Consistency of the Semi-Parametric MLE under the Piecewise Proportional Hazards Models with Interval-Censored Data

Qiqing Y\(^1\)* and Diao Q\(^2\)

Department of Mathematical Sciences, SUNY, USA

Submission: August 24, 2017; Published: October 02, 2017

*Corresponding author: Qiqing Y, Department of Mathematical Sciences, SUNY, USA, Email: qpy@math.binghamton.edu

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 is studied by Wong et al. [1]. 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. They proposed a modification to the likelihood function so that its SMLE is unique. Under certain regularity conditions, we show that the SMLE is consistent and is asymptotically normally distributed.

Keywords: Cox model; Time-dependent covariates; Semi-parametric MLE; Identifiability; consistency; Asymptotic normality

Abbreviations: PWPH: Piecewise Proportional Hazards; IC: Interval-Censored; PH: Proportional Hazards; TIPH: Time-Independent Covariate

Introduction

We establish the consistency of the semi-parametric MLE under the piecewise proportional hazards (PWPH) model, with interval-censored (IC) continuous survival time \( Y \). The proportional hazards (PH) model specifies that a covariate vector \( Z \) has a proportional effect on the hazard function of \( Y \). It is a common regression model for survival analysis. The PWPH model is a special PH model.

For a random variable \( Y \), denote its survival function by 
\[
P_Y(t) = P(Y > t),
\]
its density function by 
\[
f_Y(t)
\]
and its hazard function by 
\[
h(t) = \frac{f(t)}{P(t)},
\]
where \( h(t) > 0 \). Given a covariate (vector) \( Z \) which does not depend on time \( Y \), \( (Y, Z) \) follows a time-independent covariate PH (TIPH) model or Cox’s regression model if the conditional hazard function of 
\[
\lambda(z | Y) = \lambda_0(t) e^{\beta^T Z},
\]
is a constant function. The Cox model has been extended to the time-dependent covariates proportional hazards (TDPH) model. Cox & Oak [4] give a typical example of time-dependent covariate in medical research, namely,
\[
h(t | z) = e^{\beta^T z} h_0(t), \quad t < \tau,
\]
where \( \beta = (\beta_1, \ldots, \beta_k) \) is the transpose of the vector \( \beta \), \( \tau = \sup \{t : h(t) > 0\} \), and \( h_0 \) is an unknown baseline hazard function.

IC data consist of \( n \) time intervals with the end-points 
\[
L_i = L_i, i = 1, \ldots, n,
\]
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 \( L_i = R_i \). Schick & Yu [2] proposed the mixed case interval censorship model to specify the IC data without exact observations as follows. Let \( K \) be the number of follow-up time for a patient. Conditional on 
\[
K = k, Y \text{ and } (C_k, C_{k+1})
\]
are independent, where \( C_k, \ldots, C_{k+1} \) are the \( k \) follow-up times. The observable random vector is 
\[
(C_k, C_{k+1}, Y) \sim \text{TIPH}(Y | Z)
\]
where \( C_0 = -\infty \) and \( C_{k+1} = \infty \). If \( P(K = m) = 1 \), then the mixed case model becomes the case \( m \) interval censorship model [3].

The Cox model with IC data, we assume that \( Z \) and \( (Y, K, C) \) are independent, where \( Y, K, C \) are independent, where \( C = \{C_i : i \in \{1, \ldots, k\}, k \geq 1\} \).

The Cox model has been extended to the time-dependent covariates proportional hazards (TDPH) model. Cox & Oak [4] give a typical example of time-dependent covariate in medical research, namely,
\[
h(t | z) = e^{\beta^T z} h_0(t), \quad t < \tau,
\]
and \( c \) is the admission time to a treatment for a patient. They also give another example of time-dependent covariate. The TDPH model has been commonly used for right-censored (RC) data (see, for instance, Therneau & Grambsch [5], Platt et al. [6], Stephan & Michael [7], Masaaki & Masato [8], and Leffondre et al. [9]).

Zhou formulates a PWPH model with \( k \) cut points:
Moreover, where the be the conditional and without censoring. Then the be the of ) ( ) ( ) ( . The semi-parametric ( ) ( ) . Suppose that, in the sense that if is a time-independent covariate vector. is not , and So be absolutely continuous, , maximizes L over all survival functions ( ) ( ) ( ) ( ) ( ) . For instance, Wong et al. defined in (1.4) is applicable to all IC ( ) ( ) ( ) ( ) ( ) . Assume and ( ) ( ) ( ) ( ) ( ) . Under the mixed ( ) ( ) ( ) ( ) ( ) . Let ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) . The semi-parametric problem under the PWPH model with IC data was studied by Wong et al. [1]. It turns out that under PWPH model(1) with IC data, the parameter β is not identifiable unless further assumptions are imposed (see Example 1). Moreover, in general, the SMLE of β under the likelihood function (1.4) may not be unique. Both phenomena do not occur if the covariates are time-independent . They specified the Identifiability condition for such problems and studied the estimation problem of deriving the SMLE. Their simulation results suggest that the SMLEs of So and β are consistent under the mixed case IC model [2]. We give the proof of the consistency and asymptotic normality of the SMLE in this paper.

The Main Results

We study consistency of the SMLE under the PWPH model with one cut point assuming Y is continuous in this paper. In particular, we consider the model \( h_F(t|z)=h_z(t)c(t)\beta(t \geq c) \), where \( z \) is a time-independent covariate vector (2.1). Y is subject to interval censoring under the mixed case IC model with the following up times \( C_{ki} \) and the random number of follow-up times \( K. \) We first present some preliminary results [13].

Proposition 1

Under model (2.1), if Y is continuous, then
\[
S_{F_Y}(t|\beta) = \begin{cases} S(t_0) & \text{if } t \leq t_0 \\ \left(S(t_0)\right)^{\left(c-c_0\right)} & \text{if } t > c_0 \\ \end{cases}
\]

Abusing notations, we write \( S_{F_Y}(t|\beta) \) for \( S_{F_Y}(t|\beta) \).

Given a random variable, say \( Y \), let \( S_{F_Y} \) be the support set of \( F_Y \), in the sense that if \( x \in S_{F_Y} \) then \( F_Y(x - +) - F_Y(x - -) > 0 \). SFL and SFR are defined in a similar manner.

Lemma 1: Assume the PH model \( h(t|z)=h_z(t)e^{\beta z} \), with the parameter \( (\beta, S) \) and without censoring. Then the parameter \( (\beta, S) \) is identifiable, provided \( \tau > 0 \) such that \( \tau = \sup_{\tau > 0} S_{F_Y}(t_0) > 0 \).

Lemma 2: Assume \( h(t|z)=h_z(t)e^{\beta z} \). Under the mixed case IC model and assuming that \( S \) is absolutely continuous, the parameter \( \beta \) is identifiable if
\[
3a,b \in (S_{\tau_0} \cup S_{\tau_0}) \cap \{x|a(x) \}
\]

The parameter \( S_{\tau_0} \) is identifiable if \( \beta \neq 0 \) in addition to assumption (2.2). If assumption (2.2) is violated, \( \beta \) is not identifiable, as is the case in the next example.

Example 1. Assume \( h(t|z)=h_z(t)e^{\beta z} \). Let \( Z \sim \text{bin}(1,0.5) \). Suppose that \( S_{\tau_0} \in (0,1) \) on \( (0,4) \). Moreover, assume the Case 2 model, that is, the observable random vector is \( (L,R)=(-\infty,U) \cap (V,U]\cap (V,F) \cap (V,F) \cap (V,U) \cap (V,U) \cap (V,F) \cap (V,F) \), where the censoring vector \( (U,F)=\{1,3\} \) and So be absolutely continuous, where
\[
S_{\tau_0}(1) > S_{\tau_0}(2) > S_{\tau_0}(3) > S_{\tau_0}(4) = 0
\]

Then is not identifiable. The proof is given in the Appendix.

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 $S_0$ assigns weight to the cut point $c$ under the IC model, but not under the RC model unless there exists an exact observation at $c$.

Let $A_1, \ldots, A_n$ be all the innermost intervals induced by $I_i^*$ if the covariates are time independent, it is well known that in order to maximize $L$, it suffices to put the weights of $S_0$ to the right-end points of the IIs. Let $t_{j+1}^*$ be the right-end point of the II's or $c$, or $\pm \infty$, $t_j = -\infty < t_1 < \cdots < t_n = c < t_{n+1} < \cdots < t_m = \infty$ and . Write $S_i = S_i(t_i)$. For each let $(t_i, t_{i+1})$

$$
\begin{cases}
\text{tri} \leq R_i < t_{i+1}, \\
\text{tri} \leq R_i < t_{i+1} \quad \text{if} \quad L_i < R_i < \infty
\end{cases}
$$
satisfy

$$
\begin{cases}
\text{tri} = t_m \quad \text{and} \quad t_{i+1} = L_i \quad \text{if} \quad L_i < R_i = \infty
\end{cases}
$$

Theorem 1

Suppose that $h(t, z) = h(t) e^{\beta z} (t \geq c)$, $h$ is continuous and subject to the mixed case IC model, $L(k) < \infty$, and the identifiable condition in Lemma 2 is satisfied. Then the SMLE of is consistent.

Proof. We shall give the proof in 4 steps. Abusing notation, write $S_0^{(i)}(t) = S(t | u)$ and $S_0^{(0)}(t) = S_0(t)$ . Let $\Omega$ be the sample space.

Step 1: (preliminary). Under the mixed interval censorship model, by (14), the normalized generalized log-likelihood becomes $L_n(S, b)$

$$
= \frac{1}{n} \sum_{i=1}^n \log \left( \frac{S_0^{(i)}(L_i)}{S_0^{(i)}(R_i)} \right) + \frac{1}{n} \sum_{j=1}^n \log \left( \frac{S_0^{(j)}(L_j)}{S_0^{(j)}(R_j)} \right)
$$

where $C$ is the collection of all nonincreasing functions $S$ from $[0, \infty)$ into $[0, 1]$ with $S(0) = 1$ and $S(\infty) = 0$. By the strong law of large numbers (SLLN), $L_n(S, b)$ converges almost surely to its mean $L(S, b) = E(\log(S_0^{(i)}(L_i) / S_0^{(i)}(R_i))) = E(E \left( \log \left( \frac{S_0^{(i)}(L_i)}{S_0^{(i)}(R_i)} \right) | Z \right)) | K)$, where

$$
\begin{align}
&w_{x_i} \left( C_{k}, k \right) = (1 - s_{0} \left( C_{k} \right) ) \log(1 - s_{0} \left( C_{k} \right) ) + s_{0} \left( C_{k} \right) \log s_{0} \left( C_{k} \right), \\
&\kappa = \sum_{k=1}^{n} \left( s_{0} \left( C_{k} \right) \right) - \sum_{k=1}^{n} \left( C_{k} \right) \log s_{0} \left( C_{k} \right) + \log \left( \frac{\kappa}{\kappa} \right)
\end{align}
$$

Step 2: It can be verified that $w_{x_i} \left( C_{k}, k \right)$ is maximized by a nonincreasing function $S_0^{(i)}(t) \in C$, if $S_0^{(i)}(t) \in C$, $c \in \left[ L_i, R_i \right]$ . Since $\sup \left\{ \log p : 0 \leq p \leq 1 \right\} \leq 1$, $w_{x_i}^{(0)}(C, K)$ is bounded by $K + 1$, and thus $L(S, b)$ is finite, as $\mathbb{E}K = \infty$ under the assumption in the theorem. If the identifiable conditions hold, by Lemma 2 and the Shannon-Kolmogorov inequality, we can conclude that $(S_0^{(i)}(t), S_0^{(i)}(t)) = (S_0(t), S_0(t)) \forall t \in S_i \cup S_j$ and $\forall u \in S_i$. As a consequence, for some and

$$
\frac{1}{n} \sum_{i=1}^n \log \left( \frac{S_0^{(i)}(L_i)}{S_0^{(i)}(R_i)} \right) + \frac{1}{n} \sum_{j=1}^n \log \left( \frac{S_0^{(j)}(L_j)}{S_0^{(j)}(R_j)} \right)
$$

where $c < t_1 < t_2 < \tau$ and $t_1, t_2 \in S_i \cup S_j$. If the identifiable conditions hold, by Lemma 2 and the Shannon-Kolmogorov inequality, we can conclude that $(S_0^{(i)}(t), S_0^{(i)}(t)) = (S_0(t), S_0(t)) \forall t \in S_i \cup S_j$ and $\forall u \in S_i$. As a consequence, for some and

$$
\frac{1}{n} \sum_{i=1}^n \log \left( \frac{S_0^{(i)}(L_i)}{S_0^{(i)}(R_i)} \right) + \frac{1}{n} \sum_{j=1}^n \log \left( \frac{S_0^{(j)}(L_j)}{S_0^{(j)}(R_j)} \right)
$$

Thus $\beta = \beta$. Consequently, $(S_0, \beta)$ maximizes $L(S, b)$ and any other nonincreasing function $S \in C$ and $b$ satisfying $L(S, b) = L(S_0, \beta)$ satisfy $S = S_\beta$ a.s.μ (the measure induces by dFL+dFR) and $b = \beta$.

Step 3: $\lim \inf_{n \to \infty} L_n(S, b) = \lim \inf_{n \to \infty} L_n(S, b) = L(S, \beta)$. Let $\Omega_0 = \{ \omega \in \Omega : L_0(S_0(\omega)) = L(S_0(\omega)) \}$. Then $P(\Omega_0) = 1$ by the SLLN. Hereafter, we fix an $\omega \in \Omega$ and suppress it in the expressions of most random variables. For $n > 0$, let $B_n(\omega)$ be the collection of all the distinct points $0, L_0, R_0, c$, where $1 \leq i \leq n$. Write $B_n = \{ q_{\omega,i} : 1 \leq j \leq m_n \}$, where $0 = q_{\omega,0} = q_{\omega,1} = \cdots = q_{\omega,n} = \infty$. Denote the intervals $A_{\omega,j} = ( q_{\omega,j-1}, q_{\omega,j} \}$, $1 \leq j \leq m_n$. For each $j$, let $p_{\omega,i,j} = S_{\omega}(q_{\omega,j-1}) - S_{\omega}(q_{\omega,j})$. Then $\sum_{j=1}^{m_{n,i}} p_{\omega,i,j} = 1$ and $S_{\omega}(t) = \sum_{j=1}^{m_{n,i}} p_{\omega,i,j}$ for each $t \in B_n$. Moreover, the normalized log-likelihood function with $S = S_0$ is $L(S, b)(\omega)$.
x and some sequence \( \{n^i\}_{i \in \mathbb{N}} \). Let \( S^{(n^i)}(t) \) be the pointwise limit function of \( S_n^{(i)}(t) \) for all \( i \) and for some subsequence \( \{n^i\}_{i \in \mathbb{N}} \). Helly’s selection theorem guarantees the existence of pointwise limits. Let \( B^* \) be the limiting point of \( \{B^*_n\} \) for some subsequence \( \{n^i\}_{i \in \mathbb{N}} \) of \( \{n^i\}_{i \in \mathbb{N}} \).

Since \( L(S_n, B^*) \geq L(S_n, B) \) by the definition of the GMLE, the claim in Step 3 is proved.

Step 4 (Conclusion). Let \( \hat{Q} \) denote the empirical estimator of \( Q \), the distribution of \((L, R, Z)\) and \( \alpha = \{a \in \Omega: \hat{Q}((1, r, z)) - Q((1, r, z)) \text{ pointwise in } (1, r, z)\} \). By the SLNN \( \rho(\alpha) = 1_{\Omega} - \hat{Q}(\Omega) - Q(\Omega) \), a.s. for every Borel subset \( U \) of \( \Omega = \{(1, r, z): 0 < l < r \leq \infty, \omega \in S_n\} \). Let \( S_n^* \) denote the survival function defined by \( S_n^*(x) = S_n(x; \omega), h_n \), defined by \( h_n = \hat{h}_n(\alpha) \) and \( \hat{Q} \), the measure defined by \( \hat{Q}(A) = \hat{Q}(A; \omega) \). For simplicity in notation we shall assume that \( S_n(x) = S^*(x) \) for all \( x \in R \) and \( B \to B^* \). Let \( \omega = \Omega \cap \Omega \) hereafter. Let \( \liminf_{n \to \infty} L(S_n, B^*) \geq \limsup_{n \to \infty} L(S_n, B) \) for all \( \omega \in R \) and \( B \to B^* \). We shall show that \( L(S_n, B) \leq \liminf_{n \to \infty} L(S_n, B^*) \). \( \hat{Q}(A) \) (2.3)

By the previous discussion, it suffices to prove the last inequality.

Now let \( S_n^*(t) = S_n(t)e^{-\lambda_n^u(t/\gamma_n)} \). Since \( L(S_n^*, B^*) = \int \log(S_n^*(t) - S_n^*(r)dQ(l, r, u) \),

the desired inequality is thus equivalent to

\[
\limsup_{n \to \infty} \log |S_n^*(t) - S_n^*(r)|dQ(l, r, u) \leq \log |S^*(t) - S^*(r)|dQ(l, r, u)
\]

which follows from Lemma 3. It follows from inequality (2.3) that \( L(S_n, B^*) \geq L(S_n, B) \). As \( (S_n, B) \) maximizes \( L \), we can conclude that \( S^* = S_n \), a.s. \( M \) if the identifiable conditions (2.2) hold, we have \( B^* = B \).

Lemma 3. Inequality (2.4) holds.

In order to prove the Lemma 3, we will introduce the Fatou’s Lemma with varying measures.

**Theorem 2.**

Suppose that \( M_\mu \) is a sequence of measures on the measurable space \( (S, \Sigma) \) such that \( M_\mu(B) \to M_\mu(B) \) for all \( B \in \Sigma \). Then, with \( f_\mu \) non-negative integrable functions, we have \( f = \liminf_{n \to \infty} f_\mu \).

Proof of Theorem 2: We will prove something a bit stronger here. Namely, we will allow \( f \) to converge \( \mu \) almost everywhere on a subset \( B \) of \( S \). We seek to show that \( \int_B f d\mu \leq \liminf_{n \to \infty} \int_B f_\mu d\mu \).

Recall that a simple function \( \phi \) is of the form that \( \phi(x) = \sum_{i=1}^n a_i \chi_{A_i}(x) \), where \( \chi_{A_i} \) are disjoint measurable sets. Given a simple function \( \phi \) we have \( \int_B \phi d\mu = \lim_{n \to \infty} \int_B \phi d\mu_n \).

Hence, by the definition of the LebesgueIntegral, it is enough to show that if \( \phi \) is any nonnegative simple function less than or equal to \( f \), then \( \int_B \phi d\mu \leq \liminf_{n \to \infty} \int_B f_\mu d\mu_n \).

Let \( A \) be the minimum non-negative value of \( \phi \). Define \( A = \{x \in B: \phi(x) < A\} \). We first consider the case when \( \int_B \phi d\mu = \infty \).

Then \( \mu(A) = \infty \) too. Thus, \( \mu(A) = \infty \).

Thus, \( \mu(A) = \infty \).

At the same time, \( \int_B f d\mu_n \to \mu(A) = \infty \). Thus, \( \int_B f d\mu_n \to \infty \). It suffices to prove the inequality in the case \( \mu(A) = \infty \).

To prove the inequality in the case \( \mu(A) = \infty \). We must have that \( \mu(A) = \infty \). Denote, as above, by \( M \) the maximum value of \( \phi \) and fix \( \epsilon > 0 \).

Define \( A = \{x \in B: f(x) < (1-\epsilon)\phi(x) \} \). Then \( A \) is a nested increasing sequence of sets whose union contains.

Thus, \( \mu(A) = \infty \).

A is a decreasing sequence of sets with empty intersection. Since \( A \) has finite measure (this is why we needed to consider the two separate cases), \( \lim_{n \to \infty} \mu(A) = 0 \).

Thus, there exists \( n \) such that \( \mu(A - Ak) = \epsilon \), for all \( k \geq n \). Since \( \lim_{n \to \infty} \mu(A - Ak) = \epsilon \), there exists \( n \) such that \( \mu(A - Ak) = \epsilon \).

Hence, for \( k \geq n \),

\[
\int_B f d\mu_n \geq \int_{Ak} f d\mu_n \geq (1-\epsilon)\int_{Ak} \phi d\mu_n
\]

At the same time, \( \int_B \phi d\mu_n = \int_A \phi d\mu_n - \int_{Ak} \phi d\mu_n \).

Hence, \( (1-\epsilon) \int_{Ak} \phi d\mu_n \geq (1-\epsilon) \int_B \phi d\mu_n \).

These inequalities yield

\[
\int_B \phi d\mu_n \leq (1-\epsilon) \int_B \phi d\mu_n - \epsilon \int_B \phi d\mu_n - \epsilon \int_B \phi d\mu_n + \mu(M)
\]

Hence, \( \epsilon \to 0 \) letting and taking the limit inf in \( n \), we get that \( \lim_{n \to \infty} \int_B \phi d\mu_n \geq \int_B \phi d\mu \).

Now we give the proof for the Lemma 3.

Proof of Lemma 3: Since \( \lim_{n \to \infty} \log(S_n^*(t) - S_n^*(r)) = \log(S^*(t) - S^*(r)) \),

\[
\lim_{n \to \infty} \log(S_n^*(t) - S_n^*(r)) = \log(S^*(t) - S^*(r))
\]

and \( -\log(S_n^*(t) - S_n^*(r)) = \log(S_n^*(t) - S_n^*(r)) \).

Thus, an application of Theorem 2 yields

\[
\limsup_{n \to \infty} \log(S_n^*(t) - S_n^*(r))dQ(l, r, u)
\]

\[
-\limsup_{n \to \infty} \log(S_n^*(t) - S_n^*(r))dQ(l, r, u)
\]

\[
\leq -\limsup_{n \to \infty} \log(S_n^*(t) - S_n^*(r))dQ(l, r, u)
\]
∫ log(S^u(r)^r) dQ(l, r, u)

**Theorem 3**

Suppose that the assumptions in Theorem 1 holds and the support set S_F ∪ S_z ∪ S_{z'} contains finitely many elements. Then the SMLE of (S, β) is asymptotically normally distributed.

**Proof:** By assumption S_F ∪ S_z ∪ S_{z'} = S_F' = 0 and m is finite. Then the parameter (S, β) can be represented by (S(t), ..., S(t), β), and the problem becomes an estimation problem of a multinomial distribution subject to certain constraints. Thus the asymptotic normality follows and the asymptotic covariance matrix can be estimated by the inverse of the empirical Fisher information matrix.

**Appendix**

**Proof of Example 1**

Let

s_0 = 1(Y ≤ 1 | Z = 0) + 1(Y ≤ 1 | Z = 1),

s_2 = 1(Y ∈ (1, 3] | Z = 0),

s_3 = 1(Y > 3 | Z = 1),

s_4 = 1(Y > 3 | Z = 0).

Let

p_i = F_i(1), p_i = F_i(2) − F_i(1), p_i = F_i(3), and p_i = F_i(2) − F_i(1).

Abusing notations, let u denote both the random variable and the realization. The joint density is

f = [(p_i)^y (p_i/p_i)^y (p_i/p_i)^y] (1-p_i)^y (p_i/p_i)^y (p_i/p_i)^y (1/2)

For given (p_i, p_i, p_i, β) = (p_i, p_i, p_i, β), let γ = (p_i + p_i)^y p_i^y

then remains the same if (p_i, p_i, p_i, β) = (p_i, p_i, p_i, β), where (p_i, β) satisfies

(p_i + p_i)^y p_i^y = γ^y.

The latter equation yields

β = ln p_i + p_i

where

p_i = (0, 1, p_i, p_i). Thus the β is not uniquely determined if p_i + p_i < 1. For instance, let then γ = (p_i + p_i)^y p_i^y = 0.12. Thus β = β(p_i) in Eq. (B.1) is well defined for p_i in a neighborhood of 1/8 (actually, for p_i in (0, 1/8)). Hence, the parameter β is not identifiable.

**References**

1. Wong GYC, Diao QG and Yu QQ (2017). Piece-wise Cox Models with interval censored data.
2. Schick A, Yu QQ (2000) Consistency of the GMLE with mixed case interval-censored data. Scand J Statist 27(1): 45-55.
3. Greneboom P and Wellner JA (1992) Information bounds and nonparametric maximum likelihood estimation. Birkhauser Verlag, Basel.
4. Cox DR and Oakes D (1984) Analysis of Survival Data. Chapman & Hall NY.
5. Therneau T, Grambsch P (2000) Modeling survival data: extending the Cox model. Springer.
6. Platt RW, Joseph KS, Ananth CV, Grendines J, Abrahamowicz M, et al. (2004) A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. Am J Epidemiol 160(3): 199-206.
7. Stephan L, Michael S (2007) Parsimonious analysis of time-dependent effects in the Cox model. Statistics in Medicine 26(13): 2668-2698.
8. Masaaki T, Masato S (2009) Analysis of survival data having time-dependent covariates. IEEE Trans Neural Netw 20(3): 389-394.
9. Leffondre K, Wynant W, Cao Z (2010) A weighted Cox model for modeling time-dependent exposures in the analysis of case-control studies. Stat Med 29(7-8): 839-850.
10. Zhou M (2001) Understanding the Cox regression models with time-change covariates. American Statistician 55(2): 153-155.
11. Wong GYC, Osborne MP, Diao QG, and Yu QQ (2016) Piece-wise Cox Models with right-censored data. Comm Statist Comput Simul 41(8): 1489-1500.
