Journal of Advanced Statistics, Vol. 3, No. 4, December 2018
https://dx.doi.org/10.22606/jas.2018.34002

The Proportional Hazards Model with Linearly Time-dependent Covariates and Interval-censored Data

Qiqing Yu and Qinggang Diao

Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, USA
Email: qyu@math.binghamton.edu

Abstract The semi-parametric estimation under the proportional hazards (PH) model with a linearly time-dependent covariates and with interval-censored data has not been investigated before. The partial likelihood approach does not work and one has to use the generalized likelihood function (GLF). There is a challenge in this problem. The GLF must be in the form of the baseline hazard function, rather than the baseline survival function as in the PH model with time-independent covariates, and a feasible way to specify the hazard function is a piece-wise constant function. However, several naive ways do not yield a consistent estimator. We propose proper modifications of the GLF. Simulation results suggest that our method works. The generalization to other types of time-dependent covariates is also discussed.

Keywords: Cox's model, time-dependent covariates, modified likelihood function, semi-parametric MLE.

1 Introduction

We consider the semiparametric estimation problem under the proportional hazards (PH) model with a special continuous time-dependent covariates and with interval-censored data. In this paper, we assume that the survival time Y is continuous.

Interval-censored (IC) data are (L_i, R_i), $i = 1, ..., n$, where the true survival time $Y_i \in (L_i, R_i]$. The proportional hazards model (Cox (1972)) specifies that a covariate vector $\mathbf{z}$ has a proportional effect on the hazard function of Y. This model provides powerful means for fitting failure time observations to a distribution free model and for estimating the risk for failure associated with a covariate vector $\mathbf{z}$.

For a continuous random variable Y, denote its cumulative distribution function (cdf) by F_Y, its survival function by $S_Y(t) = 1 - F_Y(t)$, its density function (df) by $f_Y(t)$, and its hazard function by $h_Y(t) = \frac{f_Y(t)}{S_Y(t)}$. We say that $(\mathbf{z}, Y)$ follows a PH model or Cox's regression model if the hazard of $Y|\mathbf{z}$ is

$$h(t|\mathbf{z}) = h_o(t)e^{\beta\mathbf{z}}, \text{ for } t < \tau, \tag{1.1}$$

where $\beta\mathbf{z} = \beta'\mathbf{z}$, β' is the transpose of the vector β, $\tau = \sup\{t : h_o(t) > 0\}$, and h_o is the hazard function of $Y|(\mathbf{z} = 0)$. The PH model has been extended to time-dependent covariates, that is, $h(t|\mathbf{z}) = e^{\beta\mathbf{z}(t)}h_o(t),\ t < \tau$, where the covariate $\mathbf{z}(t)$ is also a function of the time t. In Cox and Oakes (1984 p.113), two examples of such extension are given. The PH models in those examples can be written as the form $h(t|u) = \exp(u\beta g(t))h_o(t)$, $t < \tau$, where u is a time-independent covariate, and $g(t)$ is a function of the time t. Two particular examples presented there are $g(t) = \mathbf{1}(t \geq a)$ and $g(t) = (t - a)\mathbf{1}(t \geq a)$, respectively, where $\mathbf{1}(A)$ is the indicator function of the event A. The first example is a special case of the piecewise PH model (PWPH model) and the second one is called the PH model with linearly time-dependent covariates (LDPH model). Therneau and Grambsch (2000) provide a computer program for computing the partial likelihood estimator under the PH model with such $z_i(t)$ and with right-censored data.

The TDCPH model has been commonly used for right-censored (RC) data (see, for instance, Zhou (2001), Leffondre *et al.* (2003), Platt *et al.* (2004), Zhang and Huang (2006), Stephan and Michael (2007), Masaaki and Masato (2009) and Leffondre *et al.* (2010)). The semi-parametric estimation with IC data under the PWPH model is studied by Wong *et al.* (2018). However, the semi-parametric estimation with IC data under the LDPH model has not been studied so far.

In this paper we consider the LDPH model, that is,

$$\mathbf{z}_i(t) = \mathbf{u}_i * (t-a)\mathbf{1}(t \geq a), \text{ where } \mathbf{u}_i \text{ is a time-independent covariate vector} \quad (1.2)$$

and a is a real number. This covariate is very typical and shares the light on how to estimate under the TDCPH model with IC data and with other types of covariates.

We shall assume that the hazard function h_o is unknown. Under such semi-parametric set-up, the typical estimation approach for right-censored data is the partial likelihood estimation. This approach simplifies the estimation procedure by estimating β alone without estimating the baseline hazard function h_o in the same time. Moreover, the properties of the partial likelihood MLE are quite similar to those of the generalized likelihood MLE. However, it is well known that this approach only works for right-censored data, but does not work for IC data (see, for example, Wong and Yu (2012)). Since we are dealing with IC data, we shall study the generalized likelihood estimation procedure in this paper.

For the covariate defined in (1.2), there are several theoretical issues to be settled. First, even if the parameter $\beta \in (-\infty, \infty)$, β may not be identifiable if the support set (of the observable random vector) is discrete (to be defined in Section 2). This is quite different from the case of the PH model with time-independent covariates, under which β is identifiable even if the support set contains only one point. It is also quite different from the case of the PH model with the time-dependent covariate of the form $z(t) = u\mathbf{1}(t \geq a)$, under which β is identifiable if the support set contains at least two points in $[a, \infty)$. The identifiability condition specifies the necessary condition under which a consistent estimator of β is possible and also gives a guideline for the set-up of simulation studies.

Secondly, unlike the PH model with time-independent covariates, the generalized likelihood function needs to be modified, as it must be in the form of hazard functions under the semi-parametric set-up in (1.2). Otherwise, there is no consistent estimator of β. If the covariate $\mathbf{u}_j$ in (1.2) takes on finitely many values, there is a naive non-parametric estimator of β, called the generalized maximum likelihood estimator (GMLE). However, it is not efficient. As explained in Section 3, several naive modifications on the generalized likelihood function do not lead to consistent estimators. We propose a proper modification of the generalized likelihood to get a semi-parametric MLE (SMLE) of β in this paper (see Remarks 2 and (3.3)). Our simulation studies suggest that the SMLE of β is consistent and is more efficient than the GMLE.

We study the identifiability condition in Section 2. We study how to modify the generalized likelihood function for deriving the SMLE in Section 3. We also introduce the algorithm for obtaining the GMLE if the covariate takes on finitely many values. Simulation results for comparing the SMLE and the GMLE are presented in Section 4. The generalization to other types of time-dependent covariates is discussed in Section 5.

2 The Models and the Identifiability Condition

Under interval censoring without covariates, the observable random vector is (L, R), where $L \leq Y \leq R$. The standard IC model that does not involve exact observations is the mixed case IC model (see Schick and Yu (2000)). Its simplest special case is the Case 2 model (see Groeneboom and Wellner (1992)):

(1) The random vector (U, V) and (Y, u) are independent,

(2) $(L, R) = (-\infty, U)\mathbf{1}(Y \leq U) + (U, V)\mathbf{1}(Y \in (U, V]) + (V, \infty)\mathbf{1}(Y > V)$.

The common IC model that involves exact observations is the double censorship model (see Turnbull (1976)). In this section we assume that $(Y, \mathbf{u})$ are from the PH model as in (1.1), with $\mathbf{z} = \mathbf{u} \cdot (t-a)\mathbf{1}(t \geq a)$, Y is continuous and is subject to interval censoring, and, $\mathbf{u}$ is a non-trivial time-independent random variable (vector). Interval-censored regression data are denoted by $(L_i, R_i, \mathbf{u}_i)$, where L_i and R_i are the endpoints of the interval I_i and $Y_i \in I_i$. The generalized likelihood function with IC data (L_i, R_i)'s is given by

$$\mathcal{L}_* = \textstyle\prod_{i=1}^n \mu_{S(\cdot|\cdot)}(I_i), \text{ where } \mu_S(I_i) = \begin{cases} S(L_i) - S(R_i) & \text{if } L_i < R_i \\ S(L_i-) - S(L_i) & \text{if } L_i = R_i. \end{cases} \quad (2.1)$$

2.1 General Forms of the Survival Functions

Since the generalized likelihood (2.1) depends on the survival function $S(t|\mathbf{z}(t))$, we shall discuss the general form of $S(t|\mathbf{z}(t))$.

Theorem 1. Assume that $S(t|\mathbf{z}(t))$ satisfies the PH model and is absolutely continuous, and $\mathbf{z}(t)$ is a time-dependent covariate. Then $S(t|\mathbf{z}(t)) = \exp(-\int_{-\infty}^{t} e^{\beta \mathbf{z}(x)} h_o(x)dx)$.

Corollary 1. Under the assumptions in Theorem 1, if $\mathbf{z}(t) = (t-a)\mathbf{u}\mathbf{1}(t \geq a)$, then

$$S(t|\mathbf{z}(t)) = \begin{cases} S_o(t) & \text{if } t \leq a \text{ or } \mathbf{u} = 0 \\ S_o(a)\exp(-\int_a^t e^{\beta \mathbf{u}(x-a)} h_o(x)dx) & \text{if } t > a \text{ and } \mathbf{u} \neq 0. \end{cases}$$

In our proposed estimation method, we shall make use of the special hazard function as follows.

$$\text{For } t \geq a,\ h_o(t) = \begin{cases} h_i & \text{if } t \in [a_i, b_i),\ i \in \{1, ..., k\} \\ 0 & otherwise, \end{cases}$$

where $a \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_k < b_k \leq a_{k+1} = \infty$. Then

$$S(t|\mathbf{z}(t)) = \begin{cases} S_o(t) & \text{if } t \leq a_1 \text{ or } \mathbf{u} = 0 \\ S_o(a)\exp(-\frac{e^{-a\mathbf{u}\beta}}{\mathbf{u}\beta}[\sum_{i=1}^{j-1} h_i[e^{\mathbf{u}\beta b_i} - e^{\mathbf{u}\beta a_i}]) & \\ +h_j[e^{\mathbf{u}\beta t} - e^{\mathbf{u}\beta a_j}]]) & \text{if } t \in [a_j, b_j) \text{ and } j \leq k \\ S_o(a)\exp(-\frac{e^{-a\mathbf{u}\beta}}{\mathbf{u}\beta}\sum_{i=1}^{j} h_i[e^{\mathbf{u}\beta b_i} - e^{\mathbf{u}\beta a_i}]) & \text{if } t \in [b_j, a_{j+1}) \text{ and } j \leq k. \end{cases}$$

In particular, if $b_i \approx a_i$ for all possible i,

$$\begin{aligned} S(t|\mathbf{z}(t)) \approx & S_o(a)\exp(-\sum_{i=1}^{j} h_i e^{(b_i-a)\mathbf{u}\beta}(b_i - a_i)) \\ = & S_o(a)\exp(-\sum_{i=1}^{j} h_i e^{(b_i-a)\mathbf{u}\beta}[(b_i - a_i) + o(b_i - a_i)]) \text{ if } \mathbf{u} \neq 0,\ t = b_j,\ b_i \approx a_i,\ j \leq k. \end{aligned}$$

2.2 Identifiability Issue

We assume that the $p \times 1$ covariate vector $\mathbf{u}$ takes at least p linearly independent values. In particular in this section, without loss of generality (WLOG), we assume $p = 1$ and $u \in \{0, 1\}$. Hereafter, by abuse of notations, we write $S(t|u) = S(t|\mathbf{z}(t))$ and $h(t|u) = h_o(t)\exp(\beta(t-a)u\mathbf{1}(t \geq a))$. Since h_o (or f_o) can differ on a set A satisfying $\int_A dS_o(t) = 0$, we define $f_o(t) = \begin{cases} -S_o'(t) & \text{if } S_o'(t) \text{ exists} \\ 0 & otherwise \end{cases}$ for identifiability of f_o and h_o. Let $\mathcal{S}_F$ be the support set of the random variable with the cdf F, in the sense that

$$x \in \mathcal{S}_F \text{ iff } F(x+\epsilon) - F(x-\epsilon) > 0,\ \forall\ \epsilon > 0.$$

It is worth mentioning that if the df f_X of a random variable X exists, then the cdf F_X and the hazard function h_X of X are equivalent in the sense that f_X yields F_X, F_X yields h_X and h_X yields f_X. However, one of f_X, S_X and h_X is given on a subset A of $\mathcal{S}_{F_X}$, then it is not always true that the other two functions are known on A. A counterexample can easily be found by setting $A = \{2\}$ and $X \sim \text{Exp}(\mu)$, the exponential distribution with mean μ.

Lemma 1. *The survival function $S(t|u)$ is identifiable if $t \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$.*

Proof. We say that the parameter θ is identifiable in the sense that the values of the df $f(\cdot;\theta)$ uniquely determines the parameter θ. Under the set-up in this section, the parameter is $S(t|u)$, where u is given and the df is the df of (L, R) for the given u, say $g(l, r|u; S(\cdot|\cdot))$. WLOG, we can assume the Case 2 model with the two follow-up times U and V, such that $(Y, u) \perp (U, V)$. Then $(\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}) \cap (-\infty, \infty) = \mathcal{S}_{F_U} \cup \mathcal{S}_{F_V}$.

$$g(l, r|u; S(\cdot|\cdot)) = \begin{cases} (1 - S(r|u))f_U(r) & \text{if } l = -\infty \text{ and } f_U(r) > 0 \\ (S(l|u) - S(r|u))f_{U,V}(l, r) & \text{if } f_{U,V}(l, r) > 0 \text{ and } -\infty < l < r < \infty \\ S(l|u)f_V(l) & \text{if } f_V(l) > 0 \text{ and } r = \infty \\ 0 & otherwise. \end{cases}$$

If $f_U(t) > 0$ then $S(t|u) = 1 - g(-\infty, t|u; S(\cdot|\cdot))/f_U(t)$ is uniquely determined by g. If $f_V(t) > 0$ then $S(t|u) = g(t, \infty|u; S(\cdot|\cdot))/f_V(t)$ is also uniquely determined by g. Moreover, if $t_o \in \mathcal{S}_{F_V} \cup \mathcal{S}_{F_U}$, $\exists$ a sequence of $t_j \to t_o$ and $f_U(t_j) > 0$ or $f_V(t_j) > 0$, then $S(t_o|u)$ is also uniquely determined by g, as the $S(t|u)$ is continuous and $S(t_j|u)$ is identifiable $\forall$ j. Thus given $g(\cdot, \cdot|u; S(\cdot|\cdot))$, $S(t|u)$ is identifiable at each point t in $\mathcal{S}_{F_U} \cup \mathcal{S}_{F_V}$. □

Lemma 2. *The parameter β is not identifiable if $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ is finite.*

Proof. It suffices to give a counterexample under the semi-parametric set-up. Consider the PH model $h(t|u) = \exp((t-a)u\beta\mathbf{1}(t \geq a))h_o(t)$, where $\beta = 1$, $h_o(x) = 1(x \geq 0)$ and $u \in \{0, 1\}$. WLOG, we can assume Case 2 interval censoring, with the two follow-up times U and V, and $(Y, u) \perp (U, V)$. Moreover, we can assume that the support set $\mathcal{S}_{F_U} \cup \mathcal{S}_{F_V} = \{1, 2, 3, ..., n\}$, where $a = 1$. Then $S(t|u)$ is identifiable at $t \in \{1, ..., n\}$, where $u \in \{0, 1\}$. Thus by Corollary 1, $S_o(a)$, $S_o(a)\exp(-\int_1^j h_o(x)dx)$ and $S_o(a)\exp(-\int_1^j e^{\beta(x-1)}h_o(x)dx)$ are identifiable, where $j = 1, ..., n$. Consequently,

$$\int_j^{j+1} h_o(x)dx = c_{2j-1} \text{ and } \int_j^{j+1} e^{\beta(x-1)}h_o(x)dx = c_{2j},\ j \in \{1, ..., n-1\}, \tag{2.2}$$

where $c_{2j-1} = 1$ and $c_{2j} = e^{(j-1)\beta}\frac{e^\beta - 1}{\beta} = e^{j-1}(e-1)$ are given, but h_o and β are parameters, though their true values are $h_o(x) = 1(x \geq 0)$ and $\beta = 1$.

It is obvious that $(1(x \geq 0), 1)$ is a solution to (h_o, β). We shall show that there is another solution to (h_o, β) (where $\beta > 1$) to (2.2). That is, for $j \in \{1, ..., n-1\}$, we shall define another function h_o, say $h_2(\cdot)$ on $(j, j+1]$ such that

$$\int_j^{j+1} h_2(x)dx = c_{2j-1} = 1 \text{ and } \int_j^{j+1} e^{\beta(x-1)}h_2(x)dx = c_{2j} = e^{j-1}(e-1). \tag{2.3}$$

For each $\gamma > 1$, setting $h_o = h_2 = \gamma c_{2j-1}\mathbf{1}(x \in (j, j+1/\gamma])$, where $x \in (j, j+1]$, then (2.3) yields

$$e^{\beta(j-1)}\frac{e^{\beta/\gamma} - 1}{\beta/\gamma} = c_{2j}/c_{2j-1} = e^{j-1}(e-1),\ \gamma > 1, \tag{2.4}$$

which specifies a different solution for β, that is, $\beta \neq 1$,

In particular, for $j = 1$, (2.4) becomes $\frac{e^{\beta/\gamma}-1}{\beta/\gamma} = e - 1,\ \gamma > 1$. Its solution is $\beta = \gamma$. The range of the solution β to (2.3) for $\gamma \geq 1$ and $j = 1$ is $B_1 = [1, \infty)$. For $j \in \{2, ..., n-1\}$, letting $\gamma \to \infty$, (2.4) yields $e^{\beta(j-1)} = e^{j-1}(e-1)$, or

$$\beta = \tfrac{j-1+\ln(e-1)}{j-1} > 1,\ j \in \{2, ..., n-1\}.$$

Since for each $j \in \{2, ..., n-1\}$, the solution β to (2.4) is continuous in $\gamma \in [1, \infty)$, the range of the solution β to (2.3) for $\gamma \geq 1$ is $B_j = [1, 1 + \frac{\ln(e-1)}{j-1})$. It is easy to show that the range B_j is decreasing in j. Here, notice that for each j, we only need to modify h_o or h_2 in the interval $(j, j+1]$.

Since n is finite, $\cap_{j=1}^{n-1} B_j = [1, 1 + \frac{\ln(e-1)}{n-2})$ and $(1, 1 + \frac{\ln(e-1)}{n-2}) \neq \emptyset$, $\exists$ $\beta_o \in (1, 1 + \frac{\ln(e-1)}{n-2})$ such that $\beta = \beta_o$ is a solution to $e^{\beta(j-2)}\frac{e^{\beta/\gamma_j}-1}{\beta/\gamma_j} = e^{j-1}(e-1)$ for some γ_j, where γ_j depends on $j \in \{1, 2, ..., n-1\}$. This is the second solution of β to (2.2). □

Theorem 2. An identifiability condition for β under the mixed case IC model is that $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ contains infinitely many points $\{t_j\}_{j \geq 1}$ with a limiting point, say $t_o = \lim_{j\to\infty} t_j$ in (a, ∞), provided that $S_o'(t_o) \neq 0$.

Proof. If $t_j \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$, $S(t_j|u)$ is identifiable for $u = 0$ or 1. Since S_o and $S(t|1)$ are continuous, and $t_j \to t_o$, $S_o(t_o)$ and $S(t_o|1)$ are also identifiable, and it leads to that $(S_o'(t_o), S'(t_o|1))$ is identifiable, as $S_o'(t_o) = \lim_{t_j \to t_o}\frac{S_o(t_j)-S_o(t_o)}{t_j - t_o}$, etc.. They further lead to that $(h_o(t_o), h(t_o|1))$ is identifiable, as $h(t_o|1) = -\frac{S'(t_o|1)}{S(t_o|1)}$ etc.. Consequently β can be identified by $h(t_o|1) = e^{(t_o - a)\beta}h_o(t_o)$, as a is known. □

In the proof of Lemma 2, since $(S_o(t), S(t|1))$ are only identifiable at finitely many points and one cannot derive $(S_o'(t), S'(t|1))$ and $(h_o(t), h(t|1))$ through these finitely many points. Moreover, in view of the proof of Theorem 2, if $(S_o(t), S(t|1))$ are only identifiable at all positive integers, or $S_o'(t_o) = 0$ for the t_o defined in Theorem 2, then β is likely non-identifiable. Moreover, the interval (a, ∞) in Theorem 2 cannot be replaced by $[a, \infty)$.

Lemma 3. *The function $S_o(a)$ is not identifiable if $a \notin \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$, even if β is known and $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ contains a nonempty open interval in $[a, \infty)$.*

Proof. It suffices to give a counterexample. Consider a Case 1 IC model with the censoring variable $C \sim U(2,3)$ and $h(t|u) = e^{u(t-1)\beta\mathbf{1}(t\geq 1)}h_o(t)$, $u \in \{0,1\}$. Then β is identifiable by Theorem 2. Let $\beta = 1$. If $h_o = 1(t > 0)$, then $S_o(a) = e^{-1}$, $S_o(2) = e^{-2}$, and $S(2|1) = e^{-1}e^{-(e-1)}$. Let

$$h_0^*(x) = \begin{cases} h_1 & \text{if } x \in (0,1] \\ h_2 & \text{if } x \in (1,1.5] \\ h_3 & \text{if } x \in (1.5,2] \\ arbitrary & otherwise. \end{cases}$$

The equation represents a hyperplane in R^3 passing through $(1,1,1)$. Thus there are many solutions near $(1,1,1)$ satisfying $(h_1,h_2,h_3) \geq 0$ and $h_1 \neq 1$. Thus the solution to $S_o(a) = e^{-h_1}$ is not unique. □

Remark 1. It is interesting to notice the following facts in the univariate covariate case.

1. If the covariate is time-independent then β is identifiable even if $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ contains only one point t at which $S(t) \in (0,1)$. In fact, if $t_o \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ then $S(t_o|u) = (S_o(t_o))^{e^{\beta u}}$ is identifiable for $u \in \{0,1\}$ and thus $S(t_o|0) = S_o(t_o)$ is identifiable. Since $\beta = \ln\frac{\ln S(t_o|1)}{\ln S_o(t_o)}$ and both $S_o(t_o)$ and $S(t_o|1)$ are known, it follows that β is identifiable.
2. If the covariate is $z = u\mathbf{1}(t \geq a)$, then β and $S_o(a)$ are identifiable even if $(\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}) \cap [a,\infty)$ contains only two points t's at which $S(t) \in (0,1)$. In fact, let $a \leq b < c$, where $b, c \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ and $1 > S_o(a) \geq S_o(b) \geq S_o(c) > 0$, then $(S_o(b),\ S_o(c),\ S(b|1),\ S(c|1))$ is identifiable. Moreover, $S(t|1) = (S_o(a))^{1-e^{\beta}}(S_o(t))^{e^{\beta}}$ if $t > a$. Since $\frac{S(b|1)}{S(c|1)} = (\frac{S_o(b)}{S_o(c)})^{e^{\beta}}$, β is identifiable. Since $S(b|1) = (S_o(a))^{1-e^{\beta}}(S_o(b))^{e^{\beta}}$, $S_o(a)$ is also identifiable.
3. However, if $z = (t-a)u\mathbf{1}(t \geq a)$, then neither β nor $S_o(a)$ is identifiable even if the set $(\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}) \cap [a,\infty)$ contains countably many points in (a,∞) (see the comment after Theorem 2). Moreover, $S_o(a)$ is not identifiable even if β is known and $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ contains a non-empty open interval in $[a,\infty)$.

These facts imply that if the random vector (L,R) only takes on finitely many values, there exist consistent estimators of β under the PH model with time-independent covariates but not under the situation considered in this paper. Also, in simulation studies, one can let the censoring vector have a finite discrete distribution under the PH model with time-independent covariates, but should not try the censoring vector which takes on finitely many values if the covariate is given by (1.2).

3 Semi-parametric Estimation

We shall propose our estimation method in this section.

3.1 Preliminary

For the PH model with time-independent covariates, the generalized likelihood function with IC data is given by $\mathcal{L}_* = \prod_{i=1}^n \mu_{S(\cdot|\cdot)}(I_i)$ as in (2.1). For the time-dependent covariates $\mathbf{z}(t) = \mathbf{u}\cdot(t-a)\mathbf{1}(t \geq a)$, such a definition would lead to $\mathcal{L}_* = 0$ in view of Theorem 1, as $\mu_{S(\cdot|\cdot)}(I_i) = 0$ if $L_i = R_i$. The first modification to $\mathcal{L}_*$ is

$$\mathcal{L}_o = \prod_{i=1}^n \mu_{S(\cdot|\cdot)}(I_i^*), \text{ where } \mu_S(I_i^*) = S(L_i^*) - S(R_i), \quad L_i^* = \begin{cases} L_i & \text{if } L_i < R_i \\ L_i - \epsilon_n & \text{if } L_i = R_i \end{cases} \quad (3.1)$$

and $\epsilon_n = \frac{1}{n} \wedge \min\{|x-y| :\ x \neq y,\ x,y \in \{a, L_1, ..., L_n, R_1, ..., R_n\}\}$. **Remark 2.** Hereafter, by abuse of notations, we write $L_i = L_i^*$, $i = 1$, ..., n. Moreover, I_i's are the modified observed intervals, instead of the original ones, that is, $I_i = (L_i^*, R_i]$.

An interval A is called an innermost interval (II) if it is an intersection of the observed intervals I_i's and if $A \cap I_i = A$ or $\emptyset$ for each I_i. Notice that each exact observation $[Y_k, Y_k]$ ($= [L_k, R_k]$) is an II, however, under our modification, it is changed to an interval $(Y_k - \epsilon_n, Y_k]$. It is well known (see Wong and Yu (2017)) that under the PH model with time-independent covariates in order to maximize $\mathcal{L}_o$, it suffices to put the weights of S_o to the IIs. Moreover, the weight to each II is uniquely determined, but not the distribution of the weight within the II. Let A_1, ..., A_m be all the II's induced by I_i's and let (v_j, w_j) be the pair of endpoints of A_j, where $w_0 = -\infty < w_1 < w_2 < \cdots < w_m \le \infty$. For each i, let $\xi_i = \mathbf{1}(R_i < w_m)$ and define l_i and r_i by $w_{r_i} \le R_i < w_{r_i+1}$ and $w_{l_i} \le L_i < w_{l_i+1}$. Then the likelihood function in (2.1) becomes

$$\mathcal{L}_o(\beta, S_o) = \prod_{R_i \le a \text{ or } \mathbf{u}_i = 0} (S_o(w_{l_i}) - S_o(w_{r_i})) \tag{3.2}$$
$$\cdot \prod_{L_i < a < R_i, \mathbf{u}_i \ne 0} \{S_o(w_{l_i}) - \xi_i S_o(a) \exp(-\sum_{w_j \in [a, R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x) dx)\} \cdot \prod_{L_i > a, \mathbf{u}_i \ne 0} S_o(a)$$
$$\cdot [\exp(-\sum_{w_j \in (a, L_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x) dx) - \xi_i \exp(-\sum_{w_j \in (a, R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x) dx)].$$

However, there is still a problem in this definition, as h_o is a function of x and needs to be properly defined on $[v_j, w_j]$ for all $j < m$ (note that $S_o(w_m) = 0$). The counterexample constructed in the proof of Lemma 2 can actually be modified to show that the definition of h_o on the IIs can change the value of the "SMLE" of β. There are two naive approaches:

A1. Let h_o be constant at each A_1, ..., A_{m-1}.

A2. Let h_o be two-piecewise-constant at each A_1, ..., A_{m-1}.

We shall explain in §3.3 that the previous two naive approaches do not lead to a consistent SMLE. We propose the third approach as follows, in addition to the modification mentioned in Remark 2 about L_i's and I_i's.

A3. First, let $h_o(x) = 0$ if $x \notin \cup_k (v_k, w_k]$, where $(v_1, w_1]$, ..., $(v_m, w_m]$ are all the II's. Moreover, notice that each $(v_k, w_k]$ will be contained by several modified observed intervals I_i's (see Remark 2) with J (≥ 1) distinct values of $\mathbf{u}_i$'s, where J depends on k. There are two types of (v_k, w_k): (1) $w_k - v_k \approx 0$, or $w_k \le a$, or $w_k - a \approx 0$ and $a \in (v_k, w_k]$; (2) otherwise. For $k < m$, define $h_o(x) = \begin{cases} \text{constant on } (v_k, w_k] & \text{if } (v_k, w_k) \text{ belongs to type (1)} \\ J\text{-piece-wise constant on } (v_k, w_k] & \text{if } (v_k, w_k) \text{ belongs to type (2)} \end{cases}$ (in particular, if (v_k, w_k) belongs to type (2), then

$$h_o(x) = \sum_{j=1}^{J} h_{kj} \mathbf{1}(x \in (v_{kj}, w_{kj}]) \text{ for } x \in (v_k, w_k], \text{ where } v_k = v_{k1},\ w_{k1} = v_{k2}, \tag{3.3}$$

$w_{k2} = v_{k3}$, ..., $w_{kJ} = w_k$ and $w_{kj} - v_{kj} = \begin{cases} \frac{w_k - v_k}{J} & \text{if } a \in (v_k, w_k],\ j \in \{1, ..., J\} \\ \frac{w_k - v_k}{J-1} & \text{if } a \notin (v_k, w_k],\ j \in \{2, ..., J\} \end{cases}$). If $k = m$, simply define $S_o(w_m) = 0$ (h_o can be arbitrary on $(v_m, w_m]$, provided that $h_o \ge 0$ and $\int_{v_m}^{w_m} h_o(x) dx = \infty$).

Remark 3. By abuse of notations, we let $(a_j, b_j]$ be the interval in which h_o is constant, as specified in (3.3). Then

$$h_o(x) = \sum_j h_j \mathbf{1}(x \in (a_j, b_j]), \text{ where } (a_j, b_j] \text{ may not be an II.} \tag{3.4}$$

In view of (3.4), $\mathcal{L}_o$ in (3.2) becomes

$$\begin{aligned}\mathcal{L}_o(\beta, S_o) = & \prod_{R_i \leq a, \ or \ \mathbf{u}_i = 0} \left\{ \exp(-\sum_{b_j \leq L_i} h_j[b_j - a_j])[1 - \exp(-\sum_{b_j \in (L_i, R_i]} h_j[b_j - a_j])]^{\xi_i} \right\} \\ & \cdot \prod_{L_i \geq a, \mathbf{u}_i \neq 0} \left\{ S_o(a) \exp(-\frac{e^{-a\mathbf{u}_i\beta}}{\mathbf{u}_i\beta} \sum_{b_j \in (a, L_i]} h_j[e^{\mathbf{u}_i\beta b_j} - e^{\mathbf{u}_i\beta a_j}]) \right. \\ & \left. \cdot [1 - \exp(-\frac{e^{-a\mathbf{u}_i\beta}}{\mathbf{u}_i\beta} \sum_{b_j \in (L_i, R_i]} h_j[e^{\mathbf{u}_i\beta b_j} - e^{\mathbf{u}_i\beta a_j}])] \right\} \\ & \cdot \prod_{L_i < a < R_i, \mathbf{u}_i \neq 0} \left\{ \exp(-\sum_{b_j \leq L_i} h_j[b_j - a_j]^{\xi_i}) \right. \\ & \left. \cdot [1 - \exp(-\sum_{b_j \in (L_i, a]} h_j(b_j - a_j) - \frac{e^{-a\mathbf{u}_i\beta}}{\mathbf{u}_i\beta} \sum_{b_j \in (L_i, R_i]} h_j[e^{\mathbf{u}_i\beta b_j} - e^{\mathbf{u}_i\beta a_j}])]^{\xi_i} \right\}.\end{aligned}$$

3.2 Definition of the Modified Likelihood Function

As explained later in (3.11), it is more convenient to replace $h_o(x)e^{\beta\mathbf{u}_i x}$ by a piecewise constant function, and to modify $\mathcal{L}_o$ as

$$\begin{aligned}\mathcal{L} = & \prod_{R_i \leq a, \ or \ \mathbf{u}_i = 0} \left\{ \exp(-\sum_{b_j \leq L_i} h_j[b_j - a_j])[1 - \exp(-\sum_{b_j \in (L_i, R_i]} h_j[b_j - a_j])]^{\xi_i} \right\} \\ & \cdot \prod_{L_i \geq a, \mathbf{u}_i \neq 0} \left\{ \exp(-\sum_{b_j \leq a} (b_j - a_j)h_j - \sum_{b_j \in (a, L_i]} h_j e^{(b_j - a)\mathbf{u}_i\beta}(b_j - a_j)) \right. \\ & \left. \cdot [1 - \exp(-\sum_{b_j \in (L_i, R_i]} h_j e^{\mathbf{u}_i\beta(b_j - a)}(b_j - a_j))]^{\xi_i} \right\} \prod_{L_i < a < R_i, \mathbf{u}_i \neq 0} \left\{ \exp(-\sum_{b_j \leq L_i} h_j[b_j - a_j]) \right. \\ & \left. [1 - \exp(-\sum_{b_j \in (L_i, a]} h_j[b_j - a_j] - \sum_{b_j \in (a, R_i]} h_j e^{(b_j - a)\mathbf{u}_i\beta}[b_j - a_j])]^{\xi_i} \right\}, \\ \ln\mathcal{L} = & \sum_{R_i \leq a, \ or \ \mathbf{u}_i = 0} \{(-\sum_{b_j \leq L_i} h_j[b_j - a_j]) + \xi_i \ln[1 - \mathcal{U}_i]\} \\ & + \sum_{L_i \geq a, \mathbf{u}_i \neq 0} \{-\sum_{b_j \leq a} h_j(b_j - a_j) - \sum_{b_j \in (a, L_i]} h_j e^{(b_j - a)\mathbf{u}_i\beta}(b_j - a_j) + \xi_i \ln[1 - \mathcal{W}_i]\} \\ & + \sum_{L_i < a < R_i, \mathbf{u}_i \neq 0} \{-\sum_{b_j \leq L_i} h_j[b_j - a_j] + \xi_i \ln[1 - \mathcal{V}_i]]\}, \\ & \text{where } \mathcal{U}_i = \exp(-\sum_{b_j \in (L_i, R_i]} h_j[b_j - a_j]), \ \mathcal{W}_i = \exp(-\sum_{b_j \in (L_i, R_i]} h_j e^{\mathbf{u}_i\beta(b_j - a)}(b_j - a_j)), \\ & \text{and } \mathcal{V}_i = \exp(-\sum_{b_j \in (L_i, a]} h_j[b_j - a_j] - \sum_{b_j \in (a, R_i]} h_j e^{(b_j - a)\mathbf{u}_i\beta}[b_j - a_j]).\end{aligned}$$

The SMLE maximizes $\ln\mathcal{L}$ over all h_j's and β. It is well known that the Newton Raphson method does not work (see Wong and Yu (2012) or Appendix II). We suggest to use the steepest decent method. Thus we derive the derivatives as follows.

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{L_i \ge a, \mathbf{u}_i \ne 0} \{- \sum_{b_j \in (a, L_i]} H_{ij} + \frac{\xi_i \mathcal{W}_i}{[1-\mathcal{W}_i]} \sum_{b_j \in (L_i, R_i]} H_{ij}\} + \sum_{L_i < a < R_i, \mathbf{u}_i \ne 0} \frac{\xi_i \mathcal{V}_i}{1-\mathcal{V}_i} \sum_{b_j \in (a, R_i]} H_{ij}$$

$$(H_{ij} = h_j e^{(b_j - a)\mathbf{u}_i \beta}(b_j - a_j)(b_j - a)\mathbf{u}_i)$$

$$= - \sum_{L_i \ge a, \mathbf{u}_i \ne 0} \{ \sum_{b_j \in (a, L_i]} H_{ij} + \xi_i (1 - \frac{1}{1-\mathcal{W}_i}) \sum_{b_j \in (L_i, R_i]} H_{ij}\}$$

$$- \sum_{L_i < a < R_i, \mathbf{u}_i \ne 0} \xi_i (1 - \frac{1}{1-\mathcal{V}_i}) \sum_{b_j \in (a, R_i]} H_{ij},$$

$$\frac{\partial \ln \mathcal{L}}{\partial h_k} = -(b_k - a_k) \cdot \{ \sum_{R_i \le a, \ or \ \mathbf{u}_i = 0} \{\mathbf{1}(b_k \le L_i) - \mathbf{1}(b_k \in (L_i, R_i]) \frac{\xi_i \mathcal{U}_i}{1-\mathcal{U}_i}\}$$

$$+ \sum_{L_i \ge a, \mathbf{u}_i \ne 0} \{\mathbf{1}(b_k \le a) + \mathbf{1}(b_k \in (a, L_i]) e^{(b_k - a)u_k \beta} - \xi_i \mathbf{1}(b_k \in (L_i, R_i]) e^{u_k \beta (b_k - a)} \frac{\mathcal{W}_i}{1-\mathcal{W}_i}\}$$

$$+ \sum_{L_i < a < R_i, \mathbf{u}_i \ne 0} \{\mathbf{1}(b_k \le L_i) - \xi_i [\mathbf{1}(b_k \in (L_i, a]) + \mathbf{1}(b_k \in (a, R_i]) e^{(b_k - a)u_k \beta}] \frac{\mathcal{V}_i}{1-\mathcal{V}_i}\}\}$$

$$= -(b_k - a_k) \cdot \{ \sum_{R_i \le a, \ or \ \mathbf{u}_i = 0} \{\mathbf{1}(b_k \le L_i) + \xi_i \mathbf{1}(b_k \in (L_i, R_i])(1 - \frac{1}{1-\mathcal{U}_i})\}$$

$$+ \sum_{L_i \ge a, \mathbf{u}_i \ne 0} \{\mathbf{1}(b_k \le a) + e^{(b_k - a)u_k \beta} \big[\mathbf{1}(b_k \in (a, L_i]) + \xi_i \mathbf{1}(b_k \in (L_i, R_i])(1 - \frac{1}{1-\mathcal{W}_i})\big]\}$$

$$+ \sum_{L_i < a < R_i, \mathbf{u}_i \ne 0} \{\mathbf{1}(b_k \le L_i) + \xi_i [\mathbf{1}(b_k \in (L_i, a]) + \mathbf{1}(b_k \in (a, R_i]) e^{(b_k - a)u_k \beta}](1 - \frac{1}{1-\mathcal{V}_i}\})\big\}.$$

To estimate the covariance matrix of $\hat{\beta}$, we need to compute

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \beta^2} = - \sum_{L_i < a < R_i, \mathbf{u}_i \ne 0} \xi_i \{ \sum_{b_j \in (a, R_i]} H'_{ij} - \frac{\sum_{b_j \in (a, R_i]} H'_{ij}}{1-\mathcal{V}_i} + \frac{\mathcal{V}_i (\sum_{b_j \in (a, R_i]} H_{ij})^2}{(1-\mathcal{V}_i)^2}\}$$

$$- \sum_{i: \ L_i \ge a, \mathbf{u}_i \ne 0} \{ \sum_{b_j \in (a, L_i]} H'_{ij} + \xi_i [\sum_{b_j \in (L_i, R_i]} H'_{ij} - \frac{\sum_{b_j \in (L_i, R_i]} H'_{ij}}{1-\mathcal{W}_i} + \frac{\mathcal{W}_i (\sum_{b_j \in (L_i, R_i]} H_{ij})^2}{(1-\mathcal{W}_i)^2}]\}$$

$$(\text{ where } H'_{ij} = h_j e^{(b_j - a)\mathbf{u}_i \beta}(b_j - a_j)((b_j - a)\mathbf{u}_i)^2).$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial h_k^2} = -(b_k - a_k)^2 \{ \sum_{R_i \le a \ or \ \mathbf{u}_i = 0} \xi_i \mathbf{1}(b_k \in (L_i, R_i]) \frac{\mathcal{U}_i}{(1-\mathcal{U}_i)^2}$$

$$+ \sum_{L_i \ge a, \mathbf{u}_i \ne 0} \xi_i \mathbf{1}(b_k \in (L_i, R_i]) \frac{e^{2(b_k - a)u_k \beta} \mathcal{W}_i}{(1-\mathcal{W}_i)^2}$$

$$+ \sum_{L_i < a < R_i, \mathbf{u}_i \ne 0} \xi_i [\mathbf{1}(b_k \in (L_i, a]) + \mathbf{1}(b_k \in (a, R_i]) e^{2(b_k - a)u_k \beta}] \frac{\mathcal{V}_i}{(1-\mathcal{V}_i)^2}\}.$$

Remark 4. As explained later in Example 3.1, the semi-parametric approach specified in (3.3) essentially estimates $\mu_F(II)$ by the GMLE with given β if the length of the II is not so small. There are cases that the length of an II may not tend to 0 such as an II containing $[0, 20]$ in a mammogram data set, as the age for the first mammogram is likely to be larger than 20 years. It is a practical issue how to interpret an II being small.

3.3 Justification of $\mathcal{L}_o$ and $\mathcal{L}$

We shall also explain why the third modifications of the likelihood works through Example 3.1, as well as why the others do not work.

Example 3.1. Consider a continuous nonnegative random variable Y that satisfies the model $h(t|u) = h_o(t)e^{tu\beta\mathbf{1}(t\geq 0)}$, where $u \sim bin(m, 0.5)$, the binomial distribution with m trials and the probability of success $p = 0.5$, h_o is continuous at $c = 0.69$ and $h_o(c) \neq 0$. Let Y be subject to a Case 2 IC model, where the censoring vector $(U, V) \equiv (c, c + c_n)$, $c_n = 4n^{-1/2}$ and n is the sample size. Then the observations are of the forms $(0, c, u)$, $(c, c + c_n, u)$, or $(c + c_n, \infty, u)$, $u \in \{0, 1, ..., m\}$. The observed intervals are either $(0, c]$, $(c, c + c_n]$, or $(c + c_n, \infty)$, and they are all IIs. J defined in A3 satisfies $j = m + 1$ if n is large enough. One may think that β is not identifiable as the support set is finite, in view of Lemma 2. Notice that (U, V) is really (U_n, V_n) and one can treat the support set being $A = \cup_n(\mathcal{S}_{F_{U_n}} \cup \mathcal{S}_{F_{V_n}})$, which is not finite and has a limiting point $c > a = 0$. For a fixed n, $S(t|u)$ is identifiable at c and $c + c_n$. We shall first let $m = 1$, *i.e.*, $u \sim bin(1, 0.5)$. Then

$$\ln(S(t|u)) = -\int_0^t h_o(x)e^{\beta ux}dx,\ u \in \{0, 1\},\ \ln(S(c|u)/S(c + c_n|u)) = \int_c^{c+c_n} h_o(x)e^{\beta ux}dx,\ u \in \{0, 1\},$$

$$\lim_{n\to\infty} \frac{1}{c_n}\ln(S(c|u)/S(c + c_n|u)) = h_o(c)e^{\beta uc},\ u \in \{0, 1\},\ \text{and}\ \tfrac{1}{c}\ln\tfrac{h_o(c)e^{1\beta c}}{h_o(c)e^{0\beta c}} = \beta. \tag{3.5}$$

Thus β is identifiable. Therefore, it is possible to construct consistent estimators of β. We shall discuss four estimators as follows.

a. A GMLE approach. If n is large enough, there are three innermost intervals $(0, c]$, $(c, c + c_n]$ and $(c+c_n, \infty)$. Let n_{1u}, n_{2u} and n_{3u} be the numbers of the three types of intervals with $u = 0$ or 1. Recall that the generalized likelihood function is $\mathcal{L}_o = (1 - S(c|0))^{n_{10}}(1 - S(c|1))^{n_{11}}(S(c|0) - S(c+c_n|0))^{n_{20}}(S(c|1) - S(c + c_n|1))^{n_{21}}(S(c + c_n|0))^{n_{30}}(S(c + c_n|1))^{n_{31}}$ and the GMLE of $S(\cdot|u)$, say $\hat{S}(\cdot|u)$, is given by

$$\hat{S}(c|u) = \frac{n_{2u} + n_{3u}}{n_{\cdot u}}\ \text{and}\ \hat{S}(c + c_n|u) = \frac{n_{3u}}{n_{\cdot u}},\ u \in \{0, 1\},\ \text{where}\ u_{\cdot u} = \sum_{j=1}^{3} n_{ju}. \tag{3.6}$$

Replacing $S(\cdot|u)$ in the last two equations in (3.5) by $\hat{S}(\cdot|u)$ yields an estimator of β:

$$\hat{\beta} = \frac{1}{c}\ln\frac{\ln(\hat{S}(c|1)/\hat{S}(c + c_n|1))}{\ln(\hat{S}(c|0)/\hat{S}(c + c_n|0))} = \frac{1}{c}\ln\frac{\frac{n_{21}}{n_{\cdot 1}-n_{21}}}{\frac{n_{20}}{n_{\cdot 0}-n_{20}}}. \tag{3.7}$$

Since it makes use of the GMLE $\hat{S}(\cdot|u)$, we can say that $\hat{\beta}$ is a GMLE of β. Simulation results in §4 suggest that the GMLE $\hat{\beta}$ is consistent (see Table 1).

b. An estimator due to the first naive approach. Let h_o be constant in each of first two IIs, say it equals h_1 and h_2, respectively (*i.e.*, $h_o(x) = h_1\mathbf{1}(x \in (0, c]) + h_2\mathbf{1}(x \in (c, c + c_n])$ if $x \in (0, c + c_n]$ and arbitrary otherwise). Then $\mathcal{L}_o$ in (3.1) satisfies $\mathcal{L}_o \approx \mathcal{L}_1$, where

$$\begin{aligned}\mathcal{L}_1 =&(1 - \exp(-ch_1))^{n_{10}}[\exp(-ch_1)(1 - \exp(-c_nh_2))]^{n_{20}}(\exp(-(ch_1 + c_nh_2)))^{n_{30}}\\ &\times (1 - \exp(-h_1\frac{e^{c\beta} - 1}{\beta}))^{n_{11}}[\exp(-h_1\frac{e^{c\beta} - 1}{\beta})(1 - \exp(-h_2c_ne^{c\beta}))]^{n_{21}} \quad (\text{as } c_n \approx 0)\\ &\times (\exp(-h_1\frac{e^{c\beta} - 1}{\beta} - h_2c_ne^{c\beta}))^{n_{31}}.\end{aligned}$$

Let $\breve{\beta}$ be the value of β that maximizes $\mathcal{L}_1$ just defined. Simulation results in Table 1 of Section 4 suggest that the estimator $\breve{\beta}$ is not consistent. Thus this approach does not work. In fact the limiting points of $\breve{\beta}$ should maximize the almost sure limit

$$\begin{aligned}\lim_{n\to\infty}\frac{\ln\mathcal{L}_1(h_1, h_2)}{n} =& p_{10}\ln(1 - \exp(-ch_1)) + p_{30}\ln\exp(-(ch_1)) \qquad (\text{note } c_n \to 0)\\ &+ p_{11}\ln(1 - \exp(-h_1\frac{e^{c\beta} - 1}{\beta})) + p_{31}\ln(\exp(-h_1\frac{e^{c\beta} - 1}{\beta})),\end{aligned}$$

where p_{iu}'s are defined in an obvious way, *e.g.*, $p_{1u} = P(Y \le c|u)$. Thus the (almost sure) limiting points of the estimator of (h_1, h_2, β) should satisfy

$$p_{30} = \exp(-\int_0^c h_o(x)dx) = \exp(-ch_1),$$
$$p_{31} = \exp(-\int_0^c h_o(x)e^{\beta x}dx) = \exp(-h_1\frac{e^{c\beta}-1}{\beta}),$$

etc.. If $\check{\beta}$ is consistent, then for each $(h_o(\cdot), \beta)$, there exists h_1 satisfying the two equations related to p_{30} and p_{31}. Let $\beta = 1$ and $h_o(x) = \mathbf{1}(x \notin (c/2, c))$, then p_{30} yields $h_1 = 0.5$ and p_{31} yields $\exp(-(e^{c/2}-1)) = \exp(-0.5(e^c-1))$, a contradiction, as $c = 0.69$. Thus the first naive approach does not lead to a consistent estimator of β.

Moreover, if $u \sim bin(m, 0.5)$ with $m \ge 2$ in Example 3.1, then the second naive approach does not lead to a consistent estimator of β neither. The argument is similar to the last paragraph. Here $J = m+1 > 2$, where J is defined in A3, the example motivates Eq. (3.3).

c. A consistent estimator due to the second naive approach when $u \sim bin(1, 0.5)$. Define

$$h_o(x) = \begin{cases} h_1 & \text{if } x \in (0, c/2] \\ h_2 & \text{if } x \in (c/2, c] \\ h_3 & \text{if } x \in (c, c+c_n/2] \\ h_4 & \text{if } x \in (c+c_n/2, c+c_n]) \\ arbitrary & \text{if } x > c+c_n. \end{cases}$$

There are 5 parameters in this approach and it can be shown that the solution to $(\beta, h_1, h_2, h_3, h_4)$ that maximizes $\mathcal{L}_o$ is not unique. In fact let $\hat{S}(\cdot|u)$ be the GMLE given in (3.6) (which is uniquely defined only at 0, c and $c+c_n$), then $\forall\ \beta \in (-\infty, \infty)$, there is a unique solution of (h_1, h_2, h_3, h_4) to the system of four linear equations in (3.8) and (3.9):

$$-\ln(\hat{S}(c|u)) = h_1\int_0^{c/2} e^{\beta ux}dx + h_2\int_{c/2}^{c} e^{\beta ux}dx,\ u \in \{0,1\}, \tag{3.8}$$

$$\ln(\hat{S}(c|u)/\hat{S}(c+c_n|u)) = h_3\int_c^{c+c_n/2} e^{\beta ux}dx + h_4\int_{c+c_n/2}^{c+c_n} e^{\beta ux}dx,\ u \in \{0,1\}, \tag{3.9}$$

though the solutions to h_i's may not be nonnegative. However, for each β in a neighbourhood of the GMLE $\hat{\beta}$ defined in part a, there exists a positive solution to $(h_1, ..., h_4)$ to the four equations (corresponding to $u \in \{0,1\}$) in (3.8) and (3.9) and thus, there exist many proper solutions to $(\beta, h_1, ..., h_4)$. Each of such $(\beta, h_1, h_2, h_3, h_4)$ maximizes $\mathcal{L}_o$. Consequently, this approach is not ideal, though it can lead to a consistent estimator of β.

d. An SMLE approach. In view of (3.9), if $c_n \approx 0$, we have

$$\ln(\hat{S}(c|u)/\hat{S}(c+c_n|u)) \approx e^{\beta uc}(h_3\int_c^{c+c_n/2} dx + h_4\int_{c+c_n/2}^{c+c_n} dx),\ u \in \{0,1\}, \tag{3.10}$$

$h_3 + h_4$ and β can be uniquely determined by the two equations in (3.10). Thus the parameters are really h_1, h_2, $h_3 + h_4$ and β and the degree of freedom in the second naive approach reduces to 4. If we revise (3.8) as follows,

$$-\ln(\hat{S}(c|u)) = h_1\int_0^{c/2} e^{\beta uc/2}dx + h_2\int_{c/2}^{c} e^{\beta uc}dx,\ u \in \{0,1\},$$

which uniquely determines h_1 and h_2, the four parameters, $S(c+c_n|u)$ and $S(c+c_n|u)$ for $u \in \{0,1\}$, can be reparametrized equivalently as β and h_o which is piecewise constant on $(0, c/2]$, $(c/2, c]$ and $(c, c+c_n]$ with values h_1, h_2 and h_3, respectively (notice that the new parameter is (β, h_1, h_2, h_3), with 4 degrees of freedom again). Recall the generalized likelihood

$$\begin{aligned}\mathcal{L}_o =&(1-\exp(-\int_0^c e^{\beta x}h_o(x)dx))^{n_{11}}[\exp(-\int_0^c e^{\beta x}h_o(x)dx)(1-\exp(-\int_c^{c+c_n} e^{\beta x}h_o(x)dx))]^{n_{21}}\\ &\times(\exp(-\int_0^{c+c_n} e^{\beta x}h_o(x)dx))^{n_{31}}(1-\exp(-0.5(ch_1+ch_2)))^{n_{10}}\\ &\times(\exp(-0.5(ch_1+ch_2))(1-\exp(-(c_nh_3))))^{n_{20}}(\exp(-(0.5ch_1+0.5ch_2+c_nh_3)))^{n_{30}}.\end{aligned}$$

Replacing $e^{\beta ux}h_o(x)$ by a piecewise constant, say
$h_1e^{\beta uc_2/}\mathbf{1}(x\in(0,c/2]+h_2e^{\beta uc}\mathbf{1}(x\in(c/2,c]+h_3e^{\beta uc+c_n/2}\mathbf{1}(x\in(c,c+c_n],$
$\mathcal{L}_o$ becomes

$$\begin{aligned}\mathcal{L} =&(1-\exp(-\int_0^{c/2} e^{\beta c/2}h_1dx-\int_{c/2}^c e^{\beta c}h_2dx))^{n_{11}} \qquad (3.11)\\ &\times[\exp(-\int_0^{c/2} e^{\beta c/2}h_1dx-\int_{c/2}^c e^{\beta c}h_2dx)\underbrace{(1-\exp(-\int_c^{c+c_n} e^{\beta c}h_3dx))}_{\approx h_3e^{\beta c}c_n}]^{n_{21}}\\ &\times(\exp(-\int_0^{c+c_n} e^{\beta c}h_3dx))^{n_{31}}(1-\exp(-0.5(ch_1+ch_2)))^{n_{10}}\\ &\times(\exp(-0.5(ch_1+ch_2))\underbrace{(1-\exp(-(c_nh_3)))}_{\approx h_3c_n})^{n_{20}}(\exp(-(0.5ch_1+0.5ch_2+c_nh_3)))^{n_{30}}.\end{aligned}$$

It can be verified that for (β,h_1,h_2,h_3), there is a different vector (β,h_1,h_2^*,h_3^*) such that $\mathcal{L}_o(\beta,h_1,h_2,h_3)=\mathcal{L}(\beta,h_1,h_2^*,h_3^*)$. Thus, one can take $\mathcal{L}$ as the semi-parametric likelihood and the SMLE of β (and h_3) can be solved through the GMLE $\hat{S}(\cdot|\cdot)$, which leads to $\tilde{\beta}=\frac{1}{c}\ln\frac{\frac{n_{21}}{n_{\cdot 1}-n_{21}}}{\frac{n_{20}}{n_{\cdot 0}-n_{20}}}$, which is the same as the GMLE of β in (3.7). Notice that replacing $e^{\beta ux}h_o(x)$ by a different piecewise constant such as
$h_1e^{\beta u0}\mathbf{1}(x\in(0,c/2]+h_2e^{\beta uc/2}\mathbf{1}(x\in(c/2,c]+h_3e^{\beta uc}\mathbf{1}(x\in(c,c+c_n]$ or
$h_1e^{\beta uc}\mathbf{1}(x\in(0,c/2]+h_2e^{\beta uc}\mathbf{1}(x\in(c/2,c]+h_3e^{\beta uc}\mathbf{1}(x\in(c,c+c_n],$
the limit of the SMLE will be the same. Notice that $J=2$ (see (3.3)) here, as $u\sim bin(1,0.5)$. This case also motivates Eq. (3.3) and $\mathcal{L}$ in §3.2. This concludes Example 3.1.

Remark 5. The discussion in Example 3.1 suggests that if $u\in\{0,1\}$, then one can get a GMLE of β based on $\mathcal{L}_*$ in (2.1) as follows.

1. First obtain the GMLE of S_o and $S(\cdot|1)$, based on the samples with $u_i=0$ and $u_i=1$, respectively, by the self-consistent algorithm (see Turnbull (1976)).

2. Let $\hat{\beta}=\frac{1}{m}\sum_{j=1}^m\frac{1}{b_{k_j}-a}\ln\frac{\ln(S(b_{k_j}|1)/S(a_{k_j}|1))}{\ln(S(b_{k_j}|0)/S(a_{k_j}|0))}$, where $(a_{k_j},b_{k_j}]$, $j=1,\ldots,m$, are all the II's that satisfy $b_{k_j}-b_{k_j-1}\approx 0$, $a_{k_j}>a$, and $S(b_{k_j}|u)>S(a_{k_j}|u)$ for $u\in\{0,1\}$.

If the covariate u takes on finitely many values, the approach is applicable after minor modifications. The drawback of this approach is that it does not work if there are very few ties in the covariate u_i's.

Remark 6. If $\tau<\infty$ and $\sup\mathcal{S}_{F_L}<\infty$, then $b_m<\infty$, where b_m is the largest among the right-end points of the innermost intervals. If one defines h_o to be piecewise constant on the II's, then such h_o does not lead to a proper survival function. However it is seen from Example 3.1 that there is no need to define h_o on the interval $(a_m,b_m]$ in the likelihood function $\mathcal{L}$, as long as $S_o(b_m)=0$. In the latter case, the likelihood will remain the same for any reasonable definition of h_o on $[a_m,b_m]$.

4 Simulation Studies

The next table presents the simulation results for the naive estimator $\check{\beta}$ and the SMLE $\hat{\beta}$ under the assumption in Example 3.1. The simulation results suggest that $\hat{\beta}$ is consistent, but not $\check{\beta}$.

We carried out simulation studies under the mixed case IC model. The mixed case IC model is implemented by $(L_i,R_i)=(W_{i-1},W_i)$ if $Y\in(W_{i-1},W_i]$, where $W_0=0$, $W_i=iV$, $i\geq 1$, V is from

Table 1. Simulation Results for the naive estimator $\check{\beta}$ and the SMLE $\hat{\beta}$

sample size	β	$\check{\beta}$	$sd_{\check{\beta}}$	$\hat{\beta}$	$sd_{\hat{\beta}}$
2000	0.5	0.284	0.086	0.564	0.299
20000	0.5	0.266	0.031	0.519	0.162
2000000	0.5	0.259	0.009	0.493	0.086

$U(0, 0.4)$. In addition if there is a fixed **II** $(b, c]$, then $L_i = a$ if $a \leq L_i \leq b$ and $R_i = b$ if $a \leq R_i \leq b$. We carried out simulation under the following setups.

(1) $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ is dense in (c, τ). $u \sim U[0, 1)$, in addition to $u \sim bin(1, 0.5)$.

(2) A population **II** is $(b, c]$, where $a < b < c$.

(3) A population **II** is $(b, c]$, where $b < a < c$.

We generated data with 5000 replications each for sample sizes n = 100, 200 and 400. The cut point is 0.5 and $\lambda = 0.2$. Table 2 displays the results. Our simulation study suggests that the SMLE $\hat{\beta}$ is consistent and the convergence rate is $n^{\frac{1}{2}}$.

Table 2. Simulation Results for the SMLE

sample size	β	h_o	a	u	(a, b)	$\hat{\beta}$	$SD_{\hat{\beta}}$
100	0.5	1	0.2	$bin(1, 0.5)$	(0.3,0.8)	0.512	0.238
200	0.5	1	0.2	$bin(1, 0.5)$	(0.3,0.8)	0.516	0.183
400	0.5	1	0.2	$bin(1, 0.5)$	(0.3,0.8)	0.489	0.111
100	0.5	1	0.2	$U(0, 1)$	(0.3,0.8)	0.542	0.423
200	0.5	1	0.2	$U(0, 1)$	(0.3,0.8)	0.491	0.298
400	0.5	1	0.2	$U(0, 1)$	(0.3,0.8)	0.469	0.196
100	1	$\exp(\lambda t)$	0.5	$U(0, 1)$	(0.4,0.6)	1.243	0.897
200	1	$\exp(\lambda t)$	0.5	$U(0, 1)$	(0.4,0.6)	1.132	0.612
400	1	$\exp(\lambda t)$	0.5	$U(0, 1)$	(0.4,0.6)	1.072	0.386
100	1	$\exp(\lambda t)$	0.5	$bin(1, 0.5)$	(0.4,0.6)	1.072	0.532
200	1	$\exp(\lambda t)$	0.5	$bin(1, 0.5)$	(0.4,0.6)	1.074	0.296
400	1	$\exp(\lambda t)$	0.5	$bin(1, 0.5)$	(0.4,0.6)	1.025	0.204
100	1	$\exp(\lambda t)$	0.5	$bin(1, 0.5)$	(0.6,0.8)	1.057	0.534
200	1	$\exp(\lambda t)$	0.5	$bin(1, 0.5)$	(0.6,0.8)	1.040	0.302
400	1	$\exp(\lambda t)$	0.5	$bin(1, 0.5)$	(0.6,0.8)	1.016	0.221
100	2	$\exp(\lambda t)$	0.5	$U(0, 1)$	(0.6,0.8)	2.178	0.651
200	2	$\exp(\lambda t)$	0.5	$U(0, 1)$	(0.6,0.8)	2.114	0.523
400	2	$\exp(\lambda t)$	0.5	$U(0, 1)$	(0.6,0.8)	2.058	0.359

5 Concluding Remark

Even though we only consider case that the covariate is of the form $\mathbf{z}(t) = (t-a)\mathbf{u}\mathbf{1}(t \geq a)$, the result can be generalized to the case of other time-dependent covariates, such as the form $\mathbf{z}(t) = \mathbf{u}g(t)$, where $\mathbf{u}$ is a covariate and $g(t)$ is a function. For instance, the two modification are trivially applicable to the case that $g(t) = \mathbf{1}(t \geq a)$, though (3.3) is not necessary. However, if $g(t) = (t-a)^2$ then the modification in (3.3) is necessary and works.

Acknowledgments. The authors thank the editor and a referee for their invaluable comments.

References

1. D.R. Cox, "Regression models and life tables," *Journal of Royal Statistics Society, B,* vol. 34, pp. 187-220, 1972.
2. D.R. Cox, and D. Oakes, "Analysis of Survival Data," Chapman & Hall, 1984.
3. K. Leffondre, M. Abrahamowicz, and J. Siemiatycki, "Evaluation of Cox's model and logistic regression for matched case-control data with time-dependent covariates: a simulation study," *Statistics in Medicine*, vol. 22, pp. 3781-3794, 2003.
4. K. Leffondre, W. Wynant, Z. Cao, et al. "A weighted Cox model for modeling time-dependent exposures in the analysis of case-control studies," *Statistics in Medicine*, vol. 29, pp. 839-850, 2010.
5. T. Masaaki, and S. Masato, "Analysis of survival data having time-dependent covariates," *IEEE Transaction on Neural Networks*, vol. 20, pp. 389-394, 2009.
6. R.W. Platt, K. S. Joseph3, C. V. Ananth, J. Grondines1, M. Abrahamowicz, and M. S. Kramer, "A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death," *American Journal of Epidemiology*, vol. 160, pp. 199-206, 2004.
7. A. Schick, and Q.Q. Yu, "Consistency of the GMLE with mixed case interval-censored data," *Scandinavian Journal of Statistics*, vol. 27 pp. 45-55, 2000.
8. L. Stephan, and S. Michael, "Parsimonious analysis of time-dependent effects in the Cox model," *Statistics in Medicine*, vol. 26, pp. 2686-2698, 2007.
9. T. Therneau, and P. Grambsch, "Modeling survival data : extending the Cox model," Springer. 2000.
10. B. W. Turnbull, "The empirical distribution function with arbitrary grouped, censored and truncated data," *Journal of Royal Statistics Society, B,* vol. 38, pp. 290-295, 1976.
11. G.Y.C. Wong, Q.G. Diao, and Q.Q. Yu, "Piecewise proportional hazards models with interval-censored data," *Journal of Statistical Computation and Simulation*, vol. 88, pp.140-155, 2018.
12. G. Y. C. Wong, M. P. Osborne, Q.G. Diao, and Q.Q. Yu, Piece-wise Cox Models With Right-Censored Data," *Communications in Statistics – Simulation and Computation*, vol. 46, pp. 7894-7908, 2017.
13. G.Y.C. Wong, and Q.Q. Yu, "Estimation under the Lehmann regression model with interval-censored data," *Communications in Statistics – Simulation and Computation*, 41 1489-1500, 2012.
14. H.M. Zhang, and C.F. Huang, "Nonparametric survival analysis on time-dependent covariate effects in case-cohort sampling design," *Statistica Sinica*, vol. 16, pp. 267-285. 2006.
15. M. Zhou, "Understanding the Cox regression models with time-change covariates," *American Statistian*, vol. 55, pp. 153-155, 2001.

