z_i$'s and $W_i$'s, the augmented data likelihood in this stage is

$$\mathcal{L}_2 = \prod_{i=1}^n \{\Lambda_0'(L_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_{i0}} \exp\{-\Lambda_0(L_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i + \psi_i)\delta_{i0}\}$$
$$\times \mathcal{P}(Z_i)^{1-\delta_0} \mathcal{P}(W_i)^{\delta_{i2}+\delta_{i3}} \exp(-\psi_i\delta_{i0}) \exp(-\phi_i) \tag{5}$$

where $\mathcal{P}(\cdot)$ denote the Poisson probability mass function. In this augmented likelihood, the latent variables $Z_i$'s and $W_i$'s are subject to the following constraints: $Z_i > 0$ if $\delta_{i1} = 1$; $W_i > 0$ and $Z_i = 0$ if $\delta_{i2} = 1$; and $W_i = Z_i = 0$ if $\delta_{i3} = 1$. Integrating out the augmented likelihood with respect to $Z_i$'s and $W_i$'s leads to the second augmented likelihood function in equation (4).

To fully taking advantage of the additive form of the monotone spline representation (2), we decompose both $Z_i$ and $W_i$ as a sum of K conditionally independent Poisson random variables given $\phi_i$ as follows $Z_i = \sum_{l=1}^K Z_{il}$ and $W_i = \sum_{l=1}^K W_{il}$, where

$$Z_{il} \sim \text{Poisson}(\gamma_l b_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i),$$

$$W_{il} \sim \text{Poisson}(\{b_l(t_{i2}) - b_l(t_{i1})\}\gamma_l \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i),$$

for each $i$ with $\delta_{i0} = 0$. Treating all $Z_{il}$'s and $W_{il}$'s as missing data, the augmented likelihood in this stage takes the following form

$$\mathcal{L}_3 = \prod_{i=1}^n \{\Lambda_0'(L_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_{i0}} \exp\{-\Lambda_0(L_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i + \psi_i)\delta_{i0}\} \exp(-\psi_i\delta_{i0}) \exp(-\phi_i)$$
$$\times \prod_{l=1}^K \mathcal{P}(Z_{il})^{1-\delta_{i0}} \mathcal{P}(W_{il})^{\delta_{i2}+\delta_{i3}}, \tag{6}$$

where the latent variables are subject to the following constraints: $\sum_{l=1}^{K} Z_{il} > 0$ when $\delta_{i1} = 1$; $\sum_{l=1}^{K} W_{il} > 0$ and $Z_{il} = 0$ for $l = 1, \cdots, K$ when $\delta_{i2} = 1$; , $Z_{il} = 0$ and $W_{il} = 0$ for $l = 1, \cdots, K$ when $\delta_{i3} = 1$.

Notice that the first multiplicative term in the augmented likelihood (6) involves the summation of M-splines for exactly observed observations. In order to facilitate the computation and get rid of the summation, we introduce a multinomial latent vector $(U_{i1}, ..., U_{iK}) \sim Multinomial\{1, (\frac{1}{K}, ..., \frac{1}{K})\}$ in the forth stage of our augmentation for subject $i$ with $\delta_{i0} = 1$. It is clear that summing $\prod_{l=1}^{K} [\gamma_l M_l(t)]^{u_l}$ over all possible combinations of $(u_1, ..., u_k)$ leads to $\sum_{l=1}^{K} \gamma_l M_l(t)$. With these new latent variables, the augmented data likelihood function now has the following multiplicative form:

$$\mathcal{L}_c = \prod_{i=1}^{n} \exp(\mathbf{x}_i' \boldsymbol{\beta} \delta_{i0}) \exp\{-\sum_{l=1}^{K} \gamma_l b_l(t_{i1}) \exp(\mathbf{x}_i' \boldsymbol{\beta})(\phi_i + \psi_i)\delta_{i0}\} \exp(-\psi_i \delta_{i0}) \exp(-\phi_i)$$
$$\times \prod_{l=1}^{K} \{\gamma_l M_l(R_i)\}^{U_{il} \delta_{i0}} P(Z_{il})^{1-\delta_{i0}} P(W_{il})^{\delta_{i2}+\delta_{i3}},$$

subject to the same constrains for the augmented likelihood $\mathcal{L}_3$. The augmented data likelihood (7) is extremely appealing because it only contains multiplicative terms of simple functions and will be viewed as the complete data likelihood for the derivation of our EM algorithm below.

## 2.4 EM Algorithm

The basic idea of EM algorithm is obtaining MLE through maximizing the surrogate function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ rather than the actual observed likelihood function. Let $\mathcal{D}$ denote all the observed data, $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\gamma}^{(d)'})'$ denote the current parameter estimate, then $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = E[\log\{L_c(\boldsymbol{\theta})\}]$ is the expectation of logarithm of the complete data likelihood

with respect to the latent variables condition on the observed data $\mathcal{D}$ and the current parameter estimate $\boldsymbol{\theta}^{(d)}$. To provide a concise expression, we will omit $\mathcal{D}$ and $\boldsymbol{\theta}^{(d)}$ in all the conditional expectation in the rest of the paper. Benefit to the well-designed data augmentation, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ has not only a linear form, but all the conditional expectations of the latent variables in it have explicit forms.

$$
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n} \left\{ \mathbf{x}_i' \boldsymbol{\beta} \delta_{i0} - E(\psi_i)\delta_{i0} - E(\phi_i) + \sum_{l=1}^{K} \left[ \{ E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2} \} \log(\gamma_l) \right. \right.
$$
$$
+ \{ E(\psi_i)\delta_{i0}b_l(L_i) + E(\phi_i)(b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})) \} \exp(\mathbf{x}_i'\boldsymbol{\beta})\gamma_l
$$
$$
\left. \left. + \{ E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2} \} \mathbf{x}_i'\boldsymbol{\beta} \right] \right\} + g(\boldsymbol{\theta}^{(d)})
$$

Firstly, the complete likelihood implies $E(U_{il})$ only exist when $\delta_{i0} = 1$ (i.e., data is exact observed). Since $(U_{i1}, ..., U_{iK}) \sim \mathcal{M}ultinomial\{1, (p_{i1}, ..., p_{iK})\}$ where $p_{il} = \frac{\gamma_l M_l(L_i)}{\Lambda_0'(L_i)}$. We have

$$
E(U_{il}) = \frac{\gamma_l^{(d)} M_l(L_i)}{\sum_{l=1}^{K} \gamma_l^{(d)} M_l(L_i)}.
$$

Secondly, from the design of third augmentation, it is easy to see that the conditional distribution of $Z_{il}$'s and $W_{il}$'s given $Z_i$ and $W_i$, respectively, are multinomial distributions. By applying the law of iterated expectations, one can obtain the expectations

$$
\begin{aligned}
E(Z_{il}) &= \frac{\gamma_l^{(d)} b_l(R_i)}{\sum_{j=1}^{K} \gamma_j^{(d)} b_j(R_i)} E(Z_i) \\
E(W_{il}) &= \frac{\gamma_l^{(d)} [b_l(R_i) - b_l(L_i)]}{\sum_{j=1}^{K} \gamma_j^{(d)} [b_j(R_i) - b_j(L_i)]} E(W_i).
\end{aligned}
$$

Thirdly, $Z_i$'s and $W_i$'s are the Poisson random variables introduced in the second data augmentation. By taking into account of the constrains in the augmented likelihood, $Z_i$'s and $W_i$'s conditionally follow truncated Poisson distribution given $\mathcal{D}$ and $\phi_i$'s. Therefore,

we have to use iterated rule again to get $E(Z_i)$. Noting the distribution function of $\phi_i$'s are needed in this part, but the pdfs are not available by extracting $\phi_i$'s kernel directly from the complete likelihood since that kernel has $Z_{il}$ and $W_{il}$ in it. So we integrated $Z_{il}$ and $U_{il}$ out from the complete likelihood function or, equivalently, extracted $\phi_i$'s kernel from the first augmented likelihood (Equation 4). After normalizing the kernel, i.e., find the constant part of the distribution function, we plugged in the distribution functions of $\phi_i$'s while using iterated rule. In this way, the expected values of $Z_i$ and $W_i$ given $\mathcal{D}$ and $\boldsymbol{\theta}^{(d)}$ have the form :

$$E(Z_i) = \Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1,$$

$$E(W_i) = \frac{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}$$

Lastly, as the probability distribution function of $\phi_i$ is known by now, we have the following conclusion:

$$E(\phi_i) = \begin{cases} \frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1} & \delta_{i0} = 1 \text{ or } \delta_{i3} = 1 \\ \frac{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+2}{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1} & \delta_{i1} = 1 \\ \frac{\{\Lambda_0^{(d)}(L_i)+\Lambda_0^{(d)}(R_i)\}\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+2}{\{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1\}\{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1\}} & \delta_{i2} = 1 \end{cases}$$

Similarly, we get $E(\psi_i) = \frac{1}{\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})+1}$ which only exists when data is exactly observed. The M-step in the EM algorithm devotes to find $\boldsymbol{\theta}^{(d+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Consider

the partial derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$:

$$\frac{\partial Q}{\partial \gamma_l} = \sum_{i=1}^{n} \left( \{E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}\} \frac{1}{\gamma_l} \right.$$

$$\left. - \left[ E(\psi_i)\delta_{i0}b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \right] \exp(\mathbf{x}_i'\boldsymbol{\beta}) \right)$$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left\{ \delta_{i0} + E(Z_i)\delta_{i1} + E(W_i)\delta_{i2} \right.$$

$$\left. - \sum_{l=1}^{K} \gamma_l \exp(\mathbf{x}_i'\boldsymbol{\beta}) \left[ E(\psi_i)\delta_{i0}b_l(L_i) + E(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \right] \right\} \mathbf{x}_i$$

Therefore, $\boldsymbol{\theta}^{(d+1)}$ is solution to the system of equations: $\frac{\partial Q}{\partial \boldsymbol{\beta}} = 0$ and $\frac{\partial Q}{\partial \gamma_l} = 0$ for $l = 1, ..., K$. Solving $\frac{\partial Q}{\partial \gamma_l} = 0$ for $\gamma_l$ offers a closed-form expression for $\gamma_l^{(d+1)}$ in terms of $\boldsymbol{\beta}^{(d+1)}$ for each $l$. Thus, by plug in this expression of $\gamma_l$ in $\frac{\partial Q}{\partial \boldsymbol{\beta}} = 0$ one can directly obtain $\boldsymbol{\beta}^{(d+1)}$, which then allows for the direct calculation of $\gamma_l^{(d+1)}$.

Although the derivation seems sophisticated, the EM algorithm that is actually used to fit the model turns out to be quite succinct and easy to operate. The whole process can be summarized as follows. First set $d = 0$ and initialize $\boldsymbol{\theta}^{(d)} = \boldsymbol{\theta}(\boldsymbol{\beta}^{(d)}, \gamma^{(d)})$. Then repeat the following two steps until convergence:

   **1.** Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following system of p equations

$$\sum_{i=1}^{n} \{\delta_{i0} + E(Z_i)\delta_{i1} + E(W_i)\delta_{i2}\} \mathbf{x}_i$$

$$= \sum_{i=1}^{n} \sum_{l=1}^{K} \gamma_l^* \left[ E(\psi_i)\delta_{i0}b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \right] \exp(\mathbf{x}_i'\boldsymbol{\beta}) \mathbf{x}_i \tag{7}$$

where

$$\gamma_l^*(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} \{E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}\}}{\left[ E(\psi_i)\delta_{i0}b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \right] \exp(\mathbf{x}_i'\boldsymbol{\beta})}. \tag{8}$$

**2.** Let $\gamma_l^{(d+1)} = \gamma_l^*(\boldsymbol{\beta}^{(d+1)})$ and increase d by 1.

Solving the system of equations in the first step of the iteration part can be accomplished using standard root finding routines, available in practically all existing statistical software packages. The second step of the iteration part is a simple updating of $\gamma_l^{(d)}$ in closed form. Thus, the implementation of the EM algorithm is straightforward and computationally inexpensive. Moreover, it can be shown that $\gamma_l^{(d+1)}$ is the unique global maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ and $\hat{\boldsymbol{\theta}}$ solves the score equations based on the observed likelihood. The proof is shown in Appendix B.

## 2.5   Asymptotic Properties and Variance Estimation

We provide an asymptotic variance-covariance matrix estimate for the proposed estimator in this section. Since our estimator is actually an MLE, it has all the good properties of MLE in general. Suppose the number and position of the knots are pre-specified and do not depend on the sample size, under the standard regularity conditions, as $n \to \infty$, $\hat{\boldsymbol{\theta}} \overset{a}{\sim} N(\boldsymbol{\theta}, \{I(\boldsymbol{\theta})\}^{-1})$. In this expression, the Fisher information matrix $I(\boldsymbol{\theta})$ can be easily got by using Louis's method (Louis (1982)) as shown in Equation 9.

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\{\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} | \mathcal{D}, \hat{\boldsymbol{\theta}}\} \tag{9}$$

Depending on the missing information principle, Louis's method is one of the most commonly used ways to get the Fisher information matrix and further get an asymptotic estimator of the variance-covariance matrix while using EM algorithm. In addition, benefit from the exquisite four-step augmentation, all the entries in $\text{var}\{\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} | \mathcal{D}, \hat{\boldsymbol{\theta}}\}$ and $\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ have explicit forms. Thus Wald inference and the according confidence intervals for the regression coefficients $\boldsymbol{\beta}$

along with the baseline odds function can be obtained easily. The details pertaining to the calculation of the two matrix on the right hand side are provided, in closed-form, in Appendix A.

# 3  Simulation

A simulation study is provided to evaluate the proposed method. We also compared it with *ic_np* function in *Icenreg* in this part. We adopted the same scheme as Lin et al. (2015) to generate observation times. First of all, each individual failure time $T$ was generated from

$$F(t|\mathbf{x}) = \frac{\Lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2)}{1 + \Lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2)},$$

where $\Lambda_0(t)$ is the baseline odds function, $x_1$ is a $N(0,1)$ random variable and $x_2$ is a *Bernoulli*(0.5) random variable. Obviously, by letting $\beta_i = \{-1, 0, 1\}$ for $i = 1, 2$, we can easily assess characteristics of the method under continuous and categorical data. Moreover, to conduct a detail and thorough assessment, two different baseline functions are employed to generate data in two groups of simulations separately. One baseline odds has the form

$$\Lambda_0(t) = \log(1 + t) + t^3 + sin(t)$$

another has the form

$$\Lambda_0(t) = \log(1 + t) + t^{1.5}.$$

The former has more complicated form but the latter allows $T$ having a heavier tail distribution.

In the second step of data generation, indicators follow *Bernoulli*(0.3) were generated for all the failure times. If an observation's indicator equals 1 we say its failure time is

exactly observed, so that the observed interval has lower bound and upper bound both equal to the true failure time. Otherwise, a series of examination times will be assigned to this individual. We allowed each individual to have a random number of observation times, which determined by 1 plus a Poisson random variable with mean 6. The gap times between adjacent observation times were generated from independent exponential distributions with mean 0.2. Subsequently, the observation times were given by the cumulative sums of the gap times. Naturally, the observed interval was determined by the two adjacent observation times (including 0 and $\infty$ ) that bracket the true failure time. According to this data generating scheme, we have around 30% of exactly observed data, 20% of left censored data, 25% of interval censored data and 25% of right censored data in each data set. In this way, 500 independent data sets with sample size $n = 200$ were generated for each configuration.

We applied the proposed method and $ic\_np$ to each simulated data set under PO model. In the aspect of estimating the unknown baseline odds functions, we fixed the degree of I-spline to 3 and took 9 equally spaced knots between 0 and maximum of the finite endpoints of all the observed intervals. The simulation results from both methods are shown side by side in Table 1, Table 2 and Table 3. Table 1 and Table 2 illustrate the frequentist operating characteristics of the estimates of the regression parameters from both methods. They summarized the average empirical bias (BIAS), the average of the 500 estimated standard errors (ESD), sample standard deviation of the 500 point estimates (SSD), and the empirical coverage probability associated with 95% Wald confidence intervals for each of the regression parameters with different baseline odds functions.

As shown in Table 1 and Table 2, the small values of BIAS suggest that the point estimate of regression parameters obtained by the proposed method are all close to their corresponding

true values. Additionally, ESD and SSD are in agreement in every configurations, which means the asymptotic approximation of the variance-covariance matrix obtained from Louis's method performs well for finite samples. And the empirical coverage probabilities (CP95) for the confidence intervals for the regression parameters are close to the nominal level 0.95, which indicates that the Wald-type inference is reliable for estimates obtained by the EM algorithm. Furthermore, comparing the proposed approach to $ic\_sp$, we can see the proposed methodology performed as well, if not better, than $ic\_sp$ both in terms of parameter estimation and inferential characteristics across all considered configurations. In addition, since the proposed approach provides a closed-form expression of the asymptotic variance estimates while $ic\_sp$ needs to use bootstrap standard errors, our method has a full advantage in the model fitting times. Specifically, we ran all the simulations in R on a computer with a 3.60 GHz processor and 32.0 GB of memory. On average, $ic\_sp$ took 3 times longer to complete model fitting when compared to the proposed approach. This advantage could render the proposed approach preferable when analyzing larger data sets. The reason is that the model fitting times for both methods increase as the sample size become larger, but the bootstrap in $ic\_sp$ significantly magnified the prolonged fitting time thus makes it much slower than the proposed method.

Table 1: Simulation results of estimated regression parameters when baseline odds is $\Lambda_0(t) = \log(1+t) + t^3 + \sin(t)$.

| Method | $\beta_1$ | BIAS | ESD | SSD | CP95 | $\beta_1$ | BIAS | ESD | SSD | CP95 |
|---|---|---|---|---|---|---|---|---|---|---|
| EM | 1 | 0.0222 | 0.1552 | 0.1727 | 0.9360 | 1 | 0.0439 | 0.2768 | 0.2875 | 0.9520 |
| Icenreg | | 0.0259 | 0.1642 | 0.1743 | 0.9420 | | 0.0455 | 0.2880 | 0.2887 | 0.9540 |
| EM | 0 | -0.0086 | 0.1320 | 0.1353 | 0.9520 | 1 | 0.0123 | 0.2730 | 0.2810 | 0.9440 |
| Icenreg | | -0.0089 | 0.1374 | 0.1355 | 0.9560 | | 0.0120 | 0.2779 | 0.2813 | 0.9480 |
| EM | 1 | 0.0212 | 0.1538 | 0.1606 | 0.9540 | 0 | -0.0055 | 0.2671 | 0.2754 | 0.9500 |
| Icenreg | | 0.0236 | 0.1617 | 0.1615 | 0.9620 | | -0.0050 | 0.2757 | 0.2759 | 0.9600 |
| EM | 0 | -0.0074 | 0.1298 | 0.1284 | 0.9460 | 0 | -0.0139 | 0.2582 | 0.2646 | 0.9400 |
| Icenreg | | -0.0072 | 0.1340 | 0.1283 | 0.9620 | | -0.0148 | 0.2668 | 0.2652 | 0.9480 |
| EM | 1 | 0.0184 | 0.1567 | 0.1611 | 0.9500 | -1 | -0.0068 | 0.2782 | 0.2882 | 0.9480 |
| Icenreg | | 0.0198 | 0.1640 | 0.1610 | 0.9540 | | 0.0073 | 0.2895 | 0.2882 | 0.9540 |
| EM | -1 | -0.0276 | 0.1546 | 0.1605 | 0.9380 | 1 | 0.0298 | 0.2766 | 0.2913 | 0.9320 |
| Icenreg | | -0.0297 | 0.1614 | 0.1617 | 0.9440 | | 0.0319 | 0.2862 | 0.2928 | 0.9460 |
| EM | 0 | 0.0059 | 0.1324 | 0.1347 | 0.9540 | -1 | -0.0263 | 0.2694 | 0.2660 | 0.9560 |
| Icenreg | | 0.0066 | 0.1377 | 0.1350 | 0.9480 | | -0.0281 | 0.2807 | 0.2650 | 0.9640 |
| EM | -1 | -0.0261 | 0.1542 | 0.1571 | 0.9440 | 0 | 0.0021 | 0.2668 | 0.2680 | 0.9480 |
| Icenreg | | -0.0295 | 0.1636 | 0.1575 | 0.9580 | | 0.0024 | 0.2780 | 0.2692 | 0.9520 |
| EM | -1 | -0.0234 | 0.1611 | 0.1514 | 0.9440 | -1 | -0.0204 | 0.2899 | 0.2941 | 0.9220 |
| Icenreg | | -0.0265 | 0.1651 | 0.1523 | 0.9540 | | -0.0221 | 0.2904 | 0.2935 | 0.9480 |

Table 2: Simulation results of estimated regression parameters when baseline odds is $\Lambda_0(t) = \log(1 + t) + t^{1.5}$.

| Method | $\beta_1$ | BIAS | ESD | SSD | CP95 | $\beta_2$ | BIAS | ESD | SSD | CP95 |
|---|---|---|---|---|---|---|---|---|---|---|
| EM | 1 | 0.0061 | 0.1554 | 0.1499 | 0.9580 | 1 | 0.0074 | 0.2781 | 0.2763 | 0.9540 |
| Icenreg | | 0.0086 | 0.1643 | 0.1509 | 0.9600 | | 0.0097 | 0.2903 | 0.2782 | 0.9580 |
| EM | 0 | -0.0015 | 0.1329 | 0.1313 | 0.9540 | 1 | 0.0151 | 0.2719 | 0.2735 | 0.9480 |
| Icenreg | | -0.0016 | 0.1383 | 0.1318 | 0.9620 | | 0.0166 | 0.2815 | 0.2742 | 0.9480 |
| EM | 1 | 0.0025 | 0.1538 | 0.1516 | 0.9500 | 0 | 0.0050 | 0.2674 | 0.2667 | 0.9520 |
| Icenreg | | 0.0053 | 0.1600 | 0.1519 | 0.9580 | | 0.0049 | 0.2749 | 0.2673 | 0.9640 |
| EM | 0 | 0.0047 | 0.1301 | 0.1359 | 0.9360 | 0 | -0.0116 | 0.2593 | 0.2467 | 0.9680 |
| Icenreg | | 0.0046 | 0.1357 | 0.1360 | 0.9460 | | -0.0108 | 0.2667 | 0.2481 | 0.9640 |
| EM | 1 | 0.0212 | 0.1569 | 0.1607 | 0.9580 | -1 | -0.0169 | 0.2779 | 0.2907 | 0.9400 |
| Icenreg | | 0.0237 | 0.1655 | 0.1612 | 0.9520 | | -0.0203 | 0.2875 | 0.2914 | 0.9480 |
| EM | -1 | 0.0004 | 0.1551 | 0.1546 | 0.9580 | 1 | 0.0341 | 0.2783 | 0.2798 | 0.9560 |
| Icenreg | | -0.0045 | 0.1626 | 0.1551 | 0.9620 | | 0.0380 | 0.2918 | 0.2807 | 0.9580 |
| EM | 0 | 0.0074 | 0.1313 | 0.1375 | 0.9420 | -1 | -0.0078 | 0.2706 | 0.2741 | 0.9520 |
| Icenreg | | 0.0072 | 0.1365 | 0.1374 | 0.9500 | | -0.0077 | 0.2760 | 0.2738 | 0.9540 |
| EM | -1 | -0.0160 | 0.1550 | 0.1605 | 0.9420 | 0 | -0.0086 | 0.2677 | 0.2658 | 0.9500 |
| Icenreg | | -0.0189 | 0.1629 | 0.1623 | 0.9540 | | -0.0084 | 0.2763 | 0.2665 | 0.9520 |
| EM | -1 | -0.0319 | 0.1574 | 0.1608 | 0.9500 | -1 | -0.0415 | 0.2788 | 0.2819 | 0.9480 |
| Icenreg | | -0.0337 | 0.1640 | 0.1623 | 0.9500 | | -0.0425 | 0.2895 | 0.2824 | 0.9520 |

Table 3 shows the comparison of mean squared errors (MSE) for regression parameters between two methods. If look closely, our approach has advantage in this regard which indicates the proposed approach is able to offer a better fit. Meanwhile, Table 3 illustrates the performance of both methods in estimating the baseline odds functions. In order to show the estimation accuracy globally, we calculated the mean squared errors (MSE) of the estimates of the baseline survival function $S_0(t)$ at a set of pre-specified grid points (119 evenly-spaced points between 0 and 6). The mean and the maximum of these local MSEs are calculated and named meanMSE and maxMSE in Table 3, respectively. As shown in

Table 3: MSE of regression parameters, mean and maximum of local MSEs for the estimated survival function $S_0(t)$.

| $(\beta_1, \beta_2)$ | M | $\Lambda_0(t) = \log(1+t) + t^{1.5}$ | | | | $\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathrm{MSE}_{\beta_1}$ | $\mathrm{MSE}_{\beta_2}$ | meanMSE | maxMSE | $\mathrm{MSE}_{\beta_1}$ | $\mathrm{MSE}_{\beta_2}$ | meanMSE | maxMSE |
| (1,1) | EM | 0.0225 | 0.0763 | 0.0020 | 0.0027 | 0.0302 | 0.0844 | 0.0021 | 0.0029 |
| | Icenreg | 0.0228 | 0.0773 | 0.0028 | 0.0044 | 0.0310 | 0.0853 | 0.0012 | 0.0032 |
| (0,1) | EM | 0.0172 | 0.0749 | 0.0020 | 0.0027 | 0.0183 | 0.0789 | 0.0024 | 0.0034 |
| | Icenreg | 0.0173 | 0.0753 | 0.0029 | 0.0047 | 0.0184 | 0.0791 | 0.0013 | 0.0028 |
| (1,0) | EM | 0.0229 | 0.0709 | 0.0016 | 0.0025 | 0.0262 | 0.0757 | 0.0017 | 0.0028 |
| | Icenreg | 0.0230 | 0.0713 | 0.0023 | 0.0047 | 0.0266 | 0.0760 | 0.0011 | 0.0033 |
| (0,0) | EM | 0.0184 | 0.0609 | 0.0018 | 0.0025 | 0.0165 | 0.0702 | 0.0016 | 0.0026 |
| | Icenreg | 0.0185 | 0.0616 | 0.0025 | 0.0041 | 0.0165 | 0.0704 | 0.0011 | 0.0029 |
| (1,-1) | EM | 0.0262 | 0.0845 | 0.0016 | 0.0028 | 0.0262 | 0.0829 | 0.0015 | 0.0028 |
| | Icenreg | 0.0263 | 0.0831 | 0.0022 | 0.0086 | 0.0254 | 0.0825 | 0.0011 | 0.0034 |
| (-1,1) | EM | 0.0239 | 0.0793 | 0.0021 | 0.0029 | 0.0265 | 0.0856 | 0.0022 | 0.0030 |
| | Icenreg | 0.0265 | 0.0852 | 0.0028 | 0.0043 | 0.0270 | 0.0866 | 0.0013 | 0.0033 |
| (0,-1) | EM | 0.0189 | 0.0750 | 0.0015 | 0.0023 | 0.0181 | 0.0708 | 0.0015 | 0.0026 |
| | Icenreg | 0.0189 | 0.0749 | 0.0021 | 0.0046 | 0.0182 | 0.0709 | 0.0010 | 0.0029 |
| (-1,0) | EM | 0.0260 | 0.0706 | 0.0017 | 0.0027 | 0.0252 | 0.0720 | 0.0018 | 0.0027 |
| | Icenreg | 0.0267 | 0.0709 | 0.0025 | 0.0073 | 0.0256 | 0.0723 | 0.0011 | 0.0031 |
| (-1,-1) | EM | 0.0268 | 0.0810 | 0.0016 | 0.0029 | 0.0234 | 0.0867 | 0.0014 | 0.0028 |
| | Icenreg | 0.0274 | 0.0814 | 0.0024 | 0.0350 | 0.0238 | 0.0865 | 0.0010 | 0.0031 |

Table 3, the proposed method has an excellent performance in estimating the baseline odds function for all the simulation setups.

# 4   Real Data Application

To demonstrate that the algorithm finds the correct solution, results obtained from EM algorithm were compared with *icenReg*'s *ic_sp* function on three real datasets. Since *icenReg* needs bootstrap for estimating the standard errors, we fixed the number of bootstrap samples to 100 in all three analysis. The first dataset is a mixture of exact observed, left-, interval- and

right censored data, which dose not have any covariate; the second dataset is an example of type-2 interval censored data with one covariate; the third data set contains exactly observed and interval-censored data with one covariate.

## 4.1  First Use of Marijuana

In 1975, on the Stanford-Palo Alto Peer Counseling Program, Hamburg et al. studied drug use in a representative sample of suburban junior and senior high school students. In this study, they found a distinctive age-related pattern of drug use among students. Then in 1987, Turnbull & Weiss extracted a set of failure time data of marijuana use from Hamburg et al.'s study. As shown in Table 7, this data set summarizes the answer of 191 California high school students to the question "When did you first use marijuana?". A direct answer give rise to an exact observation. If the student answered, "I have never used it," then this gives rise to an observation which is censored on the right at his/her present age. The final possibility was someone who answered, "I have used it but cannot recall just when the first time was." This gives rise to a left censored observation where age of first use is known only to be prior to the student's current age.

Since there is no covariate in this dataset, we simply illustrate the estimated survival function in Figure 1. In Figure 1 the smooth dotted line and the solid step line are survival functions obtained by our EM algorithm and *ic_sp*, respectively. As can be seen from this figure, the two methods match each other very well.

Table 4: Ages in years to the first use of marijuana

| Age | No. of exact observation | No. of left-censored observations | No. of right-censored observations |
|---|---|---|---|
| 10 | 4 | 0 | 0 |
| 11 | 12 | 0 | 0 |
| 12 | 19 | 2 | 0 |
| 13 | 24 | 15 | 1 |
| 14 | 20 | 24 | 2 |
| 15 | 13 | 18 | 3 |
| 16 | 3 | 14 | 2 |
| 17 | 1 | 6 | 3 |
| 18 | 0 | 0 | 1 |
| >18 | 4 | 0 | 0 |

Figure 1: Time to the first use of marijuana
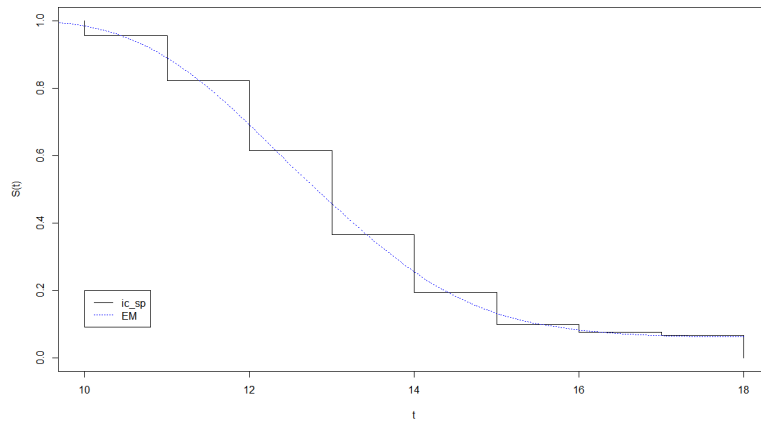


## 4.2 HIV

The dataset *Hemophilia* from r package **ICsurv** was collected in 1980's as part of a multi-center prospective study. Patients with hemophilia need blood products made from donors

plasma, so they are at risk of HIV-1 infection. This study was conducted to quantify the dose effect of blood products. Specifically, it aimed at assessing the HIV-1 infection rate in hemophilia patients with different average annual dose of blood products. In this study, 544 patients were classified into high, medium, low, or no dose group based on their average annual dose of blood products. The exact HIV-1 infection times were never observed and only observed intervals are available. Among all the patients, 63 of them are left-censored, 204 are interval-censored, and 277 are right-censored. Please refer to Goedert et al. (1989) and Kroner et al. (1994) for more detail about this study. This typical interval-censored data set has also been analyzed by Sun (2006), and Lin & Wang (2010). We specified monotone splines by fixing degree to 3 and taking 10 equally spaced interior knots within (0, 57.01).

The estimated regression coefficients for low, medium and high average annual doses obtained by the proposed EM algorithm and $ic\_sp$ are summarized in Table 5. From the Table 5, those two methods give comparable results. These results suggest that there is a significant dose effect between each dose group and the non-dose group. In particular, under the proportional odds assumption the odds of HIV infection for patients using low-, medium- and high- dose of blood products is estimated to be approximately 9.88, 68.92 and 174.69 times that for patients who do not use blood products.

In Figure 2 ∼ Figure 4, we superimposed the estimated survival functions obtained from the proposed approach and $ic\_np$ for all the dose groups. The difference among these survival functions is clearly seen by comparing the three plots. Meanwhile, the estimated survival function obtained by our method is consistent with the nonparametric estimate offered by $ic\_np$.

Table 5: HIV data analysis: estimated regression coefficients for the average annual blood products does level.

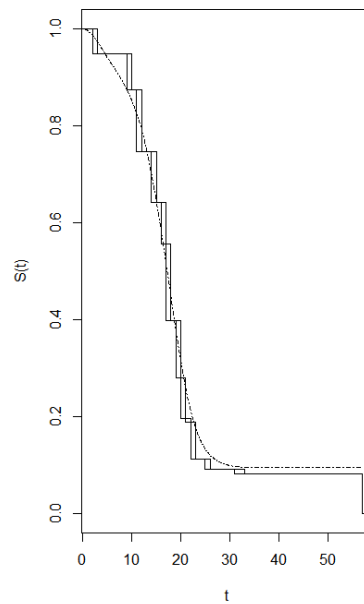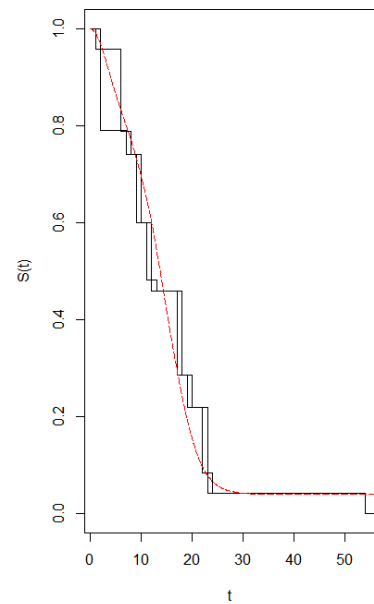| Dose-Level | Method | Estimate | Exp(Est) | Std. Error | z-value | p-value |
|---|---|---|---|---|---|---|
| Low | EM | 2.291 | 9.8848 | 0.2620 | 8.744 | 0.000 |
| | ic_sp | 2.278 | 9.7571 | 0.2682 | 8.494 | 0.000 |
| Medium | EM | 4.233 | 68.9237 | 0.3149 | 13.442 | 0.000 |
| | ic_sp | 4.226 | 68.4429 | 0.2772 | 15.250 | 0.000 |
| High | EM | 5.163 | 174.6877 | 0.3683 | 14.019 | 0.000 |
| | ic_sp | 5.163 | 174.6877 | 0.3775 | 13.680 | 0.000 |
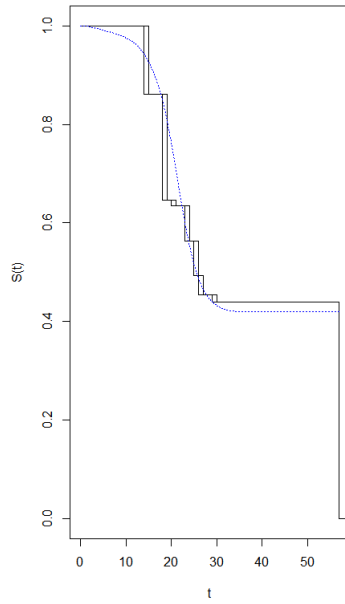


Figure 2: Low

Figure 3: Medium

Figure 4: High

## 4.3  Diabetes

The *IR_diabetes* dataset from **icenReg** was analyzed by Anderson-Bergman (2017) with function *ic_sp*. We reanalyzed it for the purpose of comparison. This dataset is initially introduced by Zhao & Sun (2015) in their package **glrt**. The data frame is based on a study conducted at the Steno Memorial Hospital in Denmark from 1933-1984. In this study, the time from onset of diabetes to onset of diabetic nephropathy is the response time of interest. It contains data from 731 patients (454 males and 277 females), for many of the patients (595), the event time was known exactly but for others (136) the event time was known only up to an interval due to limited follow up. The dataset contains three variables: left, right and gender. The variables left and right represent the observational interval and the effect of gender will be examined by the proposed EM algorithm under PO model. In this part, we also fixed the degree of I-spine to 3 for adequate flexibility and took 10 equally spaced interior knots within (0, 44.01).

The estimated regression coefficient for gender obtained by the proposed EM algorithm and *ic_sp* is summarized in Table 6. The proposed procedure resulted in practically identical estimates of the regression coefficients and inferential conclusions with *ic_sp*. From the result, it can be seen that there is a statistically significant difference in the odds of having experienced diabetic nephropathy at a given time after diabetes between men and women in the study. As the female is treated as baseline group, it is estimated that the odds of onset of diabetic nephropathy for men at any given time will be almost 0.681 times lower than for women under the assumption of proportional odds model.

Figure 5 plots the estimated survival functions for male and female obtained from the proposed method and *ic_sp*. The smooth dotted lines represent the estimated survival func-

tions obtained by the proposed EM algorithm and the solid step lines are the estimated survival functions obtained by *ic_sp*. Figure 5 indicates the odds of survival for males is estimated to be higher than females at all times.

Table 6: IR_diabetes data analysis: estimated regression coefficients for gender.

|  | Estimate | Exp(Est) | Std. Error | z-value | p-value | Time |
|---|---|---|---|---|---|---|
| EM | -0.3840 | 0.681 | 0.1728 | -2.223 | 0.0262 | 1.65 |
| *ic_sp* | -0.4013 | 0.669 | 0.1407 | -2.851 | 0.0044 | 24.14 |

Figure 5: Diabetes
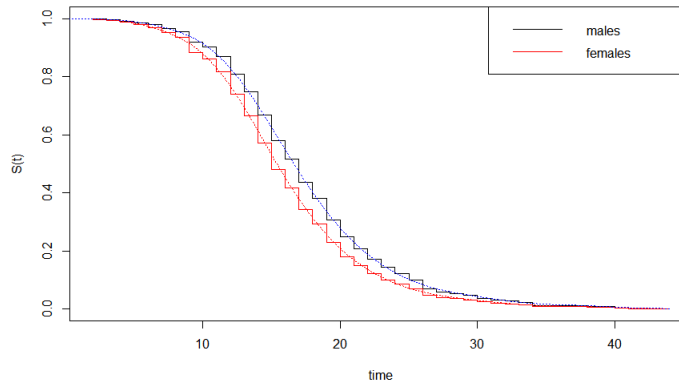


# 5  Discussion

In this article a new method for analyzing arbitrary censored data is proposed under the proportional odds model. After a reparameterization basing on the semi-parametric framework, an approximation of the nondecreasing baseline odds function has been achieved by using monotone splines, therefore leading to a finite number of parameters to estimate. The

EM algorithm developed in this paper can be used to find the maximum likelihood estimates of the baseline odds function and regression parameters simultaneously and to provide a closed-form of the asymptotic variance-covariance matrix. The key step in the derivation of the proposed EM algorithm is a four-step data augmentation which expanded the observed data likelihood to a complete data likelihood. The expanding process involves the relationship between PO model and frailty PH model along with the first failure time under the PH model with a latent non-homogeneous Poisson process. Simulation study and real dataset applications shown that the proposed method can provide accurate estimation results, takes less time and effort. It does not require model assumptions about the observational process and thus is widely applicable. In summary, the proposed approach can be easily applied to various situations, it is accurate, reliable, and computationally efficient. We expect our approach to be widely used for analyzing any arbitrarily censored data under proportional odds model.

# References

Alioum, A. & Commenges, D. (1996), 'A proportional hazards model for arbitrarily censored and truncated data', *Biometrics* **52**, 512524.

Anderson-Bergman, C. (2017), 'icenreg: Regression models for interval censored data in r', *Journal of Statistical Software* **81**.

Bennett, S. (1983a), 'Analysis of survival data by the proportional odds model', *Statistics in Medicine* **2**, 273–277.

Cai, B., Lin, X. & Wang, L. (2011), 'Bayesian proportional hazards model for current status data with monotone splines', *Computational Statistical Data Analysis* **55**, 26442651.

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

Hamburg, A. B., Kraemer, C. & Jahnke, W. (1975), 'A hierarchy of drug use in adolescence: Behavioral and attitudinal correlates of substantial drug use', *Am J Psychiatry* **132**, 1155–1163.

Huang, J. & Rossini, J. (1997), 'Sieve estimation for the proportional odds model with interval-censoring', *Journal of American Statistical Association* **92**, 960–967.

Lin, X., Cai, B., Wang, L. & Zhang, Z. (2015), 'A bayesian proportional hazards model for general interval-censored data', *Lifetime Data Analysis* **21**, 470–490.

Lin, X. & Wang, L. (2010), 'A semiparametric probit model for case 2 interval-censored failure time data', *Statistics in Medicine* **29**, 972–981.

Lin, X. & Wang, L. (2011), 'Bayesian proportional odds models for analyzing current status data: Univariate, clustered, and multivariate', *Communications in Statistics - Simulation and Computation* **40**, 1171–1181.

Louis, T. (1982), 'Finding the observed information matrix when using the em algorithm', *Journal of the Royal Statistical Society* **44**, 226–233.

McMahan, C., Wang, L. & Tebbs, J. (2013), 'Regression analysis for current status data using the em algorithm', *Statistics in Medicine* **32**, 4452–4466.

Murphy, S., Rossini, A. & van der Vaart, A. W. (1997), 'Maximum likelihood estimation in the proportional odds model', *Journal of the American Statistical Association* **92**, 968–976.

Pan, W. (1999), 'Extending the iterative convex minorant algorithm to the cox model for interval-censored data', *Journal of Computational and Graphical Statistics* **8(1)**, 109–120.

Pan, W. & Chappell, R. (1998), 'A nonparametric estimator of survival functions for arbitrarily truncated and censored data', *Lifetime Data Analysis* **4**, 187–202.

Peto, R. (1973), 'Experimental survival curves for interval-censored data', *Applied Statistics* **22**, 86–91.

Ramsay, J. O. (1988), 'Monotone regression splines in action', *Statistical Science* **3**, 425–461.

Rosenberg, P. S. (1995), 'Hazard function estimation using b-splines', *Biometrics* **51**, 874887.

Rossini, A. J. & Tsiatis, A. A. (1996), 'A semiparametric proportional odds regression model for the analysis of current status data', *Journal of the American Statistical Association* **91**, 713–721.

Royston, P. & Parmar, M. K. B. (2002), 'Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects', *Statistics in Medicine* **21**, 2175–2197.

Shen, X. T. (1998), 'Proportional odds regression and sieve maximum likelihood estimation', *Biometrika* **85**, 165–177.

Sun, J. (2006), 'The statistical analysis of interval-censored data', *Springer* .

Tu, X. M., Meng, X. & Pagano, M. (1993), 'The aids epidemic: Estimating survival after aids diagnosis from surveillance data', *Journal of the American Statistical Association* **88**, 26–36.

Turnbull, B. W. (1974), 'Nonparametric estimation of a survivorship function with doubly censored data', *Journal of the American Statistical Association* **69**, 169–173.

Turnbull, B. W. (1976), 'The empirical distribution function with arbitrarily grouped, censored and truncated data', *Journal of the Royal Statistical Society, Series B* **38**, 290–295.

Turnbull, B. W. & Weiss, L. (1978), 'A likelihood ratio statistic for testing goodness of fit with randomly censored data', *Biometrics* **34**, 367–375.

Wang, L. & Dunson, D. (2011), 'Semiparametric bayes proportional odds models for current status data with underreporting', *Biometrics* **67**, 1111–1118.

Wang, L.and McMahan, C., Hudgens, M. & Qureshi, Z. (2016), 'A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data', *Biometrics* **72**, 222–231.

Yang, S. & Prentice, R. L. (1999), 'Semiparametric inference in the proportional odds regression model', *Journal of the American Statistical Association* **94**, 125–136.

Zhao, Q. & Sun, J. (2015), 'glrt: Generalized logrank tests for interval-censored failure time data', *R package version 2.0* pp. URL https://CRAN.R–project.org/package=glrt.

# Appendix A. Asymptotic Variance-Covariance Matrix Estimation

Once applied the four-step data augmentation and get the complete likelihood function, it is easily to see that as $n \to \infty$, $\hat{\boldsymbol{\theta}} \overset{a}{\sim} N(\boldsymbol{\theta}, \{I(\boldsymbol{\theta})\}^{-1})$, under regularity conditions. Based on Louis's method:

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\{\frac{\partial \log L_c(\theta)}{\partial \boldsymbol{\theta}} | \mathcal{D}, \hat{\theta}\}$$

All the necessary entries in $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ are:

$$
\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\sum_{i=1}^{n} \sum_{l=1}^{K} \left[ E(\psi_i) b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \right] e^{\mathbf{x}_i'\boldsymbol{\beta}} \gamma_l \mathbf{x}_i \mathbf{x}_i' \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \gamma_l} &= -\sum_{i=1}^{n} \left[ E(\psi_i) b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \right] e^{\mathbf{x}_i'\boldsymbol{\beta}} \mathbf{x}_i \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \gamma_l \partial \gamma_l} &= -\sum_{i=1}^{n} \frac{E(U_{il}) + E(Z_{il}) + E(W_{il})}{\gamma_l^{(d)2}} \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \gamma_l \partial \gamma_k} &= 0, \qquad k \neq l
\end{aligned}
\tag{10}
$$

All the necessary entries in $\text{var}\{\partial \log L_c(\theta)/\partial \boldsymbol{\theta} | \mathcal{D}, \hat{\theta}\}$ are:

$$
\begin{aligned}
\text{var}(\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} | \mathcal{D}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \Big[ &\{\Lambda_0(L_i)\delta_{i0} \exp(\mathbf{x}_i'\boldsymbol{\beta})\}^2 var(\psi_i) \\
&+ \{\Lambda_0(L_i)(\delta_{i0} + \delta_{i3}) + \Lambda_0(R_i)(\delta_{i1} + \delta_{i2})\}^2 \exp(2\mathbf{x}_i'\boldsymbol{\beta}) var(\phi_i) \\
&+ var(Z_i)\delta_{i1} + var(W_i)\delta_{i2} - 2\text{cov}(\phi_i, Z_i)\delta_{i1}\Lambda_0(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta}) \\
&- 2\text{cov}(\phi_i, W_i)\delta_{i2}\Lambda_0(R_i) \exp(\mathbf{x}_i'\boldsymbol{\beta}) \Big] \mathbf{x}_i \mathbf{x}_i'
\end{aligned}
$$

$$\text{var}(\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \gamma_l}|\mathcal{D}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \Big( \frac{\text{var}(U_{il})\delta_{i0} + \text{var}(W_{il})\delta_{i2} + \text{var}(Z_{il})\delta_{i1}}{\gamma_l^2}$$

$$+ \big[\delta_{i0}b_l(L_i)^2\text{var}(\psi_i) + \{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\}^2\text{var}(\phi_i)\big] \exp(2\mathbf{x}_i'\boldsymbol{\beta})$$

$$- 2\frac{\delta_{i2}\exp(\mathbf{x}'\boldsymbol{\beta})b_l(R_i)}{\gamma_l}\text{cov}(W_{il}, \phi_i)$$

$$- 2\frac{\delta_{i1}\exp(\mathbf{x}'\boldsymbol{\beta})b_l(R_i)}{\gamma_l}\text{cov}(Z_{il}, \phi_i)\Big)$$

$$\text{cov}(\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}, \frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \gamma_l}|\mathcal{D}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \Big[ \Lambda_0(L_i)\exp(2\mathbf{x}_i\boldsymbol{\beta})b_l(L_i)\delta_{i0}\text{var}(\psi_i)$$

$$+ \{\Lambda_0(L_i)(\delta_{i0} + \delta_{i3}) + \Lambda_0(R_i)(\delta_{i1} + \delta_{i2})\}\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\}\exp(2\mathbf{x}_i'\boldsymbol{\beta})\text{var}(\phi_i)$$

$$+ \frac{\delta_{i1}\text{cov}(Z_i, Z_{il})}{\gamma_l} + \frac{\delta_{i2}\text{cov}(W_i, W_{il})}{\gamma_l}$$

$$- \frac{\delta_{i2}\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\text{cov}(\phi_i, W_{il})}{\gamma_l} - \frac{\delta_{i1}\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\text{cov}(\phi_i, Z_{il})}{\gamma_l}$$

$$- \exp(\mathbf{x}_i'\boldsymbol{\beta})b_l(R_i)\{\text{cov}(Z_i, \phi_i)\delta_{i1} + \text{cov}(W_i, \phi_i)\delta_{i2}\}\Big]\mathbf{x}_i$$

$$\text{cov}(\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \gamma_l}, \frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \gamma_k}|\mathcal{D}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \Big( \frac{\text{cov}(U_{il}, U_{ik})\delta_{i0} + \text{cov}(W_{il}, W_{ik})\delta_{i2} + \text{cov}(Z_{il}, Z_{ik})\delta_{i1}}{\gamma_l\gamma_k}$$

$$+ \big[b_l(L_i)b_k(L_i)\text{var}(\psi_i)\delta_{i0} + \{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\}$$

$$\times \{b_k(L_i)(\delta_{i0} + \delta_{i3}) + b_k(R_i)(\delta_{i1} + \delta_{i2})\}\text{var}(\phi_i)\big]\exp(2\mathbf{x}_i'\boldsymbol{\beta})$$

$$- \frac{b_k(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\gamma_l}\{\text{cov}(Z_{il}, \phi_i)\delta_{i1} + \text{cov}(W_{il}, \phi_i)\delta_{i2}\}$$

$$- \frac{b_l(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\gamma_k}\{\text{cov}(Z_{ik}, \phi_i)\delta_{i1} + \text{cov}(W_{ik}, \phi_i)\delta_{i2}\}\Big)$$

All the conditional expectations, variances and covariances in the above equations are given

by:

$$E(\psi_i) \;=\; \frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1}I(\delta_{i0}=1)$$

$$\mathrm{var}(\psi_i) \;=\; E(\psi_i)^2$$

$$\mathrm{var}(U_{il}) \;=\; \frac{\gamma_l^{(d)}M_l(L_i)}{\Lambda_0'^{(d)}(L_i)}\{1-\frac{\gamma_l^{(d)}M_l(L_i)}{\Lambda_0'^{(d)}(L_i)}\}I(\delta_{i0}=1)$$

$$\mathrm{var}(\phi_i) \;=\; \begin{cases} \{\frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1}\}^2 & \delta_{i0}=1 \quad\text{or}\quad \delta_{i3}=1 \\[2ex] 1+\{\frac{1}{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1}\}^2 & \delta_{i1}=1 \\[2ex] \{\frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1}\}^2+\{\frac{1}{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1}\}^2 & \delta_{i2}=1 \end{cases}$$

When $\delta_{i1}=1$

$$E(Z_i) \;=\; \Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1$$

$$\mathrm{var}(Z_i) \;=\; \{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})+1\}\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})$$

$$\mathrm{var}(Z_{il}) \;=\; E(Z_i)\frac{\gamma_l^{(d)}b_l(R_i)}{\Lambda_0^{(d)}(R_i)}\{1-\frac{\gamma_l^{(d)}b_l(R_i)}{\Lambda_0^{(d)}(R_i)}\}+\mathrm{var}(Z_i)\{\frac{\gamma_l^{(d)}b_l(R_i)}{\Lambda_0^{(d)}(R_i)}\}^2$$

$$\mathrm{cov}(Z_i,Z_{il}) \;=\; \frac{b_l(R_i)\mathbf{x}_i'\boldsymbol{\beta}^{(d)}}{\Lambda_0^{(d)}(R_i)}\mathrm{var}(Z_i)$$

$$\mathrm{cov}(Z_{il},Z_{ik}) \;=\; \frac{\gamma_l^{(d)}\gamma_k^{(d)}b_l(R_i)b_k(R_i)}{\{\Lambda_0^{(d)}(R_i)\}^2}\{\mathrm{var}(Z_i)-E(Z_i)\}$$

$$\mathrm{cov}(\phi_i,Z_i) \;=\; \Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})$$

$$\mathrm{cov}(\phi_i,Z_{il}) \;=\; \gamma_l^{(d)}b_l(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})$$

When $\delta_{i2} = 1$

$$
\begin{aligned}
E(W_i) &= \frac{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1} \\
\mathrm{var}(W_i) &= E(W_i)\{E(W_i) - 1\} \\
\mathrm{var}(W_{il}) &= E(W_i)\{1 - \frac{\gamma_l^{(d)}[b_l(R_i) - b_l(L_i)]}{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)}\} + \mathrm{var}(W_i)[\frac{\gamma_l^{(d)}\{b_l(R_i) - b_l(L_i)\}}{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)}]^2 \\
\mathrm{cov}(W_i, W_{il}) &= \mathrm{var}(W_i)\frac{\gamma_l^{(d)}\{b_l(R_i) - b_l(L_i)\}}{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)} \\
\mathrm{cov}(W_{il}, W_{ik}) &= \frac{\gamma_l^{(d)}\gamma_k^{(d)}\{b_l(R_i) - b_l(L_i)\}\{b_k(R_i) - b_k(L_i)\}}{\{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)\}^2}\{\mathrm{var}(W_i) - E(W_i)\} \\
\mathrm{cov}(W_i, \phi_i) &= \frac{\{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)\}\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})}{\{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1\}^2} \\
\mathrm{cov}(W_{il}, \phi_i) &= \frac{\gamma_l^{(d)}\{b_l(R_i) - b_l(L_i)\}\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})}{\{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1\}^2}
\end{aligned}
$$

# Appendix B. Proof that $\hat{\boldsymbol{\theta}}$ is the unique global maximizer of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$

In each iteration of the EM algorithm, the M-step seeks the maximizer of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$, i.e., $\boldsymbol{\theta}^{(d+1)} = (\boldsymbol{\beta}^{(d+1)\prime}, \boldsymbol{\gamma}^{(d+1)\prime})' = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. This maximization can be accomplished through a two step procedure. Firstly, get $\boldsymbol{\gamma}^{(d+1)} = \arg\max_{\boldsymbol{\gamma}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. The maximizer $\boldsymbol{\gamma}^*(\boldsymbol{\beta})$ can be obtained as the solution to the system of equations $\partial\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\gamma} = 0$ and can be expressed in the form of equation ( 8). Because $\mathcal{Q}^2(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'$, the Hessian matrix of $\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$, is a diagonal matrix with the $l$th diagonal element takes the form in equation (10) in Appendix A. We can easily verify it is negative definite for all $\boldsymbol{\gamma}$ by noting $\gamma_1 > 0$, for $l = 1, ..., K$, so that $\mathcal{Q}^2(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})/\partial\gamma_l^2$ is strictly less than 0. In this way,

we can see for each value of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}^*(\boldsymbol{\beta})$ is the unique maximizer. Secondly, we need to get

$\boldsymbol{\beta}^{(d+1)} = \arg\max_{\boldsymbol{\beta}} \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$. Again, we can show $\boldsymbol{\beta}^{(d+1)}$ is the unique maximizer of

$\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ by showing the Hessian matrix of $\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ is negative definite for

all $\boldsymbol{\beta}$.

For notational convenience, let's first define

$$
\begin{aligned}
A_{il} &= E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2} \\
B_{il} &= E(\psi_i)\delta_{i0}b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \\
C_{il} &= E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}
\end{aligned}
$$

Based on the previous result, we have

$$
\begin{aligned}
\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n} &\Big\{ \mathbf{x}_i'\boldsymbol{\beta}\delta_{i0} - E(\psi_i)\delta_{i0} - E(\phi_i) \\
&+ \sum_{l=1}^{k} \Big[ A_{il} \log(\sum_{i=1}^{n} A_{il}) - A_{il} \log\{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})\} \\
&- B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta}) \frac{\sum_{i=1}^{n} A_{il}}{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})} + C_{il}\mathbf{x}_i'\boldsymbol{\beta} \Big] \Big\} + g(\boldsymbol{\theta}^d),
\end{aligned}
$$

and

$$
\frac{\partial \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \delta_{i0}\mathbf{x}_i - \sum_{i=1}^{n}\sum_{l=1}^{k} \frac{\sum_{i=1}^{n} A_{il}\{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\}}{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})} + \sum_{i=1}^{n}\sum_{l=1}^{k} C_{il}\mathbf{x}_i.
$$

Then if we denote $\sum_{i=1}^{n} A_{il} = A_l$ and $B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta}) = D_{il}$. The second order derivative of

$\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\beta}$ is

$$
\frac{\partial^2 \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\sum_{l=1}^{k} A_l \frac{(\sum_{i=1}^{n} D_{il}\mathbf{x}_i\mathbf{x}_i')(\sum_{i=1}^{n} D_{il}) - (D_{il}\mathbf{x}_i)(\sum_{i=1}^{n} D_{il}\mathbf{x}_i')}{(\sum_{i=1}^{n} D_{il})^2}.
$$

For any $\mathbf{z} \in \mathbb{R}^P$, where p is the dimension of $\boldsymbol{\beta}$. Let $H_l = \frac{A_l}{\{\sum_{i=1}^n D_{il}\}^2}$, then

$$\mathbf{z}'\frac{\partial^2 \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\mathbf{z} = -\sum_{l=1}^k H_l\Big[\{\sum_{i=1}^n D_{il}\mathbf{z}'\mathbf{x}_i\mathbf{x}_i'\mathbf{z}\}\{\sum_{i=1}^n D_{il}\} - \{\sum_{i=1}^n D_{il}\mathbf{z}'\mathbf{x}_i\}\{\sum_{i=1}^n D_{il}\mathbf{x}_i'\mathbf{z}\}\Big]$$

$$= -\sum_{l=1}^k H_l\Big[\{\sum_{i=1}^n D_{il}(\sum_{j=1}^p z_j x_{ij})^2\}\{\sum_{i=1}^n D_{il}\} - \{\sum_{i=1}^n D_{il}(\sum_{j=1}^p z_j x_{ij})\}^2\Big]$$

$$= -\sum_{l=1}^k H_l\Big[\sum_{i<h} D_{il}D_{hl}\{(\sum_{j=1}^p z_j x_{ij})^2 + (\sum_{j=1}^p z_j x_{hj})^2\}$$

$$- 2\sum_{i<h} D_{il}D_{hl}(\sum_{j=1}^p z_j x_{ij})(\sum_{j=1}^p z_j x_{hj})\Big]$$

$$= -\sum_{l=1}^k H_l\sum_{i<h} D_{il}D_{hl}\{(\sum_{j=1}^p z_j x_{ij}) - (\sum_{j=1}^p z_j x_{hj})\}^2$$

$$= -\sum_{l=1}^k H_l\sum_{i<h} D_{il}D_{hl}\{\mathbf{z}'(\mathbf{x}_j - \mathbf{x}_h)\}^2$$

For nonzero $\mathbf{z}$, $\mathbf{z}'\frac{\partial^2 \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\mathbf{z} = 0$ only when $\mathbf{z}'(\mathbf{x}_j - \mathbf{x}_h) = 0$ for all $i \neq h$. This only happens when all subjects have the same value for a particular covariate. In this situation the corresponding regression parameter and $\Lambda_0$ are not identifiable. In other words, as long as the model is identifiable, we have $\mathbf{z}'\frac{\partial^2 \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\mathbf{z} < 0$ for all $\mathbf{z} \in \mathbb{R}^P\{0\}$. This shows that $\frac{\partial^2 \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is negative definite. Thus, $\boldsymbol{\beta}^{(d+1)} = \arg\max_{\boldsymbol{\beta}} \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ is the unique maximizer. Consequently, $\boldsymbol{\theta}^{(d+1)}$ as determined by the two step process described in the EM algorithm is the unique maximizer of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$.