Accepted by "Journal of Statistical Computation and Simulation"

Technical Report of

"Piecewise Proportional Hazards Models With Interval-Censored Data"

By George Y. C. Wong,[a] Qinggang Diao[b] and Qiqing Yu[b] *

a Department of Integrative Medicine, Beth Israel Medical Center, NY 10003, USA

and

b Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, USA

Current version: 7/4/2017

AMS 1991 subject classification: Primary 62 G07; Secondary 62 J99.

Short Title: The PH model.

Key words and phrases: Cox's model, time-dependent covariates, semi-parametric MLE, identifiability.

Abstract: We consider the piecewise proportional hazards (PWPH) model with interval-censored (IC) relapse times under the distribution-free set-up. The partial likelihood approach is not applicable for IC data, and the generalized likelihood approach has not been studied in the literature. It turns out that under the PWPH model with IC data, the semi-parametric MLE (SMLE) of the covariate effect under the standard generalized likelihood may not be unique and may not be consistent. In fact, the parameter under the PWPH model with IC data is not identifiable unless the identifiability assumption is imposed. We propose a modification to the likelihood function so that its SMLE is unique. Under the identifiability assumption, our simulation study suggests that the SMLE is consistent. We apply the method to our cancer relapse time data and conclude that the bone marrow micrometastasis does not have a significant prognostic factor.

* Corresponding author's email address: qyu@math.binghamton.edu

1. Introduction. We consider the semi-parametric estimation problem under the piece-wise proportional hazards (PWPH) model, with interval-censored (IC) continuous survival time Y. The proportional hazards (PH) 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 vector of covariates. The PWPH model is a special PH model.

IC data consist of n time intervals with the end-points $L_i \le R_i$, $i = 1, ..., n$, where the true survival time Y_i falls inside the interval. Notice that (L_i, R_i) is called left-censored if $L_i = -\infty$, right-censored if $R_i = \infty$, strictly interval-censored if $0 < L_i < R_i < \infty$ and exact if $L_i = R_i$. For a random variable Y, denote its survival function by $S_Y(t) = P(Y > t)$, its density function by $f_Y(t)$, and its hazard function by $h_Y(t) = \frac{f_Y(t)}{S_Y(t-)}$. Given a covariate (vector) $\mathbf{Z}$ which does not depend on time Y, $(\mathbf{Z}, Y)$ follows a time-independent covariate PH (TICPH) model or Cox's regression model if the conditional hazard function $h_{Y|\mathbf{Z}}$ satisfies

$$h(t|\mathbf{z}) \ (= h_{Y|\mathbf{Z}}(t|\mathbf{z})) \ = h_o(t)e^{\beta'\mathbf{z}}, \text{ for } t < \tau, \tag{1.1}$$

where β' is the transpose of the $p \times 1$ vector β, $\tau = \sup\{t : h_o(t) > 0\}$, and h_o is an unknown baseline hazard function.

The Cox model has been extended to the time-dependent covariates proportional hazards (TDCPH) model. Cox and Oak (1984, p. 115) give a typical example of time dependent covariate in medical research, namely,

$$h(t|z) = e^{\beta z(t)}h_o(t), \ t < \tau, \text{ where } z = z(t) = \mathbf{1}_{(t \ge a)}, \tag{1.2}$$

and a is the admission time to a treatment for a patient. They also give another example of time-dependent covariate. Zhou (2001) formulates a PWPH model with k cut points:

$$h(t|\mathbf{z}) = \sum_{i=0}^{k} h_o(t)e^{\beta_i z_i}\mathbf{1}(t \in [a_i, a_{i+1})), \text{ where } a_0 = 0 < a_1 < \cdots < a_{k+1} = \infty, \tag{1.3}$$

$\mathbf{z} = (z_0, z_1, ..., z_k)'$ is a time-independent covariate vector. Model (1.2) is a special case of the PWPH model (1.3) with a single cut point at a. The TDCPH model has been commonly used for right-censored (RC) data (see, for instance, Therneau and Grambsch (2000), Leffondre *et al.* (2003), Platt *et al.* (2004), Zhang and Huang (2006), Stephan and Michael (2007), Masaaki and Masato. (2009), Leffondre *et al.* (2010)) and Wong *et al.* (2016). However, it has not been studied under the interval censoring.

Let $(\mathbf{Z}_1, Y_1)$, ..., $(\mathbf{Z}_n, Y_n)$ be a set of regression data. For instance, in our cancer research data set, Y_i is the relapse time of a cancer patient after surgery, $\mathbf{Z}_i$ is a vector with numerical or categorical coordinates, containing information about the age, tumor size at surgery, nodal number, bone marrow micrometastasis (bmm) or other information about the i-th patient. We are interested in the conditional survival function $S_{Y|\mathbf{Z}}$ instead of S_Y. In particular, we consider a problem of studying the relation between the covariate bmm with IC relapse time Y of a breast cancer patient after the surgery. The covariate bmm is a categorical variable taking two values, say 1 (bmm positive) and 0 (otherwise). Some medical doctors suspect that the bmm effect might depend on time T. To test this hypothesis, we consider a PWPH model as follows.

* **PWPH Model(1)**: Let $h(t|\mathbf{z}) = h_o(t)e^{\beta'\mathbf{z}}$ for $t < \tau$, where $\beta = (\beta_1, \beta_2)'$, $\mathbf{z} = (z_1, z_2)'$, $z_1 = \begin{cases} 1 & \text{if bmm}= 1 \text{ and } t < 4 \text{ years} \\ 0 & \text{ow,} \end{cases}$ and $z_2 = \begin{cases} 1 & \text{if bmm}= 1 \text{ and } t \geq 4 \text{ years} \\ 0 & \text{ow.} \end{cases}$
  Or more general $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')'$, $\mathbf{z}_1 = \mathbf{u}\mathbf{1}(t < a)$ and $\mathbf{z}_2 = \mathbf{v}\mathbf{1}(t \geq a)$, where a is a fixed constant, $\mathbf{u}$ and $\mathbf{v}$ are time-independent covariate vectors.

Under the TDCPH model with RC data, a common approach is the partial likelihood approach. However, this approach does not work even for interval censored data with time-independent covariates. Thus Finkelstein (1986) proposed the generalized likelihood function approach, making use of the generalized likelihood. Let $S_{Y|\mathbf{Z}}(t|\mathbf{z})$ or simply $S(t|\mathbf{z})$ be the conditional survival function corresponding to $h(t|\mathbf{z})$ in (1.1) and $S_o(t) = S_{Y|\mathbf{Z}}(t|0)$. Given IC data $(L_i, R_i, \mathbf{Z}_i)$ which may contain exact observations, the generalized likelihood is

$$\mathcal{L} = \mathcal{L}(\beta, S_o) = \prod_{i=1}^{n}[(S(L_i|\mathbf{Z}_i) - S(R_i|\mathbf{Z}_i))^{1-\delta_i}(S(L_i - |\mathbf{Z}_i) - S(R_i|\mathbf{Z}_i))^{\delta_i}], \tag{1.4}$$

where $\delta_i = \mathbf{1}(L_i = R_i)$. The semi-parametric maximum likelihood estimator (SMLE) of (β, S_o), denoted by $(\hat{\beta}, \hat{S}_o)$, maximizes $\mathcal{L}$ over all survival functions $S_o(\cdot)$ $(= S(\cdot|0))$ and all possible values of β. $\mathcal{L}$ defined in (1.4) is applicable to all IC data. In particular, it is applicable to the double censorship model (Turnbull (1976)), which assumes $Y \perp (\mathcal{U}, \mathcal{V})$

and $(L,R)=\begin{cases}(Y,Y) & \text{if } \mathcal{U}<Y\le\mathcal{V}\\(-\infty,\mathcal{U}) & \text{if } Y\le\mathcal{U}\\(\mathcal{V},\infty) & \text{if } Y>\mathcal{V},\end{cases}$ and it is possible that $\delta_i=1$. This is a model for a data set that contains left-censored, right-censored and exact observations. $\mathcal{L}$ is also applicable to the mixed case model (Schick and Yu (2000)) (with $\delta_i\equiv 0$), as it assumes $(L,R)=\begin{cases}(-\infty,C_1) & \text{if } Y\le C_1\\(C_{i-1},C_i) & \text{if } Y\in(C_{i-1},C_i],\ i\in\{2,...,N\}\\(C_N,\infty) & \text{if } Y>C_N,\end{cases}$ and $Y\perp\{N,C_1,C_2,...\}$, where C_i is the i-th follow-up time and N is a (random) number of follow-up times. If $P(N=m)=1$, then the mixed case model becomes the case m interval censorship model (see Groeneboom and Wellner (1992)). They are models for a data set that contains interval-censored observations, but not exact observations. Thus $\delta_i\equiv 0$.

The semi-parametric problem under the PWPH model with IC data has not been studied in the literature. Under PWPH model(1) with IC data, the parameter β is not identifiable unless further assumptions are imposed (see Example 2.1). Moreover, in general, the SMLE of β under the likelihood function (1.4) may not be unique (see Example 2.3). Both phenomena do not occur if the covariates are time-independent (see Wong and Yu (2012)). In this paper, we specify the identifiability condition for such problems. We propose an estimator of the regression parameter β based on the non-parametric MLE (NPMLE) and discrete data. We also study the estimation problem of deriving the SMLE. Under the identifiability condition, the simulation results suggest that both estimators of β are consistent under the mixed case IC model and the SMLE is more efficient. Moreover, the SMLE of $S_o(a)$ is consistent unless $\beta=0$, even if a is always censored. Under the assumption that the censoring distribution takes on finitely many values, the estimation problem becomes a multinomial distribution problem and the asymptotic properties of the SMLE can be easily established (see (Examples 2.2 and 2.3). We have completed the proof of the consistency of the SMLE in the general case and will be presented in a forthcoming paper due to the length of the proof.

The results in this paper can be extended to the general PWPH models. The main results are given in section 2. Algorithms for deriving the SMLE are discussed there. Simulation results are presented in section 3. Data analysis is in section 4. The detailed proofs of some lemma and examples are relegated to Appendix.

2. The main results. We study the estimation problem under the PWPH model assuming

Y is continuous in this section. For simplicity, we focus on the PWPH model with one cut point most of the time. Since we would make use of the generalized likelihood function of the form (1.4), we study the survival function under the model in §2.1 and investigate the identifiability condition in §2.2 and study how to modify the likelihood in (1.4) under PWPH model(1) in §2.3 and §2.4.

2.1. Survival functions. We study the general form of $S(t|\mathbf{z}(t))$ for various $\mathbf{z}(t)$ listed in section 1. Recall that $h_o(t) = h_{Y|\mathbf{Z}}(t|0) = h(t|0)$ and $S_o(t) = S_{Y|\mathbf{Z}}(t|0) = S(t|0)$.

Proposition 1. Assume that $S_{Y|\mathbf{Z}}(t|\mathbf{z}(t))$ satisfies the PH model. If $S_o(t)$ is absolutely continuous and non-negative. Then $S(t|\mathbf{z}(t)) = \exp(-\int_0^t e^{\beta'\mathbf{Z}(x)}h_o(x)dx)$, $t \geq 0$

Corollary 1. *If $S_o(t)$ is absolutely continuous and non-negative, then under model (1.3),*

$$S(t|\mathbf{z}(t)) = \begin{cases} (S_o(t))^{e^{\beta_0'\mathbf{Z}_0}} & \text{if } t \in (-\infty, a_1] \\ (S_o(a_1))^{e^{\beta_0'\mathbf{Z}_0}}(\frac{S_o(t)}{S_o(a_1)})^{e^{\beta_1'\mathbf{Z}_1}} & \text{if } t \in (a_1, a_2] \\ \prod_{j=1}^{i}(\frac{S_o(a_j)}{S_o(a_{j-1})})^{e^{\beta_{j-1}'\mathbf{Z}_{j-1}}}(\frac{S_o(t)}{S_o(a_i)})^{e^{\beta_i'\mathbf{Z}_i}} & \text{if } t \in (a_i, a_{i+1}],\ i \leq k. \end{cases} \quad (2.1)$$

Remark 1. The family of survival functions $\{S: \ S(t) = (S_o(t))^{e^{\beta'\mathbf{Z}}}, \ \beta, \mathbf{z} \in \mathcal{R}^p\}$, is called a Lehmann family or Lehmann model (Lehmann (1959)). If the covariates is not time-dependent and Y_i's are continuous, these two models are the same (see Yu (2006)). However, Corollary 1 indicates that the PH model and Lehmann model with time-dependent covariates are different even if S_o is absolutely continuous. In particular, if S_o is the survival function of the exponential distribution, that is, $S_o(t) = e^{-t}$, $t > 0$, then the PH model $h(t|\mathbf{z}(t)) = e^{\beta'\mathbf{Z}(t)}h_o(t)$ leads to (2.1), or $S(t|\mathbf{z}(t)) = (S_o(t))^{e^{\beta'\mathbf{Z}(t)}}(S_o(a))^{\mathbf{Z}(t)(1-e^{\beta'\mathbf{Z}(t)})}$, which does not lead to the Lehmann model $S(t|\mathbf{z}(t)) = (S_o(t))^{e^{\beta'\mathbf{Z}(t)}}$. Vice versa.

Corollary 2. *Let $S_o(t)$ be absolutely continuous. If the covariate vector $\mathbf{Z} = (W, Z_2, Z_3)'$, where $Z_2 = U\mathbf{1}(t < a)$, $Z_3(t) = V\mathbf{1}(t \geq a)$, and W, U and V are time-independent random variables, then* $S(t|\mathbf{z}(t)) = \begin{cases} (S_o(t))^{e^{(W,U,0)\beta}} & \text{if } t < a \\ (S_o(a))^{e^{(W,U,0)\beta}}(\frac{S_o(t)}{S_o(a)})^{e^{(W,0,V)\beta}} & \text{if } t \geq a, \end{cases}$ where $\beta = (\beta_1, \beta_2, \beta_3)'$.

The findings in this section indicate that the semi-parametric likelihood function has different forms depending on the covariates $\mathbf{z}(t)$ as well as the assumption on $S(t|\mathbf{z}(t))$. In fact, if $S(t|\mathbf{z}(t))$ is not continuous then the expressions are different.

2.2. Identifiability conditions. We shall first study the identifiability condition for the PWPH model(1), as it is related to whether we can get a consistent estimator. Without loss of generality (WLOG), we can assume that the covariates $\mathbf{u}$ and $\mathbf{v} \in \mathcal{R}^p$ and take at least p linearly independent values.

Given a random variable, say U, let $\mathcal{S}_{F_U}$ be the support set of F_U, in the sense that if $x \in \mathcal{S}_{F_U}$ then $F_U(x+\epsilon) - F_U(x-\epsilon) > 0 \ \forall \ \epsilon > 0$. Abusing notations, we write $h(t|z) = h_o(t)\exp(\beta'\mathbf{z}(t))$ and $f(t|z)$ is the conditional density function, etc..

Lemma 1. *Assume the PH model $h(t|\mathbf{u}) = h_o(t)e^{\beta'\mathbf{u}\mathbf{1}(t\geq a)}$, with the parameter (β, S_o) and without censoring. Then the parameter (β, S_o) is identifiable, provided that $\tau > a$ (see (1.1)).*

Proof. We shall show that given $f(\cdot|\cdot)$, (S_o, β) is identifiable. Note that given $f(\cdot|\cdot)$, $S(\cdot|\cdot)$ is identifiable. The covariate satisfies $\mathbf{Z}(t) = \mathbf{U}\mathbf{1}(t \geq a)$, where $\mathbf{U}$ is independent of t, and $\mathbf{U}$ and $\beta \in \mathcal{R}^p$. Moreover, $\mathbf{U}$ takes p linearly independent values. Let $\mathbf{u}$ be the realization of $\mathbf{U}$. Abusing notations, $S(t|\mathbf{u}) = \begin{cases} S_o(t) & \text{if } t < a \text{ or } \mathbf{u} = 0, \\ (S_o(a))^{1-e^{\beta'\mathbf{u}}}(S_o(t))^{e^{\beta'\mathbf{u}}} & \text{if } t \geq a \text{ and } \mathbf{u} \neq 0. \end{cases}$ Since $\mathbf{u}$ is independent of t, there exists $t_o > a$ such that $S(t_o|0) = S_o(t_o) > 0$ and $S(t_o|\mathbf{u}) = S_o(a)(\frac{S_o(t_o)}{S_o(a)})^{e^{\beta'\mathbf{u}}}$. $\ln S(t_o|\mathbf{u}) = \ln S_o(a) + e^{\beta'\mathbf{u}}\ln(\frac{S_o(t_o)}{S_o(a)})$ (which is a system of p linearly independent equations) and $S_o(t) = S(t|0)$ uniquely determined by $S(\cdot|z(\cdot))$. Thus (β, S_o) is identifiable. ▫

Lemma 2. *Assume $h(t|\mathbf{u}) = h_o(t)e^{\beta'\mathbf{u}\mathbf{1}(t\geq a)}$. Under the mixed case IC model and assuming that S_o is absolutely continuous, the parameter β is identifiable if*

$$\exists\ b, c \in (\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}) \cap [a, \infty) \ \textit{such that} \ S_o(b) > S_o(c) > 0, \tag{2.2}$$

where $\mathcal{S}_{F_L}$ and $\mathcal{S}_{F_R}$ are the support sets of the cdf's of L and R, respectively. The parameter $S_o(a)$ is identifiable if $\beta \neq 0$ in addition to assumption (2.2).

The proof is given in Appendix.

Remark 2. Lemma 2 can be extended to the PWPH model with k cut points, say $a_0 = 0 < a_1 < a_2 < \cdots < a_{k+1} = \infty$. Then for $i \in \{0, 1, ..., k+1\}$, β_i in (1.3) is identifiable if

$$\exists\ b_i, c_i \in (\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}) \cap [a_i, a_{i+1}], \text{ such that } S_o(b_i) > S_o(c_i) > 0;$$

moreover, $S_o(a_i)$ is identifiable if $\beta_i \neq 0$ in addition to the aforementioned assumption.

Notice that under the TICPH model, β is identifiable provided that there exists a point $t \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ such that $S_o(t) \in (0,1)$. Thus β_1 in model (1.3) is identifiable if there exists a point $t \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ and $t < a$ such that $S_o(t) \in (0, 1)$. If assumption (2.2) is violated, the parameter is not identifiable, as is the case in the next example.

Example 2.1. Suppose that $h_{Y|Z}(t|z) = e^{u\beta \mathbf{1}(t \geq 2)} h_o(t)$, $Z = U\mathbf{1}(t \geq 2)$ and U takes on two values. Under the mixed case interval censorship model, the parameter β is not identifiable if S_o is also an unknown parameter (see the proof in Appendix).

The non-identifiability example in the proof of Example 2.1 is mainly due to the fact that there is only one point in the set $(\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}) \cap [a, \infty)$. Notice that under right censoring with censoring variable C, and there are always two or more points in the set $\mathcal{S}_{F_C} \cap [a, \infty)$ if $\inf\{t : \ F_C(t) < 1\} > a$ and $h_o(t) > 0$ for some $t > a$. Thus the parameter is always identifiable under the right censorship model.

2.3. A non-parametric estimator of β. Under the PWPH models with the discrete covariates, there is a simple consistent estimator of β based on the NPMLE. WLOG we shall explain through PWPH Model(1)

$$h(t|u) = h_o(t) e^{\beta_1 u \mathbf{1}(t<a) + \beta_2 u \mathbf{1}(t \geq a)}$$

where the covariate u takes two discrete values, say 0 and 1. Then the simple consistent estimator of $\beta = (\beta_1, \beta_2)$ can be obtained as follows. Let $\breve{S}_o$ and $\breve{S}_1$ be the NPMLE's of $S_{Y|U}(\cdot|u)$ based on the observations with $u_i = 0$ and $u_i = 1$, respectively. Both estimators can be derived by the simple self-consistent algorithm (see Turnbull (1976)) as follows.

Recall that an intersection A of the observed intervals $I_i = \begin{cases} [L_i, R_i] & \text{if } L_i = R_i \\ (L_i, R_i] & \text{if } L_i < R_i \end{cases}$ is called an innermost interval (II) if $A \cap I_i = A$ or $\emptyset$ for each I_i (see Turnbull (1976)). The weight s_j assigned to A_j by the NPMLE satisfies $s_j = \lim_{h\to\infty} s_j^{(h)}$, where $s_j^{(0)} = 1/m$, and

$$s_j^{(h+1)} = \frac{1}{n} \sum_{i=1}^{n} \frac{s_j^{(h)} \mathbf{1}(A_j \subset I_i)}{\sum_{k=1}^{m} s_k^{(h)} \mathbf{1}(A_k \subset I_i)}, \quad h \geq 0, \ j = 1, ..., m.$$

To estimate β_1, let $q_1 < \cdots < q_m$ $(< a)$ be all the finite left-end points of the IIs induced by the observations with $u_i = 0$ and the IIs induced by the observations with $u_i = 1$. Since $\log S(t|1) = e^{\beta_1} \log S_o(t)$, $\beta_1 = \log \frac{\log S(t|1)}{\log S_o(t)}$. It is well known that the NPMLE $\breve{S}(\cdot|\cdot)$ and $\breve{S}_o$ are consistent (see Turnbull (1976)), then a consistent estimator is $\breve{\beta}_1 = \frac{1}{m} \sum_j \log \frac{\log \breve{S}(q_j|1)}{\log \breve{S}_o(q_j)}$ due to continuity.

To estimate β_2, let $(a <)\ g_1 < \cdots < g_m$ be all the finite left-end points of the IIs induced by the observations with $u_i = 0$ and the IIs induced by the observations with $u_i = 1$. Since $\log S(t|1) = \log(S_o(a))^{e^{\beta_1}} + e^{\beta_2} \log \frac{S_o(t)}{S_o(a)}$ for $t > a$, $\beta_2 = \log \frac{\log \frac{S(t|1)}{(S_o(a))^{e^{\beta_1}}}}{\log \frac{S_o(t)}{S_o(a)}}$. Then a consistent estimator is $\breve\beta_2 = \frac{1}{m}\sum_j \log \frac{\log \frac{\breve S(g_j|1)}{(\breve S_o(a))^{e^{\breve\beta_1}}}}{\log \frac{\breve S_o(g_j)}{\breve S_o(a)}}$.

Although the estimator $\breve\beta = (\breve\beta_1, \breve\beta_2)$ is consistent, but it is not efficient (see Example 2.2). It is an ideal initial value of β if one decides to use the SMLE, which can be obtained by an iterative algorithm.

2.4. The SMLE with IC data. The likelihood function with IC data is given by (1.4). In particular, $\mathcal{L} = \prod_{i=1}^n (S(L_i|\mathbf{z}_i) - S(R_i|\mathbf{z}_i))$. For the PH model, there are two differences between right censoring and interval censoring:

(a) One can show that the SMLE is unique and is consistent under the standard RC model but may not be so under the standard interval censorship model, unless further assumptions are imposed (due to identifiability).

(b) The SMLE of S_o assigns weight to the cut point a under the IC model (see the proof of Example 2.2) but not under the RC model (see Wong *et al.* (2016)).

Now we consider the PH model with IC data. Typically, let $\mathbf{z}_i = (z_{i1}, z_{i2}, z_{i3})'$, where $z_{i2} = u_i\mathbf{1}(t < a)$, $z_{i3}(t) = v_i\mathbf{1}(t \geq a)$, and $(z_{i1},\ u_i,\ v_i)$ is the i-the observation of the time-independent covariate vector. This is also the case in data analysis of our cancer data set. Notice that setting $w_i = u_i = 0$ leads to the covariate $z_i(t) = u_i\mathbf{1}(t \geq a)$.

Let A_1, ..., A_m be all the innermost intervals induced by I_i's. If the covariates are time-independent, it is well known that in order to maximize $\mathcal{L}$, it suffices to put the weights of S_o to the right-end points of the IIs. Let t_j's be the right-end point of the II's or a or $\pm\infty$ and $t_0 = -\infty < t_1 < \cdots < t_{i_a} = a < t_{i_a+1} < \cdots < t_m = \infty$. Write $S_j = S_o(t_j)$. For each i, let (l_i, r_i) satisfy $\begin{cases} t_{r_i} \leq R_i < t_{r_i+1} \text{ and } t_{l_i} \leq L_i < t_{l_i+1} & \text{if } L_i < R_i < \infty \\ t_{r_i} = t_m \text{ and } t_{l_i} \leq L_i < t_{l_i+1} & \text{if } L_i < R_i = \infty \\ t_{r_i} = R_i \text{ and } t_{l_i} = t_{r_i-1} & \text{if } R_i = L_i. \end{cases}$

Remark 3. We shall explain in the proof of Example 2.2 in Appendix that in order to maximize $\mathcal{L}$, it suffices to put the weights of S_o to t_j's. This is different from the case of the TICPH model or the case of the non-parametric likelihood. It is proved in Proposition 2 that if one puts the weights to the right-end points of the IIs only, then the resulting estimator is not consistent.

Since we shall carry out data analysis on the cancer research data making use of PWPH Model(1), by Corollary 2, the likelihood function of can be written as

$$\mathcal{L}(\beta, S_o) = \prod_{r_i \le i_a} ((S_o(t_{l_i}))^{e^{\beta' \mathbf{v}_{3i}}} - (S_o(t_{r_i}))^{e^{\beta' \mathbf{v}_{3i}}})$$
$$\cdot \prod_{l_i \le i_a < r_i} [(S_o(t_{l_i}))^{e^{\beta' \mathbf{v}_{3i}}} - (S_o(a))^{e^{\beta' \mathbf{v}_{3i}}} (\frac{S_o(t_{r_i})}{S_o(a)})^{e^{\beta' \mathbf{v}_{2i}}}]$$
$$\cdot \prod_{l_i > i_a} (S_o(a))^{e^{\beta' \mathbf{v}_{3i}}} [(\frac{S_o(t_{l_i})}{S_o(a)})^{e^{\beta' \mathbf{v}_{2i}}} - (\frac{S_o(t_{r_i})}{S_o(a)})^{e^{\beta' \mathbf{v}_{2i}}}],$$

where $\mathbf{v}_{2i} = \begin{pmatrix} w_i \\ 0 \\ v_i \end{pmatrix}$, $\mathbf{v}_{3i} = \begin{pmatrix} w_i \\ u_i \\ 0 \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$. Moreover,

$$\mathcal{L}(\beta, S_o) = \prod_{r_i < i_a} (S_o(t_{l_i})^{e^{\beta' \mathbf{v}_{3i}}} - S_o(t_{r_i})^{e^{\beta' \mathbf{v}_{3i}}}) \cdot \prod_{l_i > i_a} [S_o(t_{l_i})^{e^{\beta' \mathbf{v}_{2i}}} - S_o(t_{r_i})^{e^{\beta' \mathbf{v}_{2i}}}] \tag{2.3}$$
$$\cdot \prod_{l_i < i_a < r_i} [S_o(t_{l_i})^{e^{\beta' \mathbf{v}_{3i}}} - S_o(a)^{e^{\beta' \mathbf{v}_{3i}} - e^{\beta' \mathbf{v}_{2i}}} S_o(t_{r_i})^{e^{\beta' \mathbf{v}_{2i}}}] \cdot \prod_{l_i > i_a} S_o(a)^{e^{\beta' \mathbf{v}_{3i}} - e^{\beta' \mathbf{v}_{2i}}}$$
$$\cdot \prod_{r_i = i_a} (S_o(t_{l_i})^{e^{\beta' \mathbf{v}_{3i}}} - S_o(a)^{e^{\beta' \mathbf{v}_{3i}}}) \cdot \prod_{l_i = i_a} [S_o(a)^{e^{\beta' \mathbf{v}_{3i}}} - S_o(a)^{e^{\beta' \mathbf{v}_{3i}} - e^{\beta' \mathbf{v}_{2i}}} S_o(t_{r_i})^{e^{\beta' \mathbf{v}_{2i}}}].$$

The SMLE of (β, S_o) maximizes $\mathcal{L}(b, S)$ over all $b \in \mathcal{R}^p$ **and all survival functions** S. Thus it maximizes $\mathcal{L}(b, S)$ among all (b, S) that satisfies $\frac{\partial}{\partial \beta} \ln \mathcal{L}(\beta, S)|_{\beta = b} = 0$. In fact, for the complete data, if the sample size is not too large, one can use the Newton-Raphson method to solve the MLE numerically. If the data are interval censored, the method often does not work even with a time-independent covariate, such as the counterexample given in Appendix II. The main reason is that the maximum value of $\mathcal{L}(b, S)$ over all (b, S) without the restriction that S is a survival function is larger than the maximum value of $\mathcal{L}(b, S)$ over all (b, S) with the restriction. Thus we propose an algorithm as follows.

Assume that there are m distinct IIs, with the right-end points $t_1 < t_2 < \cdots < t_m$. Denote $S_i = S(t_i)$ with $S_m = 0$. Abusing notations, we identify S with a vector $(S_1, ..., S_m)$. Similarly, we identify $S^{(i)}$ with $(S_1^{(i)}, ..., S_m^{(i)})$.

Step 0. Let $b^{(0)} = 0$ be the initial estimate of β and the initial estimate of $S^{(0)}$ puts $1/m$ weight to each t_j's.

Step $i+1$ ($i \geq 0$). Let $b^{(i)}$ and $S^{(i)}$ be the updated values of b and S at Step i. Do b-step and S-step as follows.

* (b-**step**) With $S = S^{(i)}$ fixed, find a b so that the likelihood function $\mathcal{L}(S^{(i)}, \cdot)$ increases. Denote the up-dated estimate b by $b^{(i+1)}$. In particular, one can use the NR method to obtain the maximum point b of the likelihood function with the given $S = S^{(i)}$.

* (S-**step**) With $b = b^{(i+1)}$ fixed, search a non-increasing S so that the likelihood function $\mathcal{L}(\cdot, b^{(i+1)})$ is maximized (or increases). Denote the up-dated estimate S by $S^{(i+1)}$. In order to guarantee the up-dated S_o is nondecreasing, proceed as follows. Let $S^{(i+1),0} = S^{(i)}$. At Sub-step j ($j = 1$, ..., m), update $(S_1, ..., S_m)$ by $(S_1^{(i+1),j}, ...S_m^{(i+1),j})$, where $S_h^{(i+1),j} = S_{j,u_o}$ and $S_{j,u} = \begin{cases} \frac{S_h^{(i+1),j-1}+u}{1+u} & \text{if } h < j, \\ \frac{S_h^{(i+1),j-1}}{1+u} & \text{if } h \geq j, \end{cases}$ $h = 1, ..., m$, $u_o > 0$ is a number maximizing $\mathcal{L}(b^{(i+1)}, S_{\cdot,u})$ where $S_{\cdot,u} = (S_{1,u}, ..., S_{m,u})$.

  Note: If such u_o is difficult to choose, one may choose a u_o satisfying

$$\mathcal{L}(b^{(i+1)}, S^{(i+1),j}) > \mathcal{L}(b^{(i+1)}, S^{(i+1),j-1}). \tag{2.4}$$

  In particular, if $\frac{\partial}{\partial u}\ln\mathcal{L}(b^{(i+1)}, S_{\cdot,u})\big|_{u=0} > 0$, $u_o = c^k \frac{\partial}{\partial u}\ln\mathcal{L}(b^{(i+1)}, S_{\cdot,u})\big|_{u=0}$, where $S_{\cdot,u} = (S_{1,u}, ..., S_{m,u})$ and k is the smallest non-negative integer that is smaller than K_o such that (2.4) holds.

Stop at convergence.

Expressions of the partial derivatives can be found in Appendix I.

Remark 4. Let p_i be the weight on t_i, $p = (p_1, ..., p_m)$ and $p^{(i)}$ the updated value of p at the i-th step. Since $S(t_i) = p_{i+1} + \cdots + p_m$, the S-step can also be replaced by the p-step as follows.

* (p-**step**) With $b = b^{(i+1)}$ fixed, search a non-increasing S so that the likelihood function $\mathcal{L}(\cdot, b^{(i+1)})$ is maximized (or increases). Let $p^{(i+1),0} = p^{(i)}$. At Sub-step j ($j = 1$, ..., m), update $(p_1, ..., p_m)$ by $(p_1^{(i+1),j}, ...p_m^{(i+1),j})$, where $p_h^{(i+1),j} = p_{j,u_o}$ and $p_{j,u} = \begin{cases} \frac{p_h^{(i+1),j-1}+u}{1+u} & \text{if } h = j, \\ \frac{p_h^{(i+1),j-1}}{1+u} & \text{if } h \neq j, \end{cases}$ $h = 1, ..., m$, $u_o > 0$ is a number maximizing $L(b^{(i)}, S_{\cdot,u})$ where $S_{\cdot,u} = (S_{1,u}, ..., S_{m,u})$ and $S_{i,u} = p_{i+1,u} + \cdots + p_{m,u}$.

Moreover, the restriction $u > 0$ can be replaced by $u > -p_h^{(i+1),j-1}$.

Under the assumption that the follow-up times are discrete, the covariance matrix can be estimated by the inverse of the empirical Fisher information matrix.

A proof of the consistency of the SMLE and $\breve{\beta}$ under a simple assumption is as follows.

Example 2.2. Suppose that $h(t|u) = h_o(t)e^{\beta u \mathbf{1}(t\geq a)}$ and Y is continuous and is subject to the case m IC model, where $m \geq 3$, the follow-up time C_j's are constant and satisfy $C_1 < a < C_2$, and the covariate u takes at least two values, say 0 and 1. Then the SMLE and the estimator $\breve{\beta}$ based on the NPMLE are all consistent. Moreover, the SMLE is more efficient than $\breve{\beta}$.

One may think that Example 2.2 is trivial. It is interesting to see that if the case 3 model is replaced by the case 2 model, the conclusion is different as in the next example.

Example 2.3. Suppose that $h(t|u_i) = h_o(t)e^{\beta u_i \mathbf{1}(t\geq 2)}$ and Y is continuous, and is subject to the case 2 IC model where (C_1, C_2) only takes value $(1, 3)$, where $u_i = \mathbf{1}(i \leq n/2)$. It is proved in Appendix that the SMLE is not unique and not consistent.

Remark 5. The non-uniqueness and inconsistency of the SMLE in Example 2.3 is due to the non-identifiability of the parameter under the condition that the support set contains only one point $\geq a$, proved in Example 2.1. In general, the identifiability condition in Lemma 2 can be satisfied, as is the case in our breast cancer data. It is also worth noticing that under the TICPH model with the covariate taking at least two values, under the case 1 model with the follow-up time $C_1 = y$ w.p.1 and $S_o(y) \in (0, 1)$, the SMLE of β is consistent.

Proposition 2. Under the assumption in Example 2.2, if the estimator of $\tilde{\beta}$ is obtained by maximizing $\mathcal{L}$ with $S_o(a) = S_o(y_1)$, then $\tilde{\beta}$ is inconsistent.

3. Simulation Studies. We present simulation study results as follows.

sample size	β	S_o	a	$\overline{\hat{\beta}_1}$	$\overline{\hat{\beta}_2}$	$\text{SD}_{\hat{\beta}_1}$	$\text{SD}_{\hat{\beta}_2}$
400	$(0,2)$	Exp(1)	0.15	-0.011	2.101	0.292	0.289
800				0.001	2.046	0.207	0.183
1600				-0.002	2.020	0.149	0.143
400	$(-1,1)$	Exp(1)	0.15	-1.042	1.023	0.406	0.167
800				-1.034	1.008	0.275	0.123
1600				-0.969	1.008	0.210	0.085
400	$(0,2)$	U$(0,2)$	0.15	-0.019	2.069	0.412	0.275
800				-0.009	2.030	0.280	0.166
1600				0.009	2.009	0.199	0.110
400	$(-1,1)$	U$(0,2)$	0.15	-1.070	1.018	0.589	0.186
800				-1.040	1.011	0.409	0.126
1600				-1.011	1.005	0.265	0.090

In our studies, (Y,U) satisfies the model $h_{Y|U}(t|u_i) = e^{\beta_1 u_i 1(t\le a)+\beta_2 u_i 1(t>a)} h_o(t)$, where $a = 0.15$ and $U \sim bin(1, 0.4)$. The censorship model is the mixed case model with $P(N = 2) = 1$, which is also called the case 2 model. The follow-up times are C_1 and C_2, where $C_1 < C_2$. C_1 equals W with probability 0.2 and equals 0.15 with probability 0.8, where $W \sim \text{U}(0, 0.2)$. C_2 equals $0.25, 0.5$ and 1 with probabilities $0.25, 0.25$ and 0.5, respectively. S_o is either from Exp(1) (the exponential distribution with mean 1), or from U$(0,2)$ (the uniform distribution on the interval $(0,2)$). $\beta = (\beta_1, \beta_2)$ is either $(0,2)$ or $(-1,1)$. We generated data with 5000 replications each for sample sizes $n = 400$, 800 and 1600. Under both baseline distributions, the proportions of left-censoring and right-censoring are roughly 0.07 and 0.43, if $(\beta_1, \beta_2) = (0,2)$, and are 0.05 and 0.75 if $(\beta_1, \beta_2) = (-1,1)$. Our simulation results suggest that the SMLE of β is consistent and the convergence rate is $n^{1/2}$. We also carried out the simulation for sample sizes $n \le 200$, the sample standard deviations are very large, due to the divergence of the SMLE $\hat{\beta}$ for some samples. For instance, if $\beta = (-1,1)$ and S_o is Exp(1), the rates of divergence of $\hat{\beta}$ are 11.4%, 3% and 0.2% for $n = 50$, 100 and 200, respectively. It also happens for the partial likelihood estimator with right censored data if n is small (see Wong *et al.* (2016)). Thus we only present in the previous table the simulation results for $n \ge 400$. The proof of the consistency of the SMLE $(\hat{\beta}, \hat{S}_o)$ is under preparation and the proof of the convergent rate of $\hat{\beta}$ is our next project.

4. Data Analysis. Our data are obtained from 371 women with stages I - III unilateral invasive breast cancer surgically treated at Memorial Sloan-Kettering Cancer Center between 1985 and 2001, and up-dated more recently. We considered a problem of studying the relation between the bone marrow micrometastasis (bmm) and the relapse time Y of a breast cancer patient after the surgery. Y is interval censored. The covariate bmm is a categorical variable taking values 1 and 0. The median follow-up time is 7.4 years.

Suggested by a medical doctor, we considered PWPH model(1). We originally thought that the likelihood under the the PH model is still as in (1.4) and computed the "SMLE". The "estimates" $\tilde\beta_1 > 0$ and $\tilde\beta_2 < 0$, but only $\tilde\beta_1$ is significant. However, it turns out that the analysis is based on the Lehmann model, not the PH model. Moreover, there is a restriction of $\beta_1 \le 0$ under the Lehmann model (see Yu *et al.* (2013)).

We compute the value of β that maximizes the generalized likelihood (2.3). The SMLE based on our data satisfy

$\hat\beta_1 = 0.211420$ with a SE 0.188309 and Z-value$= 1.12$,

$\hat\beta_2 = 0.729921$ with a SE 0.483862 and Z-value$= 1.5$.

Both are not significant.

5. Appendix.

Proof of Lemma 2. WLOG, we can assume that the covariates u and β belong to $\mathcal{R}^1$, the censorship model is the case 2 model with two follow-up times C_1 and C_2, and conditional on u, X and (C_1, C_2) are independent.

Step (1). $\vdash$: $S(t|u)$ is identifiable at $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$.

It is easy to show that $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R} = ([c_1, c_2] \cap (\mathcal{S}_{F_{C_1}} \cup \mathcal{S}_{F_{C_2}})) \cup \{-\infty, \infty\}$. Under the censoring model, the density function is

$$g(l, r|u)) = \begin{cases} (1 - S(r|u))f_{C_1}(r) & \text{if } l = 0 \text{ and } f_{C_1}(r) > 0 \\ (S(l|u) - S(r|u))f_{C_1,C_2}(l, r) & \text{if } f_{C_1,C_2}(l, r) > 0 \text{ and } 0 < l < r < \infty \\ S(l|u)f_{C_2}(l) & \text{if } f_{C_2}(l) > 0 \text{ and } r = \infty. \end{cases}$$

If $t \in \mathcal{S}_{F_{C_1}}$, then we either have $f_{C_1}(t) > 0$ or $\exists$ a sequence of distinct points $t_j \in \mathcal{S}_{F_{C_1}}$ such that $t_j \to t$ and $f_{C_1}(t_j) > 0$. If $f_{C_1}(t) > 0$, then $S(t|u) = 1 - g(0, t|u)/f_{C_1}(t)$ is uniquely determined. If $f_{C_2}(t) > 0$ then $S(t|u) = g(t, \infty|u)/f_{C_2}(t)$ is uniquely determined. Moreover, if t_o is a limiting point in $\mathcal{S}_{F_{C_1}} \cup \mathcal{S}_{F_{C_2}}$, say $t_j \to t_o$, where $t_j \in \mathcal{S}_{F_{C_1}} \cup \mathcal{S}_{F_{C_2}}$, then $S(t_o|u)$ is uniquely determined as $S(t|u)$ is continuous. Thus the claim is proved.

Step (2) (Conclusion). Since u takes at least two values, WLOG, we can assume that it takes values 0 and 1. Notice that $S_o(t) = S(t|0)$. Thus $S_o(t)$ and $S(t|1)$ are identifiable at $\mathcal{S}_{F_{C_1}} \cup \mathcal{S}_{F_{C_2}}$. Now let $a \le b < c$, where $b, c \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$, then $S_o(b)$ and $S_o(c)$ and $S(b|1)$ and $S(c|1)$ are identifiable by Step (1). Moreover, $S(t|1) = (S_o(a))^{1-e^\beta}(S_o(t))^{e^\beta}$ if $t \ge a$. Since $S(b|1)/S(c|1) = (S_o(b)/S_o(c))^{e^\beta}$, β is identifiable if $S_o(b) > S_o(c) > 0$. Finally, since $S(b|1) = (S_o(a))^{1-e^\beta}(S_o(b))^{e^\beta}$, $S_o(a)$ is also identifiable if $\beta \ne 0$. ▫

Proof in Example 2.1. It suffices to give a counterexample to the identifiability of the parameters under the given assumptions. Let $U \sim bin(1, 0.5)$. Suppose that $S_o(t) \in (0,1)$ if $t \in (0,4)$. Moreover, assume the case 2 model, that is, the observable random vector is $(L,R) = (-\infty, C_1)\mathbf{1}(Y \le C_1) + (C_1, C_2)\mathbf{1}(Y \in (C_1, C_2]) + (C_2, \infty)\mathbf{1}(Y > C_2)$, where the random vector (C_1, C_2) and (U, Y) are independent.

Let the censoring vector $(C_1, C_2) \equiv (1,3)$ and S_o be absolutely continuous, where $S_o(1) > S_o(2) > S_o(3) > S_o(4) = 0$. Let $s_0 = \mathbf{1}(Y \le 1|U=0) + \mathbf{1}(Y \le 1|U=1)$, $s_1 = \mathbf{1}(Y \in (1,3]|U=1)$, $s_2 = \mathbf{1}(Y \in (1,3]|U=0)$, $s_3 = \mathbf{1}(Y > 3|U=1)$, and $s_4 = \mathbf{1}(Y > 3|U=0)$. Let $p_1 = F_o(1)$, $p_2 = F_o(3) - F_o(2)$, $p_3 = S_o(3)$, and $p_4 = F_o(2) - F_o(1)$. The density of $(s_0, s_1, ..., s_4)$ is

$$f = (p_1)^{s_0}(1 - p_1 - (p_2+p_3)^{1-e^\beta} p_3^{e^\beta})^{s_1}(1-p_1-p_3)^{s_2}((p_2+p_3)^{1-e^\beta} p_3^{e^\beta})^{s_3} p_3^{s_4}.$$

For given $(p_1, p_2, p_3, \beta) = (p_1^*, p_2^*, p_3^*, \beta^*)$, let $\gamma^* = (p_2^* + p_3^*)^{1-e^{\beta^*}} p_3^{*e^{\beta^*}}$, then f remains the same if $(p_1, p_2, p_3, \beta) = (p_1^*, p_2, p_3^*, \beta)$, where (p_2, β) satisfies $(p_2 + p_3^*)^{1-e^\beta} p_3^{*e^\beta} = \gamma^*$. The latter equation yields

$$\beta = \ln \frac{\ln \frac{\gamma^*}{p_2+p_3^*}}{\ln \frac{p_3^*}{p_2+p_3^*}} \tag{5.1}$$

where $p_2 \in (0, 1-p_1^*-p_3^*]$. Thus the β is not uniquely determined if $p_1^* + p_3^* < 1$. For instance, let $(p_1^*, p_2^*, p_3^*, \beta^*) = (1/3, 1/8, 1/8, -1.1)$, then $\gamma^*/(p_2^* + p_3^*) \approx 0.12$. Thus $\beta = \beta(p_2)$ in (5.1) is well defined for p_2 in a neighborhood of 1/8 (actually, for p_2 in $(0, 1-1/3-1/8]$). Hence, the parameter β is not identifiable. ▫

Proof of Example 2.2. WLOG, we can assume $m = 3$. That is, we can assume that the censorship model is a case 3 model with follow-up times at C_1, C_2, C_3. WLOG, we can further assume $(C_1, a, C_2, C_3) = (1, 2, 3, 4)$ and there are n IC observations under the model $h(t|u_i) = e^{u_i\beta\mathbf{1}(t\geq 2)}h_o(t)$, where $u \sim bin(1, 0.5)$. (L_i, R_i) is of the forms $(-\infty, 1)$, $(1, 3)$, $(3, 4)$ and $(4, \infty)$. Let $n_0 = \sum_{i\leq n}\mathbf{1}(1 = R_i)$, $n_1 = \sum_i \mathbf{1}((L_i, R_i, u_i) = (1, 3, 0))$, $n_2 = \sum_i \mathbf{1}((L_i, R_i, u_i) = (1, 3, 1))$, $n_3 = \sum_i \mathbf{1}((L_i, R_i, u_i) = (3, 4, 0))$, $n_4 = \sum_i \mathbf{1}((L_i, R_i, u_i) = (3, 4, 1))$, $n_5 = \sum_i \mathbf{1}(L_i = 4, u_i = 0)$, and $n_6 = \sum_i \mathbf{1}(L_i = 4, u_i = 1)$. Let $S_i = S_o(i)$. By Corollary 1, the generalized likelihood in (1.4) becomes

$$\begin{aligned}
\mathcal{L} =&(1-S_1)^{n_0}(S_1-S_3)^{n_1}(S_1-S_2(\frac{S_3}{S_2})^{e^\beta})^{n_2}\\
&(S_3-S_4)^{n_3}(S_2(\frac{S_3}{S_2})^{e^\beta}-S_2(\frac{S_4}{S_2})^{e^\beta})^{n_4}S_4^{n_5}(S_2(\frac{S_4}{S_2})^{e^\beta})^{n_6}\\
=&(1-S_1)^{n_0}(S_1-S_3)^{n_1}(S_1-S_2^{1-e^\beta}S_3^{e^\beta})^{n_2}S_2^{(1-e^\beta)(n_4+n_6)}(S_3-S_4)^{n_3}(S_3^{e^\beta}-S_4^{e^\beta})^{n_4}S_4^{n_5+n_6e^\beta}.
\end{aligned}$$

Let $s_j = S_j/S_1$ and $\mathbf{s} = (s_2, s_3, s_4)$ then

$\mathcal{L} = (1-S_1)^{n_0}S_1^{n-n_0}(1-s_3)^{n_1}(1-s_2^{1-e^\beta}s_3^{e^\beta})^{n_2}s_2^{(1-e^\beta)(n_4+n_6)}(s_3-s_4)^{n_3}(s_3^{e^\beta}-s_4^{e^\beta})^{n_4}s_4^{n_5+n_6e^\beta}$.

$$\frac{\partial \ln\mathcal{L}}{\partial(\mathbf{s},\beta)} = \begin{pmatrix} \frac{-n_2(1-e^\beta)s_2^{-e^\beta}s_3^{e^\beta}}{1-s_2^{1-e^\beta}s_3^{e^\beta}} + \frac{(1-e^\beta)(n_4+n_6)}{s_2} \\ -\frac{n_1}{1-s_3} - \frac{n_2e^\beta s_2^{1-e^\beta}s_3^{e^\beta-1}}{1-s_2^{1-e^\beta}s_3^{e^\beta}} + \frac{n_3}{s_3-s_4} + \frac{n_4e^\beta s_3^{e^\beta-1}}{s_3^\beta-s_4^{e^\beta}} \\ -\frac{n_3}{s_3-s_4} - \frac{n_4e^\beta s_3^{e^\beta-1}}{s_3^\beta-s_4^{e^\beta}} + \frac{n_5+n_6e^\beta}{s_4} \end{pmatrix}.$$

There are 5 parameters in this example, that is, $(S_1, S_2, S_3, S_4, \beta)$, where $S_i = S_o(i)$. Let $s_i = S_o(i)/S_o(a)$, and $\check{s}_i = S_i/S_1$, where $a = 2$. Then $\mathcal{L}$ can be written as

$$\mathcal{L} =[(1-S_1)^{n_0}S_1^{n-n_0}]\cdot[(1-\check{s}_3)^{n_1}(\check{s}_3-\check{s}_4)^{n_3}\check{s}_4^{n_5}]\cdot[(1-\check{s}_2s_3^{e^\beta})^{n_2}(\check{s}_2s_3^{e^\beta}-\check{s}_2s_4^{e^\beta})^{n_4}(\check{s}_2s_4^{e^\beta})^{n_6}].$$

Notice that $\mathcal{L}$ is the product of three factors, each has its own independent parameters, say $\{S_1\}$, $\{\check{s}_3, \check{s}_4\}$, and $\{\tilde{s}_3, \tilde{s}_4\}$, where $\tilde{s}_3 = \check{s}_2s_3$, $\tilde{s}_4 = \check{s}_2s_4$. Thus the SMLE satisfies

$$\begin{aligned}
&1-S_1 = n_0/n,\\
&\tfrac{S_1-S_3}{S_1} = \tfrac{n_1}{n_1+n_3+n_5},\\
&\tfrac{S_4}{S_1} = \tfrac{n_5}{n_1+n_3+n_5},\\
&\tfrac{S_1-S_2^{1-e^\beta}S_3^{e^\beta}}{S_1} = \tfrac{n_2}{n_2+n_4+n_6},\\
&\tfrac{S_2^{1-e^\beta}S_4^{e^\beta}}{S_1} = \tfrac{n_6}{n_2+n_4+n_6}.
\end{aligned}$$

The previous equations leads to

$\hat{S}_1 = 1 - n_0/n,$

$\hat{S}_3 = \hat{S}_1(1 - \frac{n_1}{n_1+n_3+n_5}),$

$\hat{S}_4 = \hat{S}_1 \frac{n_5}{n_1+n_3+n_5},$

$\hat{S}_2^{1-e^{\hat{\beta}}} \hat{S}_3^{e^{\hat{\beta}}} = \hat{S}_1 \frac{n_4+n_6}{n_2+n_4+n_6},$

$\hat{S}_2^{1-e^{\beta}} \hat{S}_4^{e^{\beta}} = \hat{S}_1 \frac{n_6}{n_2+n_4+n_6}.$

The last two equations lead to

$(\frac{\hat{S}_3}{\hat{S}_4})^{e^{\hat{\beta}}} = \frac{n_4+n_6}{n_6}.$

which further yields

$\hat{\beta} = \log\left\{\frac{\log(\frac{\hat{S}_3}{\hat{S}_4})}{\log(\frac{n_4+n_6}{n_6})}\right\} = \log\left\{\frac{\log(\frac{n_3+n_5}{n_5})}{\log(\frac{n_4+n_6}{n_6})}\right\}.$

Finally, if $\hat{\beta} \neq 0$ then

$$\begin{aligned}
\hat{S}_2 =& (\frac{\hat{S}_1 n_6}{\hat{S}_4^{e^{\hat{\beta}}}(n_2+n_4+n_6)})^{\frac{1}{1-e^{\hat{\beta}}}} \\
=& (\frac{\hat{S}_1 (n_4+n_6)}{\hat{S}_3^{e^{\hat{\beta}}}(n_2+n_4+n_6)})^{\frac{1}{1-e^{\hat{\beta}}}} \\
=& \left(\frac{(1-\frac{n_0}{n})(n_4+n_6)}{[(1-\frac{n_0}{n})(1-\frac{n_1}{n_1+n_3+n_5})]^{\frac{\log(\frac{n_3+n_5}{n_5})}{\log(\frac{n_4+n_6}{n_6})}}(n_2+n_4+n_6)}\right)^{\frac{1}{1-\frac{\log(\frac{n_3+n_5}{n_5})}{\log(\frac{n_4+n_6}{n_6})}}}.
\end{aligned}$$

Since $(\hat{\beta}, \hat{S}_1, \hat{S}_3, \hat{S}_4)$ is a function of $(n_1, n_2, n_3, n_4, n_5, n_8)/n$, it is easy to show that the SMLE is consistent. Notice that under given assumptions, it becomes a multinomial distribution problem, which belongs to the exponential family. The SMLE of $(\beta, S_1, S_2, S_3, S_4)$ is just the MLE under the multivariate distribution. If $\beta > 0$, then one can further show that the SMLE of β is asymptotically normally distributed and obtains the Cramer-rao lower bound. Thus $\hat{\beta}$ is the efficient estimator. Moreover, if $\beta \neq 0$, one can also show that $\hat{S}_o(a)$ is consistent and asymptotically normally distributed. It is worth mentioning that $P(\hat{\beta} = 0) \to 0$ a.s.. Hence $\hat{S}_2$ is well-defined most of the time. Thus the SMLE of S_o assigns weight to a $(= 2$ here). However, $\hat{S}_o(a)$ is not consistent if $\beta = 0$.

We now discuss $\breve{\beta}$. It is interesting to notice that $\mathcal{L}$ can be written as

$$\mathcal{L} = (1-S_1)^{n_0}(S_1-S_3)^{n_1} g(S_2)(S_3-S_4)^{n_3}(S_3^{e^{\beta}} - S_4^{e^{\beta}})^{n_4} S_4^{n_5+n_6 e^{\beta}},$$

where $g(x) = (t - x^y b)^{n_2} x^s$, where $t = S_1$, $x = S_o(a)$, $y = 1 - e^\beta$, $b = S_3^{e^\beta}$ and $s = (1 - e^\beta)(n_4 + n_6)$. $(\log g)' = 0$ yields the unique zero point of $(\log g)'$:

$$x = x_o = (\frac{st/b}{n_2 y + s})^{1/y} = (\frac{S_1(n_4 + n_6)/S_3^{e^\beta}}{n_2 + n_4 + n_6})^{1/(1-e^\beta)} \text{ if } \hat{\beta} \neq 0.$$

The expression is the same as $\hat{S}_2$ if one replaces S_i by $\hat{S}_i$

Under the given assumptions in this example and the non-parametric setup, there is a closed form solution to the parameters based on the sample with $u_i = 0$ and $u_i = 1$, respectively. Let $n_{00} = \sum_{i=1}^n \mathbf{1}(R_i = 1, u_i = 0)$ and $n_{01} = \sum_{i=1}^n \mathbf{1}(R_i = 1, u_i = 1)$. In this set-up, the degree of freedom for the parameters is 6 under the non-parametric set-up. We first estimate the 6 parameters $S(i|u)$ for $i \in \{2, 3, 4\}$ and $u \in \{0, 1\}$. Then estimate $S_o(a)$ and β based on the 6 parameters. Thus the non-parametric MLE satisfies

$1 - S_1 = \frac{n_{00}}{n_{00}+n_1+n_3+n_5}$,

$S_1 - S_3 = \frac{n_1}{n_{00}+n_1+n_3+n_5}$,

$S_4 = \frac{n_5}{n_{00}+n_1+n_3+n_5}$,

$1 - S(1|1) = \frac{n_{01}}{n_{01}+n_2+n_4+n_6}$,

$S(1|1) - S(3|1) = \frac{n_2}{n_{01}+n_2+n_4+n_6}$,

$S(4|1) = \frac{n_6}{n_{01}+n_2+n_4+n_6}$,

The first 3 equations lead to an estimate of (S_1, S_3, S_4), that is,

$\breve{S}_1 = 1 - \frac{n_{00}}{n_{00}+n_1+n_3+n_5}$, $\breve{S}_3 = \breve{S}_1 - \frac{n_1}{n_{00}+n_1+n_3+n_5}$, $\breve{S}_4 = \frac{n_5}{n_{00}+n_1+n_3+n_5}$.

Notice that the last 3 equations lead to a different estimator of S_1: $\breve{S}_1 = \frac{n_{01}}{n_{01}+n_2+n_4+n_6}$.

Moreover, they lead to $S_a^{1-e^\beta} S_3^{e^\beta} = \frac{n_4+n_6}{n_{01}+n_2+n_4+n_6}$, $S_a^{1-e^\beta} S_4^{e^\beta} = \frac{n_6}{n_{01}+n_2+n_4+n_6}$.

Thus $\breve{\beta} = \log \frac{\log \frac{n_4+n_6}{n_6}}{\log \frac{\breve{S}_3}{\breve{S}_4}}$ and $\breve{S}_a = \left\{ \frac{\frac{n_4}{n_{01}+n_2+n_4+n_6}}{S_3^{e^\beta} - S_4^{e^\beta}} \right\}^{\frac{1}{1-e^\beta}}$. It can all be viewed as a non-parametric MLE of S_i and β. Then the SMLE of (S_o, β) $(\hat{S}_o, \hat{\beta})$ satisfies $\mathcal{L}(\breve{S}_o, \breve{\beta}) > \mathcal{L}(\hat{S}_o, \hat{\beta})$. In fact,

$$\begin{aligned}
\mathcal{L}(\hat{S}_o, \hat{\beta}) =& (\frac{n_0}{n})^{n_o}(1 - \frac{n_0}{n})^{n-n_o}(\frac{n_1}{n_1 + n_3 + n_5})^{n_1}(\frac{n_3}{n_1 + n_3 + n_5})^{n_3} \\
&\cdot (\frac{n_5}{n_1 + n_3 + n_5})^{n_5}(\frac{n_2}{n_2 + n_4 + n_6})^{n_2}(\frac{n_4}{n_2 + n_4 + n_6})^{n_4}(\frac{n_6}{n_2 + n_4 + n_6})^{n_6} \\
\leq& (\frac{n_{00}}{n_{00} + n_1 + n_3 + n_5})^{n_{00}}(\frac{n_1}{n_{00} + n_1 + n_3 + n_5})^{n_1}(\frac{n_3}{n_{00} + n_1 + n_3 + n_5})^{n_3} \\
&\cdot (\frac{n_5}{n_{00} + n_1 + n_3 + n_5})^{n_5}(\frac{n_{01}}{n_{01} + n_1 + n_3 + n_5})^{n_{01}}
\end{aligned}$$

$$\cdot (\frac{n_2}{n_{01}+n_2+n_4+n_6})^{n_2}(\frac{n_4}{n_{01}+n_2+n_4+n_6})^{n_4}(\frac{n_6}{n_{01}+n_2+n_4+n_6})^{n_6} = \mathcal{L}(\breve{S}_o, \breve{\beta}).$$

It can be verified that $\lim_{n\to\infty} \sigma^2_{\breve{\beta}} n$ does not equal the Cramer-Rao lower bound and thus the SMLE is more efficient than $\breve{\beta}$. □

Proof of Example 2.3. By the given assumptions, (L_i, R_i) is of the forms $(-\infty, 1)$, $(1, 3)$ and $(3, \infty)$. To show that the SMLE is not unique, let $n = 12$. Let

$$n_0 = \sum_{i\le n} \mathbf{1}(1 = R_i) = 4, \; n_1 = \sum_{i\le n/2} \mathbf{1}((L_i, R_i) = (1,3)) = 1,$$
$$n_2 = \sum_{i> n/2} \mathbf{1}((L_i, R_i) = (1,3)) = 2, \; n_3 = \sum_{i\le n/2} \mathbf{1}(L_i = 3) = 3, \; n_4 = \sum_{i>n/2} \mathbf{1}(L_i = 3) = 2.$$

Let $p_1 = F_o(1)$, $p_2 = F_o(3) - F_o(2)$, $p_3 = S_o(3)$, and $p_4 = F_o(2) - F_o(1)$. By Corollary 1, the generalized likelihood in (1.4) becomes

$$\mathcal{L} = (p_1)^{n_0}(1-p_1-(p_2+p_3)^{1-e^\beta}p_3^{e^\beta})^{n_1}(1-p_1-p_3)^{n_2}((p_2+p_3)^{1-e^\beta}p_3^{e^\beta})^{n_3}p_3^{n_4}$$
$$= (p_1)^{n_0}(1-p_1)^{n-n_0}(1-\frac{(p_2+p_3)^{1-e^\beta}p_3^{e^\beta}}{1-p_1})^{n_1}(1-\frac{p_3}{1-p_1})^{n_2}(\frac{(p_2+p_3)^{1-e^\beta}p_3^{e^\beta}}{1-p_1})^{n_3}(\frac{p_3}{1-p_1})^{n_4}.$$

The maximum point (p_1, p_2, p_3, β) of $\mathcal{L}$ is not unique. In particular, the maximum value of $\mathcal{L}$ is $(\frac{n_0}{n})^{n_0}(1-\frac{n_0}{n})^{n-n_0}(\frac{n_1}{n_1+n_3})^{n_1}(\frac{n_2}{n_2+n_4})^{n_2}(\frac{n_3}{n_1+n_3})^{n_3}(\frac{n_4}{n_2+n_4})^{n_4}$, which can be obtained at $p_1 = \frac{n_0}{n} = 1/3$, $p_3 = (1-p_1)(\frac{n_4}{n_2+n_4}) = 1/3$, and $(p_2+p_3)^{1-e^\beta}p_3^{e^\beta} = (1-p_1)(\frac{n_3}{n_1+n_3}) = 1/2$. These equations yield $\beta = \ln\frac{\ln\frac{1/2}{p_2+1/3}}{\ln\frac{1/3}{p_2+1/3}}$. If $p_2 \in (1/6, 1/3]$ and $\beta \in (-\infty, \ln\frac{\ln\frac{4}{3}}{\ln 2}]$, then $p_2 + p_4 = 1/3$. Thus each $(p_1, p_3, p_2, \beta) = (1/3, 1/3, p_2, \ln\frac{\ln\frac{1/2}{p_2+1/3}}{\ln\frac{1/3}{p_2+1/3}})$, $p_2 \in (1/6, 1/3]$, is a solution to the SMLE.

Note that the expression of the SMLE is valid even if n is arbitrary. It can be shown that each SMLE is not consistent by the strong law of large numbers.

Proof of Proposition 2. Recall that under the TICPH model, the weights of the SMLE are assigned to the right-end points of the innermost intervals. If this is true also for the TDCPH model, then the likelihood becomes

$$\mathcal{L}_o = (1-S_1)^{n_0}S_1^{n-n_0}(1-\frac{S_3}{S_1})^{n_1}(\frac{S_3}{S_1}-\frac{S_4}{S_1})^{n_3}(\frac{S_4}{S_1})^{n_5}$$
$$(1-(\frac{S_3}{S_1})^{e^\beta})^{n_2}((\frac{S_3}{S_1})^{e^\beta}-(\frac{S_4}{S_1})^{e^\beta})^{n_4}((\frac{S_4}{S_1})^{e^\beta})^{n_6}.$$

An estimate based on the NPMLE yields

$$(\frac{\hat{S}_4}{\hat{S}_1})^{e^{\hat{\beta}}} = \frac{n_6}{n_2+n_4+n_6}, \ \hat{S}_1 = 1 - n_0/n \text{ and } \hat{S}_4/\hat{S}_1 = \frac{n_5}{n_1+n_3+n_5}.$$

Then $e^{\hat{\beta}} = \frac{\log \frac{n_6}{n_2+n_4+n_6}}{\log \frac{\hat{S}_4}{\hat{S}_1}} \to e^{\beta} + \frac{\log S_2}{\log \frac{S_4}{S_1}}$ a.s.. Thus it does lead to a consistent estimator of β. $\square$

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

References.

Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data. *Chapman & Hall* NY.

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

Groeneboom, P. and Wellner, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

Leffondre K., Wynant W., Cao Z., et al. (2010). A weighted Cox model for modeling time-dependent exposures in the analysis of case-control studies. *Statistics in Medicine*, 29 839-850.

Lehmann, E. L. (1959). Testing statistical hypotheses. 1st ed.

Masaaki, T. and Masato, S. (2009). Analysis of survival data having time-dependent covariates. *IEEE Transaction on Neural Networks*, 20 389-394.

Platt, R.W., Joseph3, K. S., Ananth, C. V., Grondines1, J., Abrahamowicz, M, and Kramer, M. S. (2004). A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *American Journal of Epidemiology*, 160 199-206.

Schick, A. and Yu, Q.Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.*, 27 45-55.

Stephan, L. and Michael, S. (2007). Parsimonious analysis of time-dependent effects in the Cox model. *Statistics in Medicine*, 26 2686-2698.

Therneau, T., and Grambsch, P. (2000). Modeling survival data : extending the Cox model. Springer.

Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser.* B. 38, 290-295.

Wong, G.Y.C and Yu, Q.Q (2012). Estimation under the Lehmann regression model with interval-censored data. *Comm. Statist. — Comput. Simul.* 41 1489-1500.

Wong, G. Y. C., Osborne, M. P., Diao, Q.G., and Yu, Q.Q. (2016). Piece-wise Cox Models with right-censored data. *Comm. Statist. — Comput. Simul.* (accepted).

Yu, Q.Q. (2006). A note on the proportional hazards model with discrete data. *Statistics and Probability Letters*, 77 735-739.

Zhang H.M. and Huang C.F. (2006). Nonparametric survival analysis on time-dependent covariate effects in case-cohort sampling design. *Statistica Sinica*, 16 267-285.

Zhou, M. (2001). Understanding the Cox regression models with time-change covariates. *American Statistian*, 55 153-155

Appendix I. The partial derivatives of the likelihood needed in §3 for IC data.

Noticing that $\ln(S(t_{l_i}|\mathbf{z}_i)-S(t_{r_i}|\mathbf{z}_i))=\begin{cases}\ln(1-S(t_{r_i}|\mathbf{z}_i)) & \text{if } l_i=0 \text{ and } r_i<m\\ \ln S(t_{l_i}|\mathbf{z}_i) & \text{if } r_i=m.\end{cases}$

$$\frac{\partial \ln\mathcal{L}}{\partial\beta}=\sum_{r_i\le i_a}\frac{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}(\ln S_{l_i})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}-(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}(\ln S_{r_i})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}}{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}}\qquad (=\tfrac{D_{-l_i}-D_{-r_i}}{\cdots-\cdots})$$

$$+\sum_{l_i\le i_a<r_i}\frac{1(r_i<m)(D_{-l_i}-D_{+r_i})}{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{r_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}}$$

$$+\sum_{l_i>i_a}\frac{\mathbf{1}(r_i<m)(D_{+l_i}-D_{+r_i})}{(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{l_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{r_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}}$$

$$+\sum_{l_i\le i_a,r_i=m}e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}\ln S_{l_i}+\sum_{l_i>i_a,r_i=m}((e^{\mathbf{V}'_{3i}\beta}\ln S_{i_a})\mathbf{v}_{3i}+e^{\mathbf{V}'_{2i}\beta}(\ln\frac{S_{l_i}}{S_{i_a}})\mathbf{v}_{2i})]$$

where $D_{-r_i}=(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}(\ln S_{r_i})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}$ and

$$D_{+r_i}=(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(\frac{S_{r_i}}{S_{i_a}})^{e^{\mathbf{V}'_{2i}\beta}}[(\ln S_{i_a})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}+(\ln\frac{S_{r_i}}{S_{i_a}})e^{\mathbf{V}'_{2i}\beta}\mathbf{v}_{2i}] \text{ if } r_i<m,$$

and D_{-l_i} and D_{+l_i} are defined in an obvious way.

$$\frac{\partial^2\ln\mathcal{L}}{\partial\beta\partial\beta'}=\sum_{r_i\le i_a}\frac{(B_{-l_i}-B_{-r_i})}{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}}-\frac{(D_{-l_i}-D_{-r_i})(D_{-l_i}-D_{-r_i})'}{((S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}})^2}$$

$$+\sum_{l_i\le i_a<r_i}[\frac{B_{-l_i}-B_{+r_i}}{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(\frac{S_{r_i}}{S_{i_a}})^{e^{\mathbf{V}'_{2i}\beta}}}-\frac{(D_{-l_i}-D_{+r_i})(D_{-l_i}-D_{+r_i})'}{((S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(\frac{S_{r_i}}{S_{i_a}})^{e^{\mathbf{V}'_{2i}\beta}})^2}]$$

$$+\sum_{l_i>i_a}[\frac{(B_{+l_i}-B_{+r_i})}{(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(\frac{S_{l_i}}{S_{i_a}})^{e^{\mathbf{V}'_{2i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(\frac{S_{r_i}}{S_{i_a}})^{e^{\mathbf{V}'_{2i}\beta}}}-\frac{(D_{+l_i}-D_{+r_i})(D_{+l_i}-D_{+r_i})'}{(S(L_i|\mathbf{z}(L_i))-S(R_i|\mathbf{z}(R_i)))^2}]$$

$$+\mathbf{1}(r_i=m)[\sum_{l_i\le i_a}e^{\mathbf{V}'_{3i}\beta}(\ln S_{l_i})\mathbf{v}_{3i}\mathbf{v}'_{3i}+\sum_{l_i>i_a}(e^{\mathbf{V}'_{3i}\beta}(\ln S_{i_a})\mathbf{v}_{3i}\mathbf{v}'_{3i}+e^{\mathbf{V}'_{2i}\beta}(\ln\frac{S_{l_i}}{S_{i_a}})\mathbf{v}_{2i}\mathbf{v}'_{2i})]$$

where $B_{-l_i}=(1+(\ln S_{l_i})e^{\mathbf{V}'_{3i}\beta})D_{-l_i}\mathbf{v}'_{3i}=(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}((\ln S_{l_i})e^{\mathbf{V}'_{3i}\beta}+((\ln S_{l_i})e^{\mathbf{V}'_{3i}\beta})^2)\mathbf{v}_{3i}\mathbf{v}'_{3i}$ and

$$B_{+r_i}=(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(\frac{S_{r_i}}{S_{i_a}})^{e^{\mathbf{V}'_{2i}\beta}}\{(\ln S_{i_a})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}\mathbf{v}'_{3i}+(\ln\frac{S_{r_i}}{S_{i_a}})e^{\mathbf{V}'_{2i}\beta}\mathbf{v}_{2i}\mathbf{v}'_{2i}$$

$$+[(\ln S_{i_a})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}+(\ln\frac{S_{r_i}}{S_{i_a}})e^{\mathbf{V}'_{2i}\beta}\mathbf{v}_{2i}][(\ln S_{i_a})e^{\mathbf{V}'_{3i}\beta}\mathbf{v}_{3i}+(\ln\frac{S_{r_i}}{S_{i_a}})e^{\mathbf{V}'_{2i}\beta}\mathbf{v}_{2i}]'\}\ if\ r_i<m,$$

and B_{-r_i} and B_{+l_i} are defined in an obvious way.

Given $k\in\{1,...,m-1\}$, write $U_{jk}(u)=\begin{cases}\frac{S_j}{1+u} & \text{if } k\le j\\ \frac{S_j+u}{1+u} & \text{if } k>j\end{cases}=\begin{cases}\frac{S_j}{1+u} & \text{if } k\le j\\ \frac{S_j-1}{1+u}+1 & \text{if } k>j.\end{cases}$ Let $\mathcal{L}=\mathcal{L}(S_1,...,S_{m-1})$ and $H_k(u)=\ln\mathcal{L}(U_{j1}(u),...,U_{i,m-1}(u))$. Then $\frac{\partial U_{jk}(u)}{\partial u}=\begin{cases}-\frac{S_j}{(1+u)^2} & \text{if } 0<k\le j<m\\ -\frac{S_j-1}{(1+u)^2} & \text{if } m>k>j>0.\end{cases}$ $\frac{\partial^2U_{jk}(u)}{\partial u^2}=\begin{cases}2\frac{S_j}{(1+u)^3} & \text{if } 0<k\le j<m\\ 2\frac{S_j-1}{(1+u)^3} & \text{if } m>k>j>0.\end{cases}$ Moreover, $U_{jk}(0)=S_j$, $U'_{jk}(0)=\frac{\partial U_{jk}(0)}{\partial u}=\begin{cases}-S_j & \text{if } 0<k\le j<m\\ 1-S_j & \text{if } 0<j<k<m,\\ 0 & \text{otherwise.}\end{cases}$

$$U''_{jk}(0) = \frac{\partial^2 U_{jk}(0)}{\partial u^2} = \begin{cases} 2S_j & \text{if } 0 < k \le j < m \\ 2(S_j - 1) & \text{if } m > k > j > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Abusing notations, write $S_j = S_j(u) = U_{jk}(u)$.

$$\frac{\partial H_k}{\partial u}\Big|_{u=0} = \sum_{r_i \le i_a} e^{\beta' \mathbf{v}_{3i}} \frac{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U'_{l_i k}(0) - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U'_{r_i k}(0)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}}}$$

$$+ \sum_{l_i \le i_a < r_i} \frac{e^{\beta'\mathbf{v}_{3i}} (S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U'_{l_i k}(0) - C_{r_i} U'_{i_a k}(0) - W_{r_i} U'_{r_i k}(0)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}}$$

$$+ \sum_{l_i > i_a} \frac{C_{l_i} U'_{i_a k}(0) + W_{l_i} U'_{l_i k}(0) - C_{r_i} U'_{i_a k}(0) - W_{r_i} U'_{r_i k}(0)}{(S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{l_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}}$$

$$+ \mathbf{1}(r_i = m)[\sum_{l_i \le i_a} \frac{e^{\beta'\mathbf{v}_{3i}} U'_{l_i k}(0)}{S_{l_i}} + \sum_{l_i > i_a} (\frac{(e^{\beta'\mathbf{v}_{3i}} - e^{\beta'\mathbf{v}_{2i}}) U'_{i_a k}(0)}{S_{i_a}} + \frac{e^{\beta'\mathbf{v}_{2i}} U'_{l_i k}(0)}{S_{l_i}})],$$

where $C_{r_i} = (e^{\beta'\mathbf{v}_{3i}} - e^{\beta'\mathbf{v}_{2i}})(S_{i_a})^{e^{\beta'\mathbf{v}_{3i}} - e^{\beta'\mathbf{v}_{2i}} - 1} (S_{r_i})^{e^{\beta'\mathbf{v}_{2i}}} \; if \; r_i < m,$

$$W_{r_i} = (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}} - e^{\beta'\mathbf{v}_{2i}}} e^{\beta'\mathbf{v}_{2i}} (S_{r_i})^{e^{\beta'\mathbf{v}_{2i}}-1} \; if \; r_i < m,$$

and C_{l_i} and W_{l_i} are defined in an obvious way.

$$\frac{\partial^2 H_k}{\partial u^2}\Big|_{u=0} = \sum_{l_i \le i_a} \{ e^{\beta'\mathbf{v}_{3i}} \frac{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U''_{l_i k}(0) - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U''_{r_i k}(0)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}}}$$

$$+ (e^{\beta'\mathbf{v}_{3i}} - 1) e^{\beta'\mathbf{v}_{3i}} \frac{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-2} (U'_{l_i k}(0))^2 - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}-2} (U'_{r_i k}(0))^2)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}}}$$

$$- (e^{\beta'\mathbf{v}_{3i}} \frac{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U'_{l_i k}(0) - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U'_{r_i k}(0))}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{r_i})^{e^{\beta'\mathbf{v}_{3i}}}})^2 \}$$

$$+ \sum_{l_i \le i_a < r_i} \{ \frac{e^{\beta'\mathbf{v}_{3i}} (S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U''_{l_i k}(0) - C_{r_i} U''_{i_a k}(0) - W_{r_i} U''_{r_i k}(0)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}}$$

$$+ \frac{(e^{\beta'\mathbf{v}_{3i}} - 1) e^{\beta'\mathbf{v}_{3i}} (S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-2} (U'_{l_i k}(0))^2 - C'_{r_i k} U'_{i_a k}(0) - W'_{r_i k} U'_{r_i k}(0)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}}$$

$$- (\frac{e^{\beta'\mathbf{v}_{3i}} (S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}-1} U'_{l_i k}(0) - C_{r_i} U'_{i_a k}(0) - W_{r_i} U'_{r_i k}(0)}{(S_{l_i})^{e^{\beta'\mathbf{v}_{3i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}})^2$$

$$+ \mathbf{1}(r_i = m) e^{\beta'\mathbf{v}_{3i}} (\frac{U''_{l_i k}(0)}{S_{l_i}} - (\frac{U'_{l_i k}(0)}{S_{l_i}})^2) \}$$

$$+ \sum_{l_i > i_a} \{ \frac{C_{l_i} U''_{i_a k}(0) + W_{l_i} U''_{l_i k}(0) - C_{r_i} U''_{i_a k}(0) - W_{r_i} U''_{r_i k}(0)}{(S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{l_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}}$$

$$+ \frac{C'_{l_i k} U'_{i_a k}(0) + W'_{l_i k} U'_{l_i k}(0) - C'_{r_i k} U'_{i_a k}(0) - W'_{r_i k} U'_{r_i k}(0)}{(S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{l_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}} - (S_{i_a})^{e^{\beta'\mathbf{v}_{3i}}} (\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{v}_{2i}}}}$$

$$
\begin{aligned}
&-(\frac{C_{l_i}U'_{i_ak}(0)+W_{l_i}U'_{l_ik}(0)-C_{r_i}U'_{i_ak}(0)-W_{r_i}U'_{r_ik}(0)}{(S_{i_a})^{e^{\beta'\mathbf{V}_{3i}}}(\frac{S_{l_i}}{S_{i_a}})^{e^{\beta'\mathbf{V}_{2i}}}-(S_{i_a})^{e^{\beta'\mathbf{V}_{3i}}}(\frac{S_{r_i}}{S_{i_a}})^{e^{\beta'\mathbf{V}_{2i}}}})^2\\
&+\mathbf{1}(r_i=m)[(e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}})(\frac{U''_{i_ak}(0)}{S_{i_a}}-(\frac{U'_{i_ak}(0)}{S_{i_a}})^2)+e^{\beta'\mathbf{V}_{2i}}(\frac{U''_{l_ik}(0)}{S_{l_i}}-(\frac{U'_{l_ik}(0)}{S_{l_i}})^2)]\}
\end{aligned}
$$

where
$$
\begin{aligned}
C'_{r_ik}&=(e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}}-1)(e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}})(S_{i_a})^{e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}}-2}(S_{r_i})^{e^{\beta'\mathbf{V}_{2i}}}U'_{i_ak}(0)\\
&+(e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}})(e^{\beta'\mathbf{V}_{2i}}-1)(S_{i_a})^{e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}}-1}(S_{r_i})^{e^{\beta'\mathbf{V}_{2i}}-1}U'_{r_ik}(0)\ if\ r_i<m,\\
W'_{r_ik}&=(e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}})(S_{i_a})^{e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}}-1}e^{\beta'\mathbf{V}_{2i}}(S_{r_i})^{e^{\beta'\mathbf{V}_{2i}}-1}U'_{i_ak}(0)\\
&+(S_{i_a})^{e^{\beta'\mathbf{V}_{3i}}-e^{\beta'\mathbf{V}_{2i}}}(e^{\beta'\mathbf{V}_{2i}}-1)e^{\beta'\mathbf{V}_{2i}}(S_{r_i})^{e^{\beta'\mathbf{V}_{2i}}-2}U'_{r_ik}(0)\ if\ r_i<m,
\end{aligned}
$$

and C'_{l_ik} and W'_{l_ik} are defined in an obvious way.

$$
\begin{aligned}
\frac{\partial\ln\mathcal{L}}{\partial S_j}&=\sum_{r_i\le i_a}\frac{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}\frac{1(l_i=j)}{S_{l_i}}e^{\mathbf{V}'_{3i}\beta}-(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}\frac{1(r_i=j)}{S_{r_i}}e^{\mathbf{V}'_{3i}\beta}}{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}}\\
&+\sum_{l_i\le i_a<r_i}[\frac{\mathbf{1}(r_i<m)(G_{-l_ij}-G_{+r_ij})}{(S_{l_i})^{e^{\mathbf{V}'_{3i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{r_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}}+\mathbf{1}(r_i=m,l_i=j)\frac{e^{\mathbf{V}'_{3i}\beta}}{S_{l_i}}]\\
&+\sum_{l_i>i_a}[\frac{\mathbf{1}(r_i<m)(G_{+l_ij}-G_{+r_ij})}{(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{l_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}-(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{r_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}}\\
&+\mathbf{1}(r_i=m)(\mathbf{1}(i_a=j)\frac{e^{\mathbf{V}'_{3i}\beta}-e^{\mathbf{V}'_{2i}\beta}}{S_{i_a}}+\mathbf{1}(l_i=j)\frac{e^{\mathbf{V}'_{2i}\beta}}{S_{l_i}})],\\
G_{-r_ij}&=(S_{r_i})^{e^{\mathbf{V}'_{3i}\beta}}\frac{1(r_i=j)}{S_{r_i}}e^{\mathbf{V}'_{3i}\beta},\\
G_{+r_ij}&=(S_{i_a})^{e^{\mathbf{V}'_{3i}\beta}}(S_{r_i}/S_{i_a})^{e^{\mathbf{V}'_{2i}\beta}}[(e^{\mathbf{V}'_{3i}\beta}-e^{\mathbf{V}'_{2i}\beta})\frac{\mathbf{1}(i_a=j)}{S_{i_a}}+e^{\mathbf{V}'_{2i}\beta}\frac{1(r_i=j\ne i_a)}{S_{r_i}}],
\end{aligned}
$$

$G_{-l_i,j}$ and G_{+l_ij} are defined in an obvious way. Notice that in order to derive the covariance matrix, we need to compute $\frac{\partial\ln\mathcal{L}}{\partial S_j}$ for $j\in\{1,...,m-1\}$, as $\hat{C}ov=\hat{I}^{-1}$ and I is the Fisher information matrix.

Appendix II We use a simple numerical example to illustrate why the various existing algorithms do not work for the SMLE.

§**a.1.** Consider fitting time-indepemdent covarites Cox's regression model with five observations (L_i, R_i, Z_i): (2,5,0), (3,4,0), (5,9,1), (1,6,1), (7,8,0). It can be viewed as data from two groups, corresponding to $Z_i = 0$ or 1. Then, the innermost intervals are (3,4), (5,6) and (7,8). Let the weights on these innermost intervals be p_1, p_2 and p_3, with $p_1+p_2+p_3 = 1$ and $p_i \geq 0$. Note that the baseline survival function S satisfies $S(4-) = 1$, $S(4) = S(6-) = p_2 + p_3$, $S(6) = S(8-) = p_3$ and $S(8) = 0$. For this example, it is more convenient to express the likelihood as a function of p_i's rather than S. The likelihood is $L = p_1^2 p_3(1 - p_3^{e^\beta})(p_2 + p_3)^{e^\beta}$. Since $p_1 + p_2 + p_3 = 1$, in view of L, it is simpler to write the log likelihood as

$$l = \log[p_1^2 p_3 (1 - p_1)^{e^\beta}(1 - p_3^{e^\beta})]. \qquad (A.1)$$

The parameter space is $\Omega = \{(\beta, p_1, p_3) : \beta \in (-\infty, \infty), p_1 \geq 0, p_3 \geq 0, p_1 + p_3 \leq 1\}$ with $p_2 = 1 - p_1 - p_3$. For convenience, we write $\alpha = e^\beta$ hereafter. Thus,

$$l = 2 \log p_1 + \log p_3 + \alpha \log(1 - p_1) + \log(1 - p_3^\alpha).$$

Since the likelihood function has only three variables, it can be shown by direct derivation that the SMLE of (β, p_1, p_2, p_3) is approximately $(-0.461, 2/3, 0, 1/3)$.

In general, the likelihood is not so simple and one needs to compute the SMLE by numerical methods. We shall illustrate by this example that several naive numerical methods fail to yield the SMLE. They include: (a) the Newton-Raphson (NR) method; (b) the scaled NR method and (c) the profile likelihood (PL) method. Finally, we shall illustrate by this example why our new algorithm can yield the SMLE. The main difference is that the first three algorithms cannot search the SMLE along the line $p_2 = 0$ (or $p_1 + p_3 = 1$), while the new algorithm can. Note that the SMLE is on boundary $p_2 = 0$.

§**a.2.** In order to apply the NR method, we need to compute the partial derivatives.

$$\frac{\partial l}{\partial \alpha} = \log(1 - p_1) - \frac{p_3^\alpha \log p_3}{1 - p_3^\alpha}, \ \frac{\partial^2 l}{\partial \alpha^2} = -\frac{p_3^\alpha (\log p_3)^2}{(1 - p_3^\alpha)^2}, \qquad (A.2)$$

$$\frac{\partial l}{\partial p_1} = \frac{2}{p_1} - \frac{\alpha}{1-p_1} \qquad \text{and } \frac{\partial l}{\partial p_3} = \frac{1}{p_3} - \frac{\alpha p_3^{\alpha-1}}{1-p_3^{\alpha}}. \tag{A.3}$$

§**a.2.1.** (The NR method). At the SMLE $(p_1, p_3) = (2/3, 1/3)$ with $\beta = -0.461$, Equation (A.3) yields that the gradient in (p_1, p_3) is $(1.11, 1.11)$. In other words, as (p_1, p_2) moves towards outside the parameter space, the likelihood increases. Thus the maximum value of L without the restriction of the parameter space can only be achieved outside the parameter space. The NR yields the unrestricted maximum point of L. Thus the solution to the NR algorithm is not the SMLE.

§**a.2.2.** A scaled NR method is as follows.

Let $\beta = 0$ or $\alpha = 1$ be the initial value, and let the SMLE (or SCE) of (p_1, p_3) at $\beta = 0$ be the initial value of (p_1, p_3).

Step 1. Maximize L over β with given up-dated (p_1, p_3) using the NR method.

Step 2. Maximize L over (p_1, p_3) with up-dated β using a scaled NR method, that is, scale the increments $\triangle p_i$'s in the original NR algorithm by a constant c so that the updated (p_1, p_3) remains in the parameter space.

Repeat Steps 1 and 2 until convergence.

However, it does not work in this example. In particular, in the initial step. we have $\beta = 0$ (or $\alpha = 1$) and $(p_1, p_3) = (3/5, 2/5)$. In Step 1, L is maximized by $\alpha = -\frac{\log 2}{\log 0.4} \approx 0.76$ (see Eq. (A.4)). In Step 2, by Equation (A.3), the gradient at $(p_1, p_3) = (3/5, 2/5)$ is $(1.44, 0.61)$. Thus (p_1, p_3) should be up-dated to $(\frac{3}{5} + 1.44x, \frac{2}{5} + 0.61x)$ for some $x \geq 0$. If $x > 0$, it violates the constraint $p_1 + p_3 \leq 1$. Thus the algorithm stops at $S(4) = p_2 + p_3 = 2/5$ and $S(6) = p_3 = 2/5$ with $\beta = \log 0.76$ $(= -0.274)$, which is not the SMLE.

§**a.2.3.** A PL approach is as follows:

The initial step and Step 1 are the same as in the scaled NR method above.

Step 2 (p_1-substep). Maximize L over p_1 with up-dated p_3 and β.

Step 3 (p_3-substep). Maximize L over p_3 with up-dated p_1 and β.

Repeat Steps 1, 2 and 3 until convergence.

However, the PL method still does not work. In particular, at Step 1, $\alpha = 0.76$, $p_1 = 0.6$ and $p_3 = 0.4$. The gradient at $(p_1, p_3) = (0.6, 0.4)$ is $(1.44, 0.61)$. Thus we move (p_1, p_3) either to $(0.6 + 1.44x, 0.4)$ with $x \geq 0$ (p_1 substep), or to $(0.6, 0.4 + 0.61x)$ with $x \geq 0$ (p_3 substep). If $x > 0$, both the p_1-substep and the p_3-substep will move (p_1, p_3) outside the parameter space. Consequently, it will stop at the value which is not the SMLE.

§**a.2.4.** There are three line segments in the boundary of the parameter space in (p_1, p_3). They are $p_1 = 0$, $p_3 = 0$ and $p_1 + p_3 = 1$. One can find the value that maximizes the likelihood on these line segments separately, using the NR method, and then check which is the SMLE. This approach works in this example. However, if there are m p_i's, we need to consider the subsets of the boundary corresponding to the $m - 2$ cases: case (1) one $p_i = 0$, case (2) two $p_i = 0$, ..., case $(m-2)$ $(m-2)$ $p_i = 0$. Thus the order is $O(m^2/2)$. When m is large, this approach is not feasible.

§**a3.** We now illustrate why the new algorithm works. Our new algorithm is as follows.

The initial step. Let the SMLE of (p_1, p_2, p_3) *be the initial value of the* (p_1, p_2, p_3) *and* $\alpha = 1$ *the initial value of* α.

β*-step. Maximize* L *over* β *with up-dated* p_i*'s.*

S*-step. Each* S*-step consists of 3 substeps:* p_1*-substep,* p_2*-substep,* p_3*-substep.*

p_1*-substep. Consider a transformation* $p_{11}(u) = \frac{p_1+u}{1+u}$, $p_{12}(u) = \frac{p_2}{1+u}$, *and* $p_{13}(u) = \frac{p_3}{1+u}$, $u > 0$. *This transformation ensures that* $(p_{11}(u), p_{12}(u), p_{13}(u))$ *remains in the parameter space of* (p_1, p_2, p_3) *for each* $u > 0$. *Let* u_o *be the value of* u *that maximizes* $L(\beta, p_{11}(u), p_{12}(u), p_{13}(u))$ *over* $u \geq 0$, *with* β *and* p_i*'s given in the previous step. Then up-date* p_i *by* $p_i = p_{1i}(u_o)$, $i = 1, 2, 3$.

p_2*-substep. Consider another transformation* $p_{21}(u) = \frac{p_1}{1+u}$, $p_{22}(u) = \frac{p_2+u}{1+u}$, *and* $p_{23}(u) = \frac{p_3}{1+u}$. *If* $\frac{\partial}{\partial u}\ln L(\beta, p_{21}(u), p_{23}(u))\big|_{u=0} > 0$, *choose a* $u_o > 0$ *that maximizes* $L(\beta, p_{21}(u), p_{22}(u), p_{23}(u))$ *over* $u \geq 0$, *with* β *and* p_i*'s given in the previous step. Update* p_i *by* $p_i = p_{2i}(u_o)$, $i = 1, 2, 3$.

p_3-substep. Consider a further new transformation $p_{31}(u) = \frac{p_1}{1+u}$, $p_{32}(u) = \frac{p_2}{1+u}$, and $p_{33}(u) = \frac{p_3+u}{1+u}$. If $\frac{\partial}{\partial u}\ln L(\beta, p_{31}(u), p_{33}(u))\big|_{u=0} > 0$, choose a $u_o > 0$ that maximizes $L(\beta, p_{31}(u), p_{32}(u), p_{33}(u))$ over $u \geq 0$, with β and p_i's given in the previous step. Up-date p_i by $p_i = p_{3i}(u_o)$, $i = 1,\ 2,\ 3$.

At the p_1-substep of the initial iteration step, by Eq. (A.5), $\frac{\partial}{\partial u}\ln L(\beta, p_{11}(u), p_{13}(u))\big|_{u=0} = 0.51 > 0$ at $(p_1, p_3) = (0.6, 0.4)$, and $u_o \approx 0.1$ maximizes $L(\beta, p_{21}(u), p_{23}(u))$. At this step (p_1, p_2, p_3) is up-dated to $(\frac{0.7}{1.1}, 0, \frac{0.4}{1.1})$ $(= (0.636, 0.364))$. At the p_2-substep and p_3-substep, by Equations (A.6) and (A.7), $\frac{\partial}{\partial u}\ln L(\beta, p_{i1}(u), p_{i3}(u))\big|_{u=0} < 0$, $i = 2,\ 3$, thus no change is made. However, since (p_1, p_2, p_3) is changed at this S-step, β (or α) will also be change at the next β-step.

In fact, in the next β-step, β is up-dated to $\ln 0.69 = -0.371$. In the p_1-substep, $\frac{\partial}{\partial u}\ln L(\beta, p_{11}(u), p_{13}(u))\big|_{u=0} = 0.14 > 0$, L is maximized by $(p_1, p_3) = (\frac{0.742}{1.142}, \frac{0.4}{1.142}) = (0.65, 0.35)$ with $u_o = 0.042$. by Equations (A.6) and (A.7), $\frac{\partial}{\partial u}\ln L(\beta, p_{i1}(u), p_{i3}(u))\big|_{u=0} < 0$, $i = 2,\ 3$, thus no change is made. However, since (p_1, p_2, p_3) is changed at this S-step, β (or α) will also be change at the next β-step.

Iteratively repeat these two steps, the algorithm will yield the SMLE $(\beta, p_1, p_2, p_3) = (-0.461, 2/3, 0, 1/3)$.

Remark 2. Recall that $p_i^{(0)}$ is the SMLE of p_i when $\beta = 0$. Let $\hat{p}_i$ be the SMLE under Cox's model. According to our observation, it is often the case that if $p_i^{(0)} = 0$ then $\hat{p}_i = 0$ too. It is not clear that whether it is indeed true that

$$p_i^{(0)} = 0 \text{ iff } \hat{p}_i = 0.$$

If this is true, then one can delete the p_i's for which $p_i^{(0)} = 0$ in the algorithm to reduce the dimension of the parameter space. Moreover, after this elimination, the NR method will work too, since the SMLE is in the interior of the parameter space. However, both the sufficient and the necessary conditions may not hold.

§**a4.** The following is the details of deriving the SMLE directly. There are only 3 variables in L, by direct examination, one can find that the maximum value of L is outside the parameter space and the SMLE of (p_1, p_2, p_3) is on the boundary of the parameter space. Moreover, the SMLE of (p_1, p_2, p_3) is on the subspace $p_2 = 0$, as $L = 0$ if $p_1 = 0$ or $p_3 = 0$. If $p_2 = 0$, $L = (1-p_3)^2 p_3^{1+\alpha}(1-p_3^\alpha)$. $\frac{\partial l}{\partial \alpha} = \log p_3 - \frac{p_3^\alpha \log p_3}{1-p_3^\alpha} = \log p_3 \frac{1-2p_3^\alpha}{1-p_3^\alpha} = 0$ implies that unless $p_3 = 1$, we have $p_3^\alpha = 1/2$ or $p_3 = 2^{-\alpha}$ or

$$\alpha = -\log 2/\log p_3. \tag{A.4}$$

For each fixed p_3, if $\alpha = -\log 2/\log p_3$, L achieves its maximum $(1-p_3)^2 p_3^{1-\log 2/\log p_3}(1-p_3)^{-\log 2/\log p_3}$. The SMLE can be found by plotting the graph of (p_3, L).

§**a5.** In this section, we shall derive the partial derivatives needed in §3.

$$\frac{\partial}{\partial u} p_{11}(u) = \frac{1-p_1}{(1+u)^2}, \ \frac{\partial}{\partial u} p_{12}(u) = \frac{-p_2}{(1+u)^2}, \ \frac{\partial}{\partial u} p_{11}(u) = \frac{-p_3}{(1+u)^2}.$$

$$\frac{\partial}{\partial u} \ln L(\beta, p_{11}(u), p_{13}(u))\Big|_{u=0} = \frac{2(1-p_1)}{p_1} + \frac{\alpha p_3^\alpha}{1-p_3^\alpha} - \alpha - 1 \tag{A.5}$$

$$\frac{\partial}{\partial u} p_{21}(u) = \frac{-p_1}{(1+u)^2}, \ \frac{\partial}{\partial u} p_{22}(u) = \frac{1-p_2}{(1+u)^2}, \ \frac{\partial}{\partial u} p_{23}(u) = \frac{-p_3}{(1+u)^2}.$$

$$\frac{\partial}{\partial u} \ln L(\beta, p_{21}(u), p_{23}(u))\Big|_{u=0} = -2 - 1 + p_3 \frac{\alpha p_3^{\alpha-1}}{1-p_3^\alpha} + \alpha \frac{p_1}{1-p_1}. \tag{A.6}$$

$$\frac{\partial}{\partial u} p_{31}(u) = \frac{-p_1}{(1+u)^2}, \ \frac{\partial}{\partial u} p_{32}(u) = \frac{-p_2}{(1+u)^2}, \ \frac{\partial}{\partial u} p_{33}(u) = \frac{1-p_3}{(1+u)^2}.$$

$$\frac{\partial}{\partial u} \ln L(\beta, p_{31}(u), p_{33}(u))\Big|_{u=0} = -2 + \frac{1-p_3}{p_3} - \frac{\alpha p_3^{\alpha-1}(1-p_3)}{1-p_3^\alpha} + \alpha \frac{p_1}{1-p_1} \tag{A.7}$$