Maximum likelihood estimation in the proportional hazards cure model

Wenbin Lu

Abstract The proportional hazards cure model generalizes Cox's proportional hazards model which allows that a proportion of study subjects may never experience the event of interest. Here nonparametric maximum likelihood approach is proposed to estimating the cumulative hazard and the regression parameters. The asymptotic properties of the resulting estimators are established using the modern empirical process theory. And the estimators for the regression parameters are shown to be semiparametric efficient.

Keywords Censoring · Maximum likelihood estimation · Mixture models · Proportional hazards cure model · Semiparametric efficiency

1 Introduction

Survival models with a cure rate have received much attention in recent years (Farewell, 1982, 1986; Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000, among others). These models are useful when a proportion of study subjects never experience the event of interest. Applications of cure models can be found in many disciplines, including biomedical sciences, economics, sociology and engineering science. Maller and Zhou (1996) contains a list of such applications.

Mixture modelling approach is commonly used to formulate a cure model, which assumes that the underlying population is a mixture of susceptible and nonsusceptible subjects. All susceptible subjects would eventually experience

W. Lu (✉)
Department of Statistics, North Carolina State University, 2501 Founders Drive,
Raleigh, NC 27695, USA
e-mail: wlu4@stat.ncsu.edu

the event if there were no censoring, while the nonsusceptible ones are immune from the event. Under mixture modelling approach, a decomposition of the event time is given by

$$T = \eta T^* + (1 - \eta)\infty, \tag{1}$$

where $T^* < \infty$ denotes the failure time of a susceptible subject and η indicates, by the value 1 or 0, whether the sampled subject is susceptible or not. Thus, one can model separately the survival distribution for susceptible individuals and the fraction of nonsusceptible ones.

Parametric mixture models were explored earlier by a number of authors. Berkson and Gage (1952) used a mixture of exponential distributions and a constant cure fraction to fit survival data from studies of breast cancer and stomach cancer. This parametric approach was extended by Farewell (1982, 1986) to Weibull regression for survival and logistic regression for the cure fraction. Theoretical and empirical properties of the Weibull extension were fully studied there via the parametric maximum likelihood method.

More recent attention has been focused on semiparametric mixture modelling approaches. Kuk and Chen (1992) proposed the so-called proportional hazards cure model in which the proportional hazards regression (Cox 1972) models the survival times of susceptible subjects while the logistic regression models the cure fraction.

The model is specified by the following two terms:

$$\lambda(t|Z, X) = \lim_{dt \to 0^+} P(t \le T^* < t + dt | T^* \ge t, Z, X)/dt = \lambda(t) \exp(\beta' Z), \tag{2}$$

$$P(\eta = 1|X, Z) = \frac{\exp(\gamma' X)}{1 + \exp(\gamma' X)}, \tag{3}$$

where $\lambda(t|Z, X)$ is the hazard function for a susceptible subject with p-dimensional covariates Z and q-dimensional covariates X, and $\lambda(t)$ the completely unspecified baseline hazard function, and β and γ the unknown regression parameter vectors of primary interest. Recall that Z and X may share some common components and X includes 1 so that γ contains the intercept term. Furthermore, we assume that the censoring time C is independent of T^* and η conditional on Z and X. Define $\tilde{T} = \min\{T, \min(C, \tau)\}$ and $\delta = I\{T \le \min(C, \tau)\}$, where τ denote the total follow-up of the study. Then the observations consist of $(\tilde{T}_i, \delta_i, Z_i, X_i)$, $i = 1, \ldots, n$, which are independent copies of $(\tilde{T}, \delta, Z, X)$. And the observed likelihood function can be written as

$$L_n(\Lambda, \theta) = \prod_{i=1}^n \left[\left\{ \pi(\gamma' X_i) \lambda(\tilde{T}_i) e^{\beta' Z_i} e^{-\Lambda(\tilde{T}_i) \exp(\beta' Z_i)} \right\}^{\delta_i} \right.$$

$$\left. \times \left\{ 1 - \pi(\gamma' X_i) + \pi(\gamma' X_i) e^{-\Lambda(\tilde{T}_i) \exp(\beta' Z_i)} \right\}^{1-\delta_i} \right], \tag{4}$$

where $\theta = (\beta', \gamma')'$, $\pi(a) = \exp(a)/\{1 + \exp(a)\}$ for any real number a, and $\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s$ is the baseline cumulative hazard function.

As noted by Kuk and Chen (1992), if the cure fraction $1 - \pi(\gamma'X)$ is not equal to zero, the hazard function of T is no longer proportional and the simple form of the partial likelihood function, like that for the usual Cox proportional hazards model, can not be obtained here. To solve this difficulty, they proposed to consider the following complete but unobserved likelihood function:

$$
L_{nc}(\Lambda, \theta) = \prod_{i=1}^{n} \Bigg[\left\{ \pi(\gamma'X_i)\lambda(\tilde{T}_i)\mathrm{e}^{\beta'Z_i}\mathrm{e}^{-\Lambda(\tilde{T}_i)\exp(\beta'Z_i)} \right\}^{\delta_i \eta_i}
$$

$$
\times \left\{ \left\{ 1 - \pi(\gamma'X_i) \right\}^{1-\eta_i} \left\{ \pi(\gamma'X_i)\mathrm{e}^{-\Lambda(\tilde{T}_i)\exp(\beta'Z_i)} \right\}^{\eta_i} \right\}^{1-\delta_i} \Bigg],
$$

where η_i's are only partially observed, i.e. when $\delta_i = 1$, $\eta_i = 1$, while when $\delta_i = 0$, η_i is unobserved. Based on L_{nc}, they developed a Monte Carlo simulation-based algorithm for approximating a rank-based likelihood function, thereby enabling them to perform maximum marginal likelihood estimation. The proportional hazards cure model was further studied by Peng and Dear (2000) and Sy and Taylor (2000) among others to obtain alternative methods for computing the joint semiparametric likelihood function. Their approaches largely relied on the semiparametric EM algorithm that computes estimates for both the cumulative baseline hazard and the regression parameters. However, the theoretical properties of the resulting estimators for the proportional hazards cure model remain to be established. Fang et al. (2005) studied the large sample properties of the maximum likelihood estimators under the proportional hazards cure model.

Recently, Lu and Ying (2004) proposed an estimating equations approach for the semiparametric transformation cure models, where the class of linear transformation models are used for the failure times of susceptible subjects and the logistic regression is used for the cure fraction. Their approach was motivated by the work of Chen et al. (2002) and used the martingale integral representation to construct unbiased estimating equations. The large sample properties of the resulting estimators were also studied. However, the proposed algorithm for solving the equations may not converge and the resulting estimators for the regression parameters are not efficient, even when the model specifies the proportional hazards cure model, i.e. the error term of the linear transformation models follows the extreme value distribution.

In this paper, we propose to estimate the parametric and the nonparametric components in the proportional hazards cure model by using nonparametric maximum likelihood. The joint parametric/nonparametric likelihood approach to semiparametric problems was pioneered by Murphy (1994, 1995) for the frailty model and Scharfsten et al. (1998) for the generalized odds-rate class of regression models, among others. Here, we apply and extend their techniques to the proportional hazards cure model and make use of the modern empirical

process theory, as elucidated in van der Vaart and Wellner (1996), to drive the large sample properties of the resulting estimators. In addition, We show that the maximum likelihood estimators for the regression parameters are semiparametrically efficient (Bickel et al., 1993).

The paper is organized in a natural order. We first compute the semiparametric variance bound (Sect. 2). Then, in Sect. 3, we define our estimators and show that they are consistent, asymptotically normal, and achieve the variance bound. Consistent variance estimators are also derived here. Section 4 gives some discussions on generalization of our approach to include time dependent covariates and other semiparametric cure models. Some technical lemmas are put together and proved in the Appendix.

2 Semiparametric variance bound

First, note that the log likelihood function for a single observation is

$$l(\Lambda, \theta) = \delta[\log\{\pi(\gamma'X)\lambda(\tilde{T})\} + \beta'Z - \Lambda(\tilde{T})\exp(\beta'Z)] + (1-\delta)\log S(\tilde{T}, \Lambda, \theta), \tag{5}$$

where $S(t, \Lambda, \theta) = 1 - \pi(\gamma'X) + \pi(\gamma'X)\exp\{-\Lambda(t)\exp(\beta'Z)\}$. Then the semiparametric variance bound is computed via the semiparametric efficiency theory of Bickel et al. (1993). To be specific, a parametric submodel corresponding to a parameterization of $\lambda(t)$ is considered, say $\lambda(t, \alpha)$, where $\lambda(t, \alpha_0) = \lambda(t)$ for some α_0.

So the log likelihood for such parametric submodel is given by

$$\begin{aligned} l(\alpha, \theta) = \delta[\log\{\pi(\gamma'X)\lambda(\tilde{T}, \alpha)\} + \beta'Z - \Lambda(\tilde{T}, \alpha)\exp(\beta'Z)] \\ + (1-\delta)\log S\{\tilde{T}, \Lambda(\cdot, \alpha), \theta\}, \end{aligned}$$

where $\Lambda(t, \alpha) = \int_0^t \lambda(s, \alpha)\mathrm{d}s$. The score for θ is

$$\frac{\partial l}{\partial \theta} = \int_0^\tau W\{t, \Lambda(\cdot, \alpha), \theta\}\mathrm{d}M\{t, \Lambda(\cdot, \alpha), \theta\}, \tag{6}$$

where $M(t, \Lambda, \theta) = N(t) - \int_0^t Y(s)g(s, \Lambda, \theta)\exp(\beta'Z)\mathrm{d}\Lambda(s)$ and $W(t, \Lambda, \theta) = (Z'[1 - \{1 - g(t, \Lambda, \theta)\}\exp(\beta'Z)\Lambda(t)], X'\{1 - g(t, \Lambda, \theta)\})'$ with

$$g(t, \Lambda, \theta) = \frac{\pi(\gamma'X)\exp\{-\Lambda(t)\exp(\beta'Z)\}}{1 - \pi(\gamma'X) + \pi(\gamma'X)\exp\{-\Lambda(t)\exp(\beta'Z)\}}.$$

By the usual counting process and its associated martingale theory (Fleming and Harrington 1991; Andersen et al. 1993), $M(t, \Lambda, \theta)$ is the $\mathcal{F}_t$-counting process martingale, where $\mathcal{F}_t$ is the smallest sigma-algebra generated by $\{N(s), Y(s), 0 \le s \le t\}$. The score for α is

$$\frac{\partial l}{\partial \alpha} = \int_0^\tau \left[\frac{\frac{\partial}{\partial \alpha}\lambda(t,\alpha)}{\lambda(t,\alpha)} - \{1 - g(t,\Lambda,\theta)\} \exp(\beta'Z) \int_0^t \frac{\partial \lambda(s,\alpha)}{\partial \alpha} ds \right] dM\{t, \Lambda(\cdot,\alpha), \theta\}. \tag{7}$$

Let θ_0 and Λ_0 be the true values of θ and Λ, respectively. And denote S_θ and S_α to be the corresponding scores for θ and α evaluated at the truth. Similar to Scharfsten et al. (1998), we can define the following tangent set Θ in the nonparametric direction,

$$\begin{aligned}\Theta = \{f(\tilde{T}, \delta, Z, X) : f(\tilde{T}, \delta, Z, X) \\ = \int_0^\tau [a(t) - \{1 - g_0(t)\} \exp(\beta'_0 Z) \int_0^t a(s) d\Lambda_0(s)] dM_0(t),\end{aligned}$$

where $a(t)$ is any $(p+q)-$dimensional function of t, $E[||f(\tilde{T}, \delta, Z, X)||^2] < \infty\}$.

Here $g_0(t) = g(t, \Lambda_0, \theta_0)$ and $M_0(t) = M(t, \Lambda_0, \theta_0)$. Then the efficient score S_{eff} for θ can be defined as $S_{\text{eff}} = S_\theta - \Pi[S_\theta|\Theta]$, where $\Pi[\cdot|\cdot]$ is the projection operator.

To project S_θ onto Θ, we need to find the vector $a(t)$ such that

$$\begin{aligned}E \left(\int_0^\tau [W_0(t) - a(t) + \{1 - g_0(t)\} \exp(\beta'_0 Z) \int_0^t a(s) d\Lambda_0(s)]' dM_0(t) \right. \\ \left. \times \int_0^\tau [a^*(t) - \{1 - g_0(t)\} \exp(\beta'_0 Z) \int_0^t a^*(s) d\Lambda_0(s)] dM_0(t) \right) = 0, \tag{8}\end{aligned}$$

for all a^*. Here $W_0(t) = W(t, \Lambda_0, \theta_0)$. By some simple algebra, we can show that the vector $a(t)$ which satisfies (8) is a solution to the following Fredholm integral equation of the second kind:

$$a(t) - \int_0^\tau K(t,s) a(s) d\Lambda_0(s) = f(t), \quad t \in [0, \tau], \tag{9}$$

where for $0 \le t, s \le \tau$,

$$\begin{aligned}K(t,s) = & \frac{E[g_0(s \vee t)\{1 - g_0(s \vee t)\} \exp(2\beta'_0 Z) Y(s \vee t)]}{E\{Y(t)g_0(t) \exp(\beta'_0 Z)\}} \\ & - \frac{\int_{s \vee t}^\tau E[g_0(u)\{1 - g_0(u)\}^2 \exp(3\beta'_0 Z) Y(u)] d\Lambda_0(u)}{E\{Y(t)g_0(t) \exp(\beta'_0 Z)\}},\end{aligned}$$

$$\begin{aligned}f(t) = & \frac{E\{W_0(t) Y(t) g_0(t) \exp(\beta'_0 Z)\}}{E\{Y(t)g_0(t) \exp(\beta'_0 Z)\}} \\ & - \frac{\int_t^\tau E[W_0(s) g_0(s)\{1 - g_0(s)\} \exp(2\beta'_0 Z) Y(s)] d\Lambda_0(s)}{E\{Y(t)g_0(t) \exp(\beta'_0 Z)\}}.\end{aligned}$$

If the kernel $K(\cdot, \cdot)$ satisfies $\sup_{t \in [0,\tau]} \int_0^\tau |K(t,s)| \mathrm{d}\Lambda_0(s) < \infty$, then according to Kress (1989), there exists a solution to this integral equation. Under mild regularity conditions, which are given in the next section, we can show that the kernel K defined above satisfies the above condition. Denote the solution by $a_{\mathrm{eff}}(t)$. Then, the efficient score for θ is

$$S_{\mathrm{eff}} = \int_0^\tau [W_0(t) - a_{\mathrm{eff}}(t) + \{1 - g_0(t)\} \exp(\beta_0'Z) \int_0^t a_{\mathrm{eff}}(s)\mathrm{d}\Lambda_0(s)]\mathrm{d}M_0(t)$$

Therefore, provided that the $E(S_{\mathrm{eff}}S_{\mathrm{eff}}')$ is nonsingular, the semiparametric variance bound, Φ, is $\{E(S_{\mathrm{eff}}S_{\mathrm{eff}}')\}^{-1}$. Furthermore, we can show that

$$E(S_{\mathrm{eff}}S_{\mathrm{eff}}') = E\left(\int_0^\tau W_0(t)[W_0(t) - a_{\mathrm{eff}}(t) + \{1 - g_0(t)\} \exp(\beta_0'Z) \int_0^t a_{\mathrm{eff}}(s)\mathrm{d}\Lambda_0(s)]' \right.$$
$$\left. \times Y(t)g_0(t) \exp(\beta_0'Z)\mathrm{d}\Lambda_0(t) \right),$$

since based on (8),

$$E\left(\int_0^\tau [W_0(t) - a_{\mathrm{eff}}(t) + \{1 - g_0(t)\} \exp(\beta_0'Z) \int_0^t a_{\mathrm{eff}}(s)\mathrm{d}\Lambda_0(s)]' \right.$$
$$\left. \times [a_{\mathrm{eff}}(t) - \{1 - g_0(t)\} \exp(\beta_0'Z) \int_0^t a_{\mathrm{eff}}(s)\mathrm{d}\Lambda_0(s)]Y(t)g_0(t) \exp(\beta_0'Z)\mathrm{d}\Lambda_0(t) \right) = 0.$$

3 Estimation and asymptotic

Nonparametric maximum likelihood method is used to estimate the regression parameters and the baseline cumulative hazard function. For convenience, we assume that there are no tied death times, but our results can be easily tailored to accommodate tied death times.

The log likelihood function for observed data is given by

$$\sum_{i=1}^n \left(\delta_i[\log\{\pi(\gamma'X_i)\} + \log\lambda(\tilde{T}_i) + \beta'Z_i - \Lambda(\tilde{T}_i)\exp(\beta'Z_i)] \right.$$
$$\left. + (1 - \delta_i)\log S_i(\tilde{T}_i, \Lambda, \theta) \right), \tag{10}$$

where S_i is obtained from S by replacing Z and X by Z_i and X_i, respectively. The maximum of this function does not exists if $\Lambda(\cdot)$ is restricted to be absolutely continuous. Thus, we allow $\Lambda(\cdot)$ to be discrete and replace $\lambda(t)$ in (10) with the jump size of Λ at time t, denoted by $\Delta\Lambda(t)$. Then we maximize the following modified log likelihood function to obtain our estimates:

$$nl_n(\Lambda, \theta) = \sum_{i=1}^{n} \Big(\delta_i[\log\{\pi(\gamma'X_i)\} + \log\{\Delta\Lambda(\tilde{T}_i)\} + \beta'Z_i - \Lambda(\tilde{T}_i)\exp(\beta'Z_i)]$$
$$+ (1 - \delta_i)\log S_i(\tilde{T}_i, \Lambda, \theta)\Big). \tag{11}$$

We observe that the maximizer $\hat{\Lambda}_n$ of (11) must be a step function with jumps at all the death times. Within the class of functions of this form, we can show that the maximum likelihood estimators, $\hat{\Lambda}_n$ and $\hat{\theta}_n$, of (11) exist and are finite. Then, we show that they are consistent and asymptotically normal. And the limiting variance of $\hat{\theta}_n$ attains the semiparametric efficiency bound, Φ. To establish our claims, we need the following conditions:

Condition 1. The function $\Lambda_0(t)$ is strictly increasing and continuously differentiable, and $\Lambda_0(\tau) < \infty$.

Condition 2. θ_0 lies in the interior of a compact set $\mathcal{C}$ and the covariate vectors Z and X are bounded in the sense that $P(|Z| < m$ and $|X| < m) = 1$ for some constant $m > 0$.

Condition 3. With probability one, there exists a positive constant ε such that $P(C \geq T^* \geq \tau | Z, X) > \varepsilon$.

Condition 4. $P\{Y(t) = 1 | Z, X\}$ is continuous in t.

Remark 1 The conditions 1, 2 and 4 are the usual regularity conditions needed for establishing the large sample results for the maximum partial likelihood estimators under the proportional hazards model. Condition 3 is for the identifiability of the proportional hazards cure model on the interval $[0, \tau]$, i.e. the follow-up is sufficiently long for identifying $\Lambda(t)$ on the interval $[0, \tau]$.

3.1 Existence

Theorem 1 *Assume that conditions* 1–2 *hold. Then the maximum likelihood estimators of $l_n(\Lambda, \theta)$, $(\Lambda, \theta) = (\hat{\Lambda}_n, \hat{\theta}_n)$ exists and is finite.*

Proof Since l_n is a continuous function of θ and the jump sizes of Λ, it is equivalent to show that the jump sizes are finite (Scharfsten et al., 1998). To see this, let $b_1, \ldots, b_{k(n)}$ denote the jump sizes at the death times $\tilde{T}_{(1)} < \tilde{T}_{(2)} < \cdots < \tilde{T}_{(k(n))}$, where $k(n)$ is the total number of deaths. Then

$$l_n < \frac{1}{n} \sum_{i=1}^{k(n)} \left[\log\{\pi(\gamma'X_{(i)})\} + \beta'Z_{(i)} + \{\log b_i - \exp(\beta'Z_{(i)})\sum_{j=1}^{i} b_j\} \right]$$
$$< \frac{1}{n} \sum_{i=1}^{k(n)} \left[\log\{\pi(\gamma'X_{(i)})\} + \beta'Z_{(i)} + \{\log b_i - \exp(-M_0)\sum_{j=1}^{i} b_j\} \right]$$

where M_0 is a positive constant and $Z_{(i)}, X_{(i)}$ are the covariate vectors corresponding to the ith death time, $\tilde{T}_{(i)}$. The first inequality above holds since

$0 < S_i(\tilde{T}_i, \Lambda, \theta) < 1$ and the second inequality is due to condition 2. Therefore, l_n diverges to $-\infty$ if b_j tends to ∞ for some j. This implies that the jump sizes of Λ must be finite.   □

Remark 2 The maximization of $l_n(\Lambda, \theta)$ in (11) can be carried out by the Nelder-Mead simplex method. Such an algorithm is available in MATLAB software with the build-in function "fminsearch".

Remark 3 Since $(\hat{\Lambda}_n, \hat{\theta}_n)$ exists and is finite, the derivative of l_n with respect to the jump sizes of Λ should be zero. This leads to the following equation for $\hat{\Lambda}_n$:

$$
\hat{\Lambda}_n(t) = \int_0^t \frac{\mathrm{d}\bar{N}(s)}{\frac{1}{n}\sum_{j=1}^n Y_j(s)\exp(\hat{\beta}'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_n, \hat{\theta})\}}, \quad 0 < t \leq \tau,
\tag{12}
$$

where $\bar{N}(t) = (1/n)\sum_{i=1}^n N_i(t)$ and g_i is obtained from g by replacing Z and X by Z_i and X_i, respectively.

3.2 Consistency

We apply the techniques used by Murphy (1994) and Scharfsten et al. (1998) here for proving consistency. Specifically, define

$$
\tilde{\Lambda}_n(t) = \int_0^t \frac{\mathrm{d}\bar{N}(s)}{\frac{1}{n}\sum_{j=1}^n Y_j(s)\exp(\beta_0'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j, \Lambda_0, \theta_0)\}}, \quad 0 < t \leq \tau,
\tag{13}
$$

which is a step function with jumps at each of the death times and converges uniformly to Λ_0 (see Lemma 2 of the Appendix).

Theorem 2 *Assume that conditions 1–4 hold. Then*

$$
\sup_{t\in[0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0 \ \ a.s. \quad and \quad ||\hat{\theta}_n - \theta_0|| \to 0 \ \ a.s.
$$

The proof of Theorem 2 is given in Appendix.

3.3 Asymptotic distribution

We extend the approach of Murphy (1995) and Scharfsten et al. (1998) to derive the asymptotic distribution of our estimators $(\hat{\Lambda}_n, \hat{\theta}_n)$. Here we also work with one-dimensional submodels through the estimators and the difference at the estimators. To be specific, set $\Lambda_d(t) = \int_0^t\{1 + dh_1(s)\mathrm{d}\hat{\Lambda}_n(s)$ and $\theta_d = dh_2 + \hat{\theta}_n$, where h_1 is a function and h_2 is a $(p+q)$-dimensional vector. Furthermore, write $h_2 = (h_{21}', h_{22}')'$, where h_{21} is the p-dimensional and

h_{22} is the q-dimensional vectors corresponding to Z and X, respectively. Let $S_n(\hat{\Lambda}_n, \hat{\theta}_n)(h_1, h_2)$ denote the derivative of l_n with respect to d and evaluated at $d = 0$. If $(\hat{\Lambda}_n, \hat{\theta}_n)$ maximizes l_n, then $S_n(\hat{\Lambda}_n, \hat{\theta}_n)(h_1, h_2) = 0$ for all (h_1, h_2). We observe that $S_n(\hat{\Lambda}_n, \hat{\theta}_n)(h_1, h_2) = S_{n1}(\hat{\Lambda}_n, \hat{\theta}_n)(h_1) + S_{n2}(\hat{\Lambda}_n, \hat{\theta}_n)(h_2)$, where

$$
S_{n1}(\hat{\Lambda}_n, \hat{\theta}_n)(h_1) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left[h_1(t) - \{1 - g_i(t, \hat{\Lambda}_n, \hat{\theta}_n)\} \exp(\hat{\beta}' Z_i) \int_0^t h_1(s) \mathrm{d}\hat{\Lambda}(s) \right]
$$
$$
\times \left\{ \mathrm{d}N_i(t) - Y_i(t)g_i(t, \hat{\Lambda}_n, \hat{\theta}_n) \exp(\hat{\beta}' Z_i) \mathrm{d}\hat{\Lambda}(t) \right\},
$$
$$
S_{n2}(\hat{\Lambda}_n, \hat{\theta}_n)(h_2) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau h_2' W_i(t, \hat{\Lambda}_n, \hat{\theta}_n)
$$
$$
\times \{ \mathrm{d}N_i(t) - Y_i(t)g_i(t, \hat{\Lambda}_n, \hat{\theta}_n) \exp(\hat{\beta}' Z_i) \mathrm{d}\hat{\Lambda}(t) \}.
$$

Let $BV[0, \tau]$ denote the space of bounded variation functions defined on $[0, \tau]$. We assume that the class of h belongs to the space $H = BV[0, \tau] \bigotimes R^{p+q}$. For $h \in H$, we define the norm on H to be $||h||_H = ||h_1||_v + |h_2|_1$, where $||h_1||_v$ is the absolute value of $h_1(0)$ plus the total variation of h_1 on the interval $[0, \tau]$ and $|h_2|_1$ is the L_1-norm of h_2. Define $H_m = \{h \in H : ||h||_H \le m\}$. If $m = \infty$, then the inequality is strict. In addition, define $\langle \Lambda, \theta \rangle(h) = \int_0^\tau h_1(t)\mathrm{d}\Lambda(t) + h_2'\theta$. The $\langle \Lambda, \theta \rangle$ indexes the space functionals

$$
\Psi = \left\{ \langle \Lambda, \theta \rangle : \sup_{h \in H_m} |\langle \Lambda, \theta \rangle| < \infty \right\}.
$$

Now $\Psi \subset l^\infty(H_m)$, where $l^\infty(H_m)$ is the space of bounded real-valued functions on H_m under the supremum norm $||U|| = \sup_{h \in H_m} |U(h)|$. The score function S_n is a random map Ψ to $l^\infty(H_m)$ for all finite m. Convergence in probability (denoted by $\mathcal{P}^*$) and weak convergence will be in terms of outer measure.

Theorem 3 *Assume that conditions* 1–4 *hold. Then*

$$
\langle \sqrt{n}(\hat{\Lambda}_n - \Lambda_0), \sqrt{n}(\hat{\theta} - \theta_0) \rangle \Rightarrow G \tag{14}
$$

weakly in $l^\infty(H_m)$, *where G is a tight Gaussian process in* $l^\infty(H_m)$ *with mean zero and covariance process*

$$
\mathrm{Cov}\{G(h), G(h^*)\} = \int_0^\tau h_1(t)\sigma_{(1)}^{-1}(h^*)(t)\mathrm{d}\Lambda_0(t) + h_2'\sigma_{(2)}^{-1}(h^*), \tag{15}
$$

where $\sigma = (\sigma_1, \sigma_2)$ is a continuous linear operator from H_∞ to H_∞ with inverse $\sigma^{-1} = (\sigma_{(1)}^{-1}, \sigma_{(2)}^{-1})$. The form of σ is given as follows:

$$\sigma_1(h)(t) = E\{V(t, \Upsilon_0)(h) Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z)\}$$
$$- E\left[\int_t^\tau V(s, \Upsilon_0)(h) Y(s) g(s, \Upsilon_0)\{1 - g(s, \Upsilon_0)\} \exp(2\beta_0' Z) d\Lambda_0(s) \right],$$

and

$$\sigma_2(h) = E\left[\int_0^\tau W(t, \Upsilon_0) V(t, \Upsilon_0)(h) Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) d\Lambda_0(t) \right],$$

where $\Upsilon_0 = (\Lambda_0, \theta_0)$ and

$$V(t, \Lambda, \theta) = h_1(t) - \{1 - g(t, \Lambda, \theta)\} \exp(\beta' Z) \int_0^t h_1(s) d\Lambda(s) + h_2' W(t, \Lambda, \theta).$$

The proof of Theorem 3 is given in the Appendix and follows the following theorem from (van der Vaart and Wellner 1996, Theorem 3.3.1). In this theorem, the parameter space Ψ is a subset of $l^\infty(H_m)$ and the score function is a random map $S_n : \Psi \to l^\infty(H_m)$. Let $\Upsilon = (\Lambda, \theta)$ and $\hat{\Upsilon}_n = (\hat{\Lambda}_n, \hat{\theta}_n)$. Furthermore, denote the asymptotic version of S_n by S, i.e $S(\Upsilon) = E\{S_n(\Upsilon)\}$. Then we have that $S_n(\hat{\Upsilon}_n) = 0$, $S(\Upsilon_0) = 0$ and $\hat{\Upsilon}_n - \Upsilon_0 = o_{\mathcal{P}*}(1)$ as elements in $l^\infty(H_m)$. The notation "*lin*" before a set denotes the set of all finite linear combinations of the elements of the set.

Theorem 4 *Assume the following*:

1. (*Asymptotic distribution of the score function*) $\sqrt{n}\{S_n(\Upsilon_0) - S(\Upsilon_0)\} \Rightarrow G^*$, *where G^* is a tight Gaussian process on $l^\infty(H_m)$.*
2. (*Fréchet differentiability of the asymptotic score*)

$$\sqrt{n}\{S(\hat{\Upsilon}_n) - S(\Upsilon_0)\} = -\sqrt{n}\dot{S}(\Upsilon_0)(\hat{\Upsilon}_n - \Upsilon_0) + o_{\mathcal{P}*}(1 + \sqrt{n}||\hat{\Upsilon}_n - \Upsilon_0||),$$

   where $\dot{S}(\Upsilon_0) : lin\{\Upsilon - \Upsilon_0 : \Upsilon \in \Psi\} \to l^\infty(H_m)$ is a continuous linear operator.
3. (*Invertibility*) $\dot{S}(\Upsilon_0)$ *is continuously invertible on its range.*
4. (*Approximation condition*) $||\sqrt{n}\{(S_n - S)(\hat{\Upsilon}_n) - (S_n - S)(\Upsilon_0)\}|| = o_{\mathcal{P}*}(1 + \sqrt{n}||\hat{\Upsilon}_n - \Upsilon_0||).$

Then, $\sqrt{n}(\hat{\Upsilon}_n - \Upsilon_0) \Rightarrow \dot{S}(\Upsilon_0)^{-1} G^*$.

Remark 4 From (15) of Theorem 3, we have

$$\text{Var}\{G(h)\} = \int_0^\tau h_1(t) \sigma_{(1)}^{-1}(h)(t) d\Lambda_0(t) + h_2' \sigma_{(2)}^{-1}(h), \qquad (16)$$

which provides the asymptotic variances of many quantities of interest. For example, if we choose $h_1(t) = 0$ for all t and $h_2 = e_i$, the ith unit vector, then we obtain the asymptotic variance of the ith element of $\hat{\theta}_n$. Alternatively, if we choose $h_1(s) = I(s \leq t)$ and $h_2 = 0$, then we obtain the asymptotic variance of $\hat{\Lambda}_n(t)$.

Remark 5 A natural estimator of the asymptotic variance of $\langle \sqrt{n}(\hat{\theta} - \theta_0), \sqrt{n}(\hat{\Lambda}_n - \Lambda_0)\rangle(h)$ can be given as $\int_0^\tau h_1 g_1 \mathrm{d}\hat{\Lambda}_n + h_2' g_2$, where $g = (g_1, g_2)$ is the solution to $h_1 = \hat{\sigma}_1(g)$ and $h_2 = \hat{\sigma}_2(g)$, with $\hat{\sigma}_1(g)$ and $\hat{\sigma}_2(g)$ defined as

$$\hat{\sigma}_1(g)(t) = \frac{1}{n} \sum_{i=1}^n V_i(t, \hat{\Upsilon}_n)(g) Y_i(t) g_i(t, \hat{\Upsilon}_n) \exp(\hat{\beta}_n' Z_i)$$

$$- \frac{1}{n} \sum_{i=1}^n \int_t^\tau V_i(s, \hat{\Upsilon}_n)(g) Y_i(s) g_i(s, \hat{\Upsilon}_n)\{1 - g_i(s, \hat{\Upsilon}_n)\} \exp(2\hat{\beta}_n' Z_i) \mathrm{d}\hat{\Lambda}_n(s)$$

$$\hat{\sigma}_2(g) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau W_i(t, \hat{\Upsilon}_n) V_i(t, \hat{\Upsilon}_n)(g) Y_i(t) g_i(t, \hat{\Upsilon}_n) \exp(\hat{\beta}_n' Z_i) \mathrm{d}\hat{\Lambda}_n(t)$$

Here $\hat{\sigma}_1(g)$ and $\hat{\sigma}_2(g)$ are the empirical versions of $\sigma_1(g)$ and $\sigma_2(g)$ with the true parameter values θ_0 and Λ_0 replaced by the maximum likelihood estimators $\hat{\theta}_n$ and $\hat{\Lambda}_n$, respectively.

In the following theorem, we want to show that the asymptotic variance estimator defined above exists and converges to the true value given in (16).

Theorem 5 *Assume that conditions 1–4 hold, then for $h \in H_m$, the solution $g = \hat{\sigma}^{-1}(h)$ exists with probability going to one as n increases. Furthermore, $\int_0^\tau h_1 g_1 \mathrm{d}\hat{\Lambda}_n + h_2' g_2$ converges to $\int_0^\tau h_1(t) \sigma_{(1)}^{-1}(h)(t) \mathrm{d}\Lambda_0(t) + h_2' \sigma_{(2)}^{-1}(h)$ in probability.*

The proof of Theorem 5 is given in the Appendix. Finally, we want to show that the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ achieve the semiparametric efficiency bound, Φ, established in Sect. 2. By the Cramer-Wold device (see Serfling 1980), it suffices to demonstrate that the asymptotic variance of $a' \sqrt{n}(\hat{\theta}_n - \theta_0)$ is equal to $a' \Phi a$, where a is any vector in R^{p+q}. To do this, we need to find an $h = (h_1, h_2)$ such that $\sigma_1(h)(t) = 0$ for all t and $\sigma_2(h) = a$. Consider the solution, $h_2 = \Phi a$ and $h_1(t) = -a_{\mathrm{eff}}'(t) A^{-1}(B\Phi - I)a$, where

$$B = E\left\{ \int_0^\tau W^{\otimes 2}(t, \Upsilon_0) Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) \mathrm{d}\Lambda_0(t) \right\},$$

$$A = E\Bigg(\int_0^\tau W(t, \Upsilon_0)[a_{eff}(t) - \{1 - g(t, \Upsilon_0)\} \exp(\beta_0' Z)$$

$$\times \int_0^t a_{\mathrm{eff}}(s) \mathrm{d}\Lambda_0(s)]' Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) \mathrm{d}\Lambda_0(t) \Bigg).$$

Note that with the h define above

$$
\begin{aligned}
\sigma_2(h) = E\Bigg(\int_0^\tau W(t, \Upsilon_0)\Big[&-a'_{\text{eff}}(t)A^{-1}(B\Phi - I)a + \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z) \\
&\times \int_0^t a'_{\text{eff}}(s)A^{-1}(B\Phi - I)a\mathrm{d}\Lambda_0(s) + W'(t, \Upsilon_0)\Phi a\Big] \\
&\times Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\Lambda_0(t)\Bigg) \\
= -AA^{-1}&(B\Phi - I)a + B\Phi a = a
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma_1(h)(t) = E\Bigg(\Big[&-a_{\text{eff}}(t) + \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t a_{\text{eff}}(s)\mathrm{d}\Lambda_0(s)\Big] \\
&\times Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\Bigg)A^{-1}(B\Phi - I)a \\
&+ E\{W(t, \Upsilon_0)Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\}\Phi a \\
&- E\Bigg(\int_t^\tau\Big[-a_{\text{eff}}(s) + \{1 - g(s, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^s a_{\text{eff}}(u)\mathrm{d}\Lambda_0(u)\Big] \\
&\times Y(s)g(s, \Upsilon_0)\{1 - g(s, \Upsilon_0)\}\exp(2\beta_0'Z)\mathrm{d}\Lambda_0(s)\Bigg)A^{-1}(B\Phi - I)a \\
&- E\Bigg[\int_t^\tau W(s, \Upsilon_0)Y(s)g(s, \Upsilon_0)\{1 - g(s, \Upsilon_0)\}\exp(2\beta_0'Z)\mathrm{d}\Lambda_0(s)\Bigg]\Phi a
\end{aligned}
$$

If $A^{-1}(B\Phi - I) = \Phi$, then

$$
\begin{aligned}
\sigma_1(h)(t) = \Bigg(E\Bigg(\Big[&-a_{\text{eff}}(t) + \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t a_{\text{eff}}(s)d\Lambda_0(s)\Big] \\
&\times Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\Bigg) + E\{W(t, \Upsilon_0)Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\} \\
&- E\Bigg(\int_t^\tau\Big[-a_{\text{eff}}(s) + \{1 - g(s, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^s a_{\text{eff}}(u)\mathrm{d}\Lambda_0(u)\Big] \\
&\times Y(s)g(s, \Upsilon_0)\{1 - g(s, \Upsilon_0)\}\exp(2\beta_0'Z)\mathrm{d}\Lambda_0(s)\Bigg) \\
&- E\Bigg[\int_t^\tau W(s, \Upsilon_0)Y(s)g(s, \Upsilon_0)\{1 - g(s, \Upsilon_0)\}\exp(2\beta_0'Z)\mathrm{d}\Lambda_0(s)\Bigg]\Bigg)\Phi a \\
= \{a_{\text{eff}}(t) &- \int_0^\tau K(t, s)a_{\text{eff}}(s)\mathrm{d}\Lambda_0(s) - f(t)\}E\{Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\}\Phi a \\
= 0
\end{aligned}
$$

the last equality above holds since $a_{\text{eff}}(t)$ is the solution to (9). To check $A^{-1}(B\Phi - I) = \Phi$, it is equivalent to show that $B - A = \Phi^{-1}$. This easily follows the definition of Φ and some simple algebra. Therefore, we have proved that the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ achieve the semiparametric efficiency bound.

4 Concluding remarks

We have developed in this paper the nonparametric maximum likelihood estimators and their associated asymptotic properties for statistical inference related to the proportional hazards cure model. And the proposed estimators for the regression parameters have been shown to be semiparametric efficient. Our methods and derivation rely on modern empirical process theory.

In this paper, a logistic regression is used for modeling the cure fraction. Other parametric models, such as the probit model, can also be easily accommodated. However, the choice of such parametric models is difficult to test in practice. In addition, the covariate vector Z in the Cox model (2) is assumed to be time independent for convenience. This assumption can be relaxed to include time dependent covariates by adding the following condition on $Z(\cdot)$ (also see condition 1 of Bilias et al. (1997)):
Condition 2*. There exists a constant B such that $||Z||_v \leq B$, where $||Z||_v$ is the absolute value of $Z(0)$ plus the total variation of Z on the interval $[0, \tau]$.

Furthermore, the methods proposed here can also be applied to analysis of other semiparametric cure models. For example, the class of transformation mixture cure models considered by Lu and Ying (2004), and the class of semiparametric nonmixture cure models (Tsodikov 1998, 2001; Tsodikov et al. 2003). In particular, Zeng et al. (2006) studied a class of transformation nonmixture cure models using the nonparametric maximum likelihood estimation. The models are specified as

$$P(T > t|X) = G\{\exp(\gamma'X)F(t)\}$$

where $F(t)$ is a completely unspecified distribution function and $G(x)$ is a known monotone decreasing transformation function with $G(0) = 1$. The cure probability for a subject with covariate vector X is then $P(T = \infty|X) = G\{\exp(\gamma'X)\}$. A main difference between the above nonmixture cure model and the proportional hazards mixture cure model considered in the paper is that the parameter γ affects both the cure fraction and the underlying survival distribution of a susceptible subject in the nonmixture cure model while it only affects the cure fraction in the mixture cure model. The goodness-of-fit tests of different types of semiparametric cure models need to be further investigated and warrant future research.

Appendix

Lemma 1 *Assume that conditions 1-4 hold, then* $\sup_n \hat{\Lambda}_n(\tau) < \infty$.

Proof We prove this result by contradiction. That is, suppose that $\sup_n \hat{\Lambda}_n(\tau)$ goes to ∞ as n increases. Then note that

$$
\begin{aligned}
0 &\le l_n(\hat{\Lambda}_n, \hat{\theta}) - l_n(\tilde{\Lambda}_n, \theta_0) \\
&= \frac{1}{n}\sum_{i=1}^n \delta_i \log \frac{\pi(\hat{\gamma}_n' X_i)}{\pi(\gamma_0' X_i)} + \frac{1}{n}\sum_{i=1}^n \delta_i(\hat{\beta}_n - \beta_0)' Z_i \\
&\quad + \frac{1}{n}\sum_{i=1}^n \delta_i \left\{ \log \frac{\Delta\hat{\Lambda}_n(\tilde{T}_i)}{\Delta\tilde{\Lambda}_n(\tilde{T}_i)} - \hat{\Lambda}_n(\tilde{T}_i)\exp(\hat{\beta}_n' Z_i) + \tilde{\Lambda}_n(\tilde{T}_i)\exp(\beta_0' Z_i) \right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^n (1-\delta_i) \log \frac{S(\tilde{T}_i, \hat{\Lambda}_n, \hat{\theta}_n)}{S(\tilde{T}_i, \tilde{\Lambda}_n, \theta_0)} \\
&= \tilde{O}_p(1) + \frac{1}{n}\sum_{i=1}^n \delta_i \left\{ \log \frac{\Delta\hat{\Lambda}_n(\tilde{T}_i)}{\Delta\tilde{\Lambda}_n(\tilde{T}_i)} - \hat{\Lambda}_n(\tilde{T}_i)\exp(\hat{\beta}_n' Z_i) \right\} \\
&= \tilde{O}_p(1) + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \log \left[\frac{1}{n}\sum_{j=1}^n Y_j(t)\exp(\beta_0' Z_j)\{\delta_j + (1-\delta_j)g(\tilde{T}_j, \Upsilon_0)\} \right] dN_i(t) \\
&\quad - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \log \left[\frac{1}{n}\sum_{j=1}^n Y_j(t)\exp(\hat{\beta}_n' Z_j)\{\delta_j + (1-\delta_j)g(\tilde{T}_j, \hat{\Upsilon}_n)\} \right] dN_i(t) \\
&\quad - \frac{1}{n}\sum_{i=1}^n \delta_i \hat{\Lambda}_n(\tilde{T}_i)\exp(\hat{\beta}_n' Z_i) \\
&= \tilde{O}_p(1) - \frac{1}{n}\sum_{i=1}^n \delta_i \hat{\Lambda}_n(\tilde{T}_i)\exp(\hat{\beta}_n' Z_i) \\
&\le \tilde{O}_p(1) - \frac{1}{n}\sum_{i=1}^n \delta_i \hat{\Lambda}_n(\tilde{T}_i) Y_i(\tau)\exp(\hat{\beta}_n' Z_i) \\
&\le \tilde{O}_p(1) - \hat{\Lambda}_n(\tau)\frac{1}{n}\sum_{i=1}^n \delta_i Y_i(\tau)\exp(\hat{\beta}_n' Z_i) \\
&\le \tilde{O}_p(1) - \exp(-c_0)\hat{\Lambda}_n(\tau)\frac{1}{n}\sum_{i=1}^n \delta_i Y_i(\tau)
\end{aligned}
$$

where c_0 is a positive constant and $\tilde{O}_p(1)$ represents quantities, which are bounded away from positive infinity with probability one as n becomes large. Since the limit of $\frac{1}{n}\sum_{i=1}^n \delta_i Y_i(\tau)$ is $E\{\pi(\gamma_0 X)P(C \ge T^* \ge \tau | Z, X)\}$, which is bigger than 0 according to conditions 2 and 3. Thus, as n goes to ∞, the right-

hand side of the above inequality diverges to $-\infty$ with probability one, which is a contradiction. $\qquad\square$

Lemma 2 *Assume that conditions* 1–4 *hold, then* $\sup_{t\in(0,\tau]} |\tilde{\Lambda}_n(t) - \Lambda_0(t)| \to 0$, *a.e., and for each* $\omega \in \Omega$,

(i) Λ^* *is absolutely continuous.*

(ii) $\sup_{t\in(0,\tau]} \left| \dfrac{d\hat{\Lambda}_{n_k}}{d\tilde{\Lambda}_{n_k}}(t) - \gamma(t) \right| \to 0$.

(iii) $\sup_{t\in(0,\tau]} |\hat{\Lambda}_{n_k}(t) - \int_0^t \gamma(s)d\Lambda_0(s)| \to 0$.

Proof Note that $M_0(t) = N(t) - \int_0^t Y(s)\exp(\beta_0'Z)g(s,\Upsilon_0)d\Lambda_0(s)$ is mean zero and $E\{Y(t)\exp(\beta_0'Z)g(t,\Upsilon_0)\} > 0$ for $t \in [0,\tau]$, we have

$$\Lambda_0(t) = \int_0^t E\{Y(s)\exp(\beta_0'Z)g(s,\Upsilon_0)\}^{-1}dE\{N(s)\}$$

where $E\{N(t)\} = E\{\int_0^t Y(s)\exp(\beta_0'Z)g(s,\Upsilon_0)d\Lambda_0(s)\}$. Then

$$\tilde{\Lambda}_n(t) - \Lambda_0(t) = \int_0^t \left[\frac{1}{n}\sum_{j=1}^n Y_j(s)\exp(\beta_0'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\Lambda_0,\theta_0)\} \right]^{-1} d\bar{N}(s)$$
$$- \int_0^t E\{Y(s)\exp(\beta_0'Z)g(s,\Upsilon_0)\}^{-1}dE\{N(s)\}$$

By the Glivenko-Cantelli theorem, $\frac{1}{n}\sum_{j=1}^n Y_j(s)\exp(\beta_0'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\Lambda_0,\theta_0)\}$ uniformly converges to $E[Y(s)\exp(\beta_0'Z)\{\delta + (1-\delta)g(\tilde{T},\Lambda_0,\theta_0)\}]$ on $[0,\tau]$. In addition,

$$E\{Y(t)\exp(\beta_0'Z)\delta\} = E[I(C \ge t)\exp(\beta_0'Z)E\{I(t \le T \le C)|Z,X,C\}]$$
$$= E\Big(I(C \ge t)\exp(\beta_0'Z)\pi(\gamma_0'X)[\exp\{-\Lambda_0(t)\exp(\beta_0'Z)\} - \exp\{-\Lambda_0(C)\exp(\beta_0'Z)\}]\Big)$$

and

$$E\{Y(t)\exp(\beta_0'Z)(1-\delta)g(\tilde{T},\Lambda_0,\theta_0)\}$$
$$= E\Big[I(C \ge t)\exp(\beta_0'Z)g(C,\Lambda_0,\theta_0)E\{I(T > C)|Z,X,C\}\Big]$$
$$= E\Big[I(C \ge t)\exp(\beta_0'Z)\pi(\gamma_0'X)\exp\{-\Lambda_0(C)\exp(\beta_0'Z)\}\Big]$$

Therefore,

$$
\begin{aligned}
E[Y(t)\exp(\beta_0'Z)\{\delta + (1-\delta)g(\tilde{T}, \Lambda_0, \theta_0)\}] \\
&= E\Big[I(C \geq t)\exp(\beta_0'Z)\pi(\gamma_0'X)\exp\{-\Lambda_0(t)\exp(\beta_0'Z)\}\Big] \\
&= E\Big[I(C \geq t)\exp(\beta_0'Z)g(t, \Lambda_0, \theta_0)E\{I(T \geq t|Z, X)\}\Big] \\
&= E\{Y(t)\exp(\beta_0'Z)g(t, \Lambda_0, \theta_0)\}
\end{aligned}
$$

Since $\inf_{t\in[0,\tau]} E\{Y(t)\exp(\beta_0'Z)g(t, \Lambda_0, \theta_0)\} > 0$, by Lemma A.2 of Tsiatis (1981), we have $\sup_{t\in(0,\tau]} |\tilde{\Lambda}_n(t) - \Lambda_0(t)| \to 0$, a.e.

For (i), let f be any non-negative, bounded, continuous function. Then

$$
\int_0^\tau f(t)\mathrm{d}\Lambda^*(t)
$$

$$
= \int_0^\tau f(t)\mathrm{d}\{\Lambda^*(t) - \hat{\Lambda}_{n_k}(t)\}
$$

$$
+ \int_0^\tau f(t)\left[\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\hat{\beta}_{n_k}'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\}\right]^{-1}\mathrm{d}\bar{N}_{n_k}(t)
$$

$$
\leq \int_0^\tau f(t)\mathrm{d}\{\Lambda^*(t) - \hat{\Lambda}_{n_k}(t)\} + M_4\int_0^\tau f(t)\left\{\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\right\}^{-1}\mathrm{d}\bar{N}_{n_k}(t)
$$

where $\bar{N}_{n_k}(t) = \sum_{i=1}^{n_k} N_i(t)$ and M_4 is a positive constant. The existence of such constant M_4 is because

$$
\begin{aligned}
\exp(\hat{\beta}_{n_k}'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\} \\
&\geq \exp(-c_0)\pi(\hat{\gamma}_{n_k}'X_j)\exp\{-\hat{\Lambda}_{n_k}(\tau)\exp(c_0)\} \\
&\geq c_1\exp(-c_0)\exp\{-c_2\exp(c_0)\}
\end{aligned}
$$

where c_1 and c_2 are two positive constants. By the Helly-Bray Lemma (p. 180 of Loeve 1963), $\int_0^\tau f(t)\mathrm{d}\{\Lambda^*(t) - \hat{\Lambda}_{n_k}(t)\} \to 0$ as $k \to \infty$. Using Lemmas of A.1 and A.2 of Tsiatis (1981), we know that

$$
\int_0^\tau f(t)\left\{\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\right\}^{-1}\mathrm{d}\bar{N}_{n_k}(t)
$$

$$
\to \int_0^\tau f(t)[E\{Y(t)\}]^{-1}E\{Y(t)\exp(\beta_0'Z)g(t, \Lambda_0, \theta_0)\}\mathrm{d}\Lambda_0(t)
$$

as $k \to \infty$. Since $E\{Y(t)\}$ is bounded away from zero for all $t \in [0, \tau]$, we know that

$$\int_0^\tau f(t)\mathrm{d}\Lambda^*(t) \le M_4 \int_0^\tau f(t)[E\{Y(t)\}]^{-1}E\{Y(t)\exp(\beta_0'Z)g(t,\Lambda_0,\theta_0)\}\mathrm{d}\Lambda_0(t)$$

By choosing f appropriately, this inequality implies that Λ^* must be continuous at the continuity points of Λ_0. Since Λ_0 is absolutely continuous, then so is Λ^*.

For (ii), since

$$\frac{\mathrm{d}\hat{\Lambda}_{n_k}(t)}{\mathrm{d}\tilde{\Lambda}_{n_k}(t)} = \frac{\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\beta_0'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\Lambda_0,\theta_0)\}}{\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\hat{\beta}_{n_k}'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})\}}$$

and we have already shown that

$$\sup_{t\in[0,\tau]} \left| \frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\beta_0'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\Lambda_0,\theta_0)\} \right.$$

$$\left. -E\{Y(t)\exp(\beta_0'Z)g(t,\Lambda_0,\theta_0)\} \right|$$

converges to 0 as k goes to ∞. Then we need to show

$$\sup_{t\in[0,\tau]} \left| \frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\hat{\beta}_{n_k}'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})\} \right.$$

$$\left. -E[Y(t)\exp(\beta^{*\prime}Z)\{\delta + (1-\delta)g(\tilde{T},\Lambda^*,\theta^*)\}] \right|$$

converges to 0 as k goes to ∞. In fact, the term in above supremum norm is bounded above by

$$\left| \frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\hat{\beta}_{n_k}'Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})\} \right.$$

$$\left. -\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\beta^{*\prime}Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\hat{\Lambda}_{n_k},\theta^*)\} \right|$$

$$+\left| \frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\beta^{*\prime}Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\hat{\Lambda}_{n_k},\theta^*)\} \right.$$

$$\left. -\frac{1}{n_k}\sum_{j=1}^{n_k} Y_j(t)\exp(\beta^{*\prime}Z_j)\{\delta_j + (1-\delta_j)g_j(\tilde{T}_j,\Lambda^*,\theta^*)\} \right|$$

$$+ \left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\beta^{*\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda^*, \theta^*)\} \right.$$

$$\left. - E[Y(t) \exp(\beta^{*\prime} Z) \{\delta + (1 - \delta) g(\tilde{T}, \Lambda^*, \theta^*)\}] \right|$$

$$\leq M_5 ||\hat{\theta}_{n_k} - \theta^*||_\infty + M_6 \sup_{t \in [0,\tau]} ||\hat{\Lambda}_{n_k}(t) - \Lambda^*(t)||$$

$$+ \left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\beta^{*\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda^*, \theta^*)\} \right.$$

$$\left. - E[Y(t) \exp(\beta^{*\prime} Z) \{\delta + (1 - \delta) g(\tilde{T}, \Lambda^*, \theta^*)\}] \right|$$

where M_5, M_6 are two positive constants. The first two terms converge to 0 by the uniform consistency of $\hat{\Lambda}_{n_k}$ and $\hat{\theta}_{n_k}$. The third term needs to be handled more carefully. As noted by Scharfsten et al. (1998), the space of absolutely, bounded, increasing functions $\{\Lambda^*(t)\}$ is separable under the supremum norm. Thus, the space has a countably dense subset. Let $\{\Lambda_l^*\}, l \geq 1$, denote this set.

Then we have

$$\sup_{t \in [0,\tau]} \left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\xi_\beta^{\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda_l^*, \xi_\theta)\} \right.$$

$$\left. - E[Y(t) \exp(\xi_\beta^{\prime} Z) \{\delta + (1 - \delta) g(\tilde{T}, \Lambda_l^*, \xi_\theta)\}] \right|$$

converges to 0 as k goes to ∞, for each rational $\xi_\theta = (\xi_\beta^{\prime}, \xi_\gamma^{\prime})^{\prime}$ and $l \geq 1$. The third term can be bounded above by

$$\left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\beta^{*\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda^*, \theta^*)\} \right.$$

$$\left. - \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\xi_\beta^{\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda^*, \xi_\theta)\} \right|$$

$$+ \left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\xi_\beta^{\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda^*, \xi_\theta)\} \right.$$

$$\left. - \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\xi_\beta^{\prime} Z_j) \{\delta_j + (1 - \delta_j) g_j(\tilde{T}_j, \Lambda_l^*, \xi_\theta)\} \right|$$

$$+ \left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\xi_\beta' Z_j)\{\delta_j + (1 - \delta_j)g_j(\tilde{T}_j, \Lambda_l^*, \xi_\theta)\} \right.$$

$$\left. - E[Y(t) \exp(\xi_\beta' Z)\{\delta + (1 - \delta)g(\tilde{T}, \Lambda_l^*, \xi_\theta)\}] \right|$$

$$\leq M_7 ||\theta^* - \xi_\theta||_\infty + M_8 \sup_{t\in[0,\tau]} ||\Lambda^*(t) - \Lambda_l^*(t)||$$

$$+ \left| \frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\xi_\beta' Z_j)\{\delta_j + (1 - \delta_j)g_j(\tilde{T}_j, \Lambda_l^*, \xi_\theta)\} \right.$$

$$\left. - E[Y(t) \exp(\xi_\beta' Z)\{\delta + (1 - \delta)g(\tilde{T}, \Lambda_l^*, \xi_\theta)\}] \right|$$

where M_7, M_8 are two positive constants. By appropriate choice of ξ and l, the right-hand side of the above inequality converges to 0 in supremum norm. So, (ii) holds.

For(iii), since

$$\left| \hat{\Lambda}_{n_k}(t) - \int_0^t \gamma(s)\mathrm{d}\Lambda_0(s) \right|$$

$$= \left| \int_0^t \left[\frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(s) \exp(\hat{\beta}_{n_k}' Z_j)\{\delta_j + (1 - \delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\} \right]^{-1} \mathrm{d}\bar{N}_{n_k}(s) \right.$$

$$\left. - \int_0^t \gamma(s)\mathrm{d}\Lambda_0(s) \right|$$

$$= \left| \int_0^t \left[\frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(s) \exp(\hat{\beta}_{n_k}' Z_j)\{\delta_j + (1 - \delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\} \right]^{-1} \mathrm{d}\bar{N}_{n_k}(s) \right.$$

$$\left. - \int_0^t \left(E[Y(s) \exp(\beta^{*'}Z)\{\delta + (1 - \delta)g^*(\tilde{T})\}] \right)^{-1} \mathrm{d}EN(s) \right|$$

$$\leq \left| \int_0^t \left\{ \left[\frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(s) \exp(\hat{\beta}_{n_k}' Z_j)\{\delta_j + (1 - \delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\} \right]^{-1} \right. \right.$$

$$\left. \left. - \int_0^t \left(E[Y(s) \exp(\beta^{*'}Z)\{\delta + (1 - \delta)g^*(\tilde{T})\}] \right)^{-1} \right\} \mathrm{d}\bar{N}_{n_k}(s) \right|$$

$$+ \left| \int_0^t \left(E[Y(s) \exp(\beta^{*'}Z)\{\delta + (1 - \delta)g^*(\tilde{T})\}] \right)^{-1} \mathrm{d}\{\bar{N}_{n_k}(s) - EN(s)\} \right|$$

$$\leq \sup_{t\in[0,\tau]} \left| \left[\frac{1}{n_k} \sum_{j=1}^{n_k} Y_j(t) \exp(\hat{\beta}_{n_k}' Z_j)\{\delta_j + (1 - \delta_j)g_j(\tilde{T}_j, \hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\} \right]^{-1} \right.$$

$$- \left(E[Y(t)\exp(\beta^{*\prime}Z)\{\delta + (1-\delta)g^*(\tilde{T})\}]\right)^{-1} \Bigg|$$

$$+ \left| \int_0^t \left(E[Y(s)\exp(\beta^{*\prime}Z)\{\delta + (1-\delta)g^*(\tilde{T})\}]\right)^{-1} \mathrm{d}\{\bar{N}_{n_k}(s) - EN(s)\} \right|$$

the first term was shown to converge to 0 in (ii) and the second term converges to zero by the Helly-Bray Lemma. In addition, pointwise convergence can be strengthened to uniform convergence by applying the same monotonicity argument used in the proof of the Glivenko-Cantelli Theorem (p. 96 of Shorack and Wellner 1986). □

Proof of Theorem 2 We first show that $(\hat{\Lambda}_n, \hat{\theta}_n)$ converges to (Λ_0, θ_0) a.s. To do this, we need to show the following three things: (i) $\sup_n \hat{\Lambda}_n(\tau) < \infty$; (ii) there exists a convergent subsequence of $(\hat{\Lambda}_n, \hat{\theta}_n)$, say $(\hat{\Lambda}_{n_k}, \hat{\theta}_{n_k}) \to (\Lambda^*, \theta^*)$ a.s.; (iii) $(\Lambda^*, \theta^*) = (\Lambda_0, \theta_0)$. The proof of (i) is given in Lemma 1 of the Appendix. For (ii), since every bounded sequence in R^{p+q} has a convergent subsequence, there exists a θ^* such that $\hat{\theta}_{m_k} \to \theta^*$. By Helly's theorem (Ash 1972), there exists a function Λ^* and a subsequence $\{\hat{\Lambda}_{n_k}\}$ of $\{\hat{\Lambda}_{m_k}\}$ such that $\hat{\Lambda}_{n_k} \to \Lambda^*$ for all $t \in [0, \tau]$ at which Λ^* is continuous. Therefore, $(\hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})$ must converge to (Λ^*, θ^*).

For (iii), we show in Lemma 2 of the Appendix that Λ^* is continuous at the continuity points of Λ_0. And we know that

$$0 \le l_{n_k}(\hat{\Lambda}_{n_k}, \hat{\theta}_{n_k}) - l_{n_k}(\tilde{\Lambda}_{n_k}, \theta_0)$$

$$= \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^\tau \log\{\chi_{n_k,i}(t)\}\{\mathrm{d}N_i(t) - Y_i(t)g_i(t, \tilde{\Lambda}_{n_k}, \theta_0)\exp(\beta_0'Z_i)\mathrm{d}\tilde{\Lambda}_{n_k}(t)\}$$

$$+ \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^\tau [\log\{\chi_{n_k,i}(t)\} - \{\chi_{n_k,i}(t) - 1\}]Y_i(t)g_i(t, \tilde{\Lambda}_{n_k}, \theta_0)\exp(\beta_0'Z_i)\mathrm{d}\tilde{\Lambda}_{n_k}(t),$$

where

$$\chi_{n_k,i}(t) = \frac{\Delta\hat{\Lambda}_{n_k}(t)g_i(\hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})\exp(\hat{\beta}_{n_k}'Z_i)}{\Delta\tilde{\Lambda}_{n_k}(t)g_i(\tilde{\Lambda}_{n_k}, \theta_0)\exp(\beta_0'Z_i)}.$$

The second term on the right-hand side of the above inequality is less or equal to zero since for $x > 0$, $\log(x) - (x-1) \le 0$. Using the results and techniques of Lemma 2 of the Appendix, the first term can be shown to converge to zero and the second term converges to

$$E\left(\int_0^\tau \left[\log\left\{ \frac{g^*(t)\exp(\beta^{*\prime}Z)}{g_0(t)\exp(\beta_0'Z)}\gamma(t) \right\} - \left\{ \frac{g^*(t)\exp(\beta^{*\prime}Z)}{g_0(t)\exp(\beta_0'Z)}\gamma(t) - 1 \right\} \right] \right.$$

$$\left. \times Y(t)g_0(t)\exp(\beta_0'Z)d\Lambda_0(t) \right), \tag{17}$$

where $g^*(t) = g(t, \Lambda^*, \theta^*)$ and

$$\gamma(t) = \frac{E[Y(t) \exp(\beta_0' Z)\{\delta + (1-\delta)g_0(\tilde{T})\}]}{E[Y(t) \exp(\beta^{*'} Z)\{\delta + (1-\delta)g^*(\tilde{T})\}]}. \tag{18}$$

Note that (17) is the negative Kullback-Leibler information, $E[l(\Lambda^*, \theta^*)] - E[l(\Lambda_0, \theta_0)]$. Due to the above inequality, the Kullback-Leibler information must equal zero. Thus, with probability one, we have

$$\int_0^\tau \log\{\lambda^*(t) \exp(\beta^{*'} Z)g^*(t)\} dN(t) - \int_0^\tau Y(t) \exp(\beta^{*'} Z)g^*(t) d\Lambda^*(t)$$

$$= \int_0^\tau \log\{\lambda_0(t) \exp(\beta_0' Z)g_0(t)\} dN(t) - \int_0^\tau Y(t) \exp(\beta_0' Z)g_0(t) d\Lambda_0(t),$$

where $\lambda^*(t) = d\Lambda^*(t)/dt$. This equality holds for the following two cases: (i) $Y(\tau) = 1, N(\tau) = 0$, and (ii) $Y(\tau) = 1, N(t-) = 0$, and $N(\tau) = 1$ for $\forall\, t \in (0, \tau]$. The difference between the equalities from these two cases entails that

$$\lambda^*(t) \exp(\beta^{*'} Z)g^*(t) = \lambda_0(t) \exp(\beta_0' Z)g_0(t), \quad \forall\, t \in (0, \tau].$$

After integrating form 0 to t on both sides of the above equality, we have

$$\log\{S(t, \Lambda_0, \theta_0)\} = \log\{S(t, \Lambda^*, \theta^*)\},$$

which implies

$$\frac{1 - \exp\{-\Lambda_0(t) \exp(\beta_0' z)\}}{1 - \exp\{-\Lambda^*(t) \exp(\beta^{*'} z)\}} = \frac{\pi(\gamma^{*'} X)}{\pi(\gamma_0' X)}, \quad \forall\, t \in (0, \tau].$$

Since the right-hand side of the above equality is independent of t, we have that $\Lambda^* = \Lambda_0$ and $\beta^* = \beta_0$. This further implies that $\gamma^* = \gamma_0$. Therefore, $(\hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})$ must converge to (Λ_0, θ_0) a.s. By Helly's theorem, we know that $(\hat{\Lambda}_n, \hat{\theta}_n)$ must converge to (Λ_0, θ_0) a.s. Furthermore, the point-wise convergence can be strengthened to uniform convergence by applying the same monotonicity argument used in the proof of the Glivenko-Cantelli Theorem (Shorack and Wellner 1986). □

Proof of Theorem 3 To prove Theorem 3, we validate each of the four conditions of Theorem 4. First, note that $S = S_1 + S_2$, where

$$S_1(\Upsilon) = E\left(\int_0^\tau \left[h_1(t) - \{1 - g(t, \Upsilon)\} \exp(\beta' Z) \int_0^t h_1(s) d\Lambda(s)\right]\right.$$

$$\times \{dN(t) - Y(t)g(t, \Upsilon) \exp(\beta' Z)d\Lambda(t)\}\Big),$$

$$S_2(\Upsilon) = E\left[\int_0^\tau h_2' W(t, \Upsilon)\{dN(t) - Y(t)g(t, \Upsilon) \exp(\beta' Z)d\Lambda(t)\}\right].$$

For $\Upsilon - \Upsilon_0 \in l^\infty(H_m)$, we need the following bounds on $||\Upsilon - \Upsilon_0||$:

$$m||\theta - \theta_0||_1 \vee m||\Lambda - \Lambda_0||_\infty \leq ||\Upsilon - \Upsilon_0|| \leq m||\theta - \theta_0||_1 \vee 2m||\Lambda - \Lambda_0||_\infty$$

Step 1. We want to establish condition 1 for all finite m. To do this, we show that the class of score function $\mathcal{S}^* \equiv \{\mathcal{S}^*(\Upsilon_0)h : h \in H_m\}$ is Donsker, where $\mathcal{S}^*(\Upsilon)h = S_\Lambda(\Upsilon)(h_1) + h_2' S_\theta(\Upsilon)$, with

$$S_\Lambda(\Upsilon)(h_1) = \int_0^\tau \left[h_1(t) - \{1 - g(t, \Upsilon)\} \exp(\beta'Z) \int_0^t h_1(s) \mathrm{d}\Lambda(s) \right]$$
$$\times \{\mathrm{d}N(t) - Y(t)g(t, \Upsilon)\exp(\beta'Z)\mathrm{d}\Lambda(t)\}$$
$$= \delta h_1(\tilde{T}) - \{\delta + (1-\delta)g(\tilde{T}, \Upsilon)\}\exp(\beta'Z) \int_0^{\tilde{T}} h_1(t)\mathrm{d}\Lambda(t),$$
$$S_\theta(\Upsilon) = \int_0^\tau W(t, \Upsilon) \left\{ \mathrm{d}N(t) - Y(t)g(t, \Upsilon)\exp(\beta'Z)\mathrm{d}\Lambda(t) \right\}$$

Conditions 1 and 2 imply that $S_\theta(\Upsilon_0)$ is a uniformly bounded function, which implies that $\{h_2' S_\theta(\Upsilon_0) : h \in R^{p+q}, |h_2|_1 \leq m\}$ is Donsker (see Example 2.10.10 of van der Vaart and Wellner 1996). Since the sum of bounded Donsker classes is Donsker, the class $\{S_\Lambda(\Upsilon_0)(h_1) : h_1 \in BV[0, \tau], ||h_1||_v \leq m\}$ is Donsker if the following two classes

$$\mathcal{F}_1 = \{\delta h_1(\tilde{T}) : h_1 \in BV[0, \tau], \ ||h_1||_v \leq m\},$$
$$\mathcal{F}_2 = \left\{ \{\delta + (1-\delta)g(\tilde{T}, \Upsilon_0)\}\exp(\beta_0'Z) \int_0^{\tilde{T}} h_1(t)\mathrm{d}\Lambda_0(t) : h_1 \in BV[0, \tau], \ ||h_1||_v \leq m \right\}$$

are Donsker and uniformly bounded. The class $\mathcal{F}_1$ is uniformly bounded and Donsker since h_1 varies over bounded variation functions (see Example 2.5.4 of van der Vaart and Wellner, 1996). In addition, $\mathcal{F}_2$ equals a uniformly bounded function times the class $\{\int_0^{\tilde{T}} h_1(t)\mathrm{d}\Lambda_0(t) : h_1 \in BV[0, \tau], ||h_1||_v \leq m\}$, and this class is Donsker because Λ_0 is a monotone function (see Example 2.10.27 of van der Vaart and Wellner 1996). Also $\mathcal{F}_2$ is uniformly bounded because h_1 varies over bounded variation functions. Thus, we conclude that $\mathcal{S}^*$ is Donsker, so the first condition holds.

Step 2. To check condition 2, it suffices to show that $||S(\Upsilon) - S(\Upsilon_0) + \dot{S}(\Upsilon_0)(\Upsilon - \Upsilon_0)||$ is $o(||\Upsilon - \Upsilon_0||)$ as $||\Upsilon - \Upsilon_0||$ goes to 0. To do this, write $S(\Upsilon)$ linearly in $d(\Lambda - \Lambda_0)$ and $\theta - \theta_0$. To be specific,

$$S_1(\Upsilon)(h) = (\theta - \theta_0)' E\Bigg(\int_0^\tau W(t, \Upsilon_0)[h_1(t) - \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)$$

$$\times \int_0^t h_1(s)\mathrm{d}\Lambda_0(s)]Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\Lambda_0(t)\Bigg)$$

$$+ E\Bigg(\int_0^\tau [h_1(t) - \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)\mathrm{d}\Lambda_0(s)]$$

$$\times Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\{\Lambda(t) - \Lambda_0(t)\}\Bigg)$$

$$- E\Bigg(\int_0^\tau [h_1(t) - \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)\mathrm{d}\Lambda_0(s)]$$

$$\times \{\Lambda(t) - \Lambda_0(t)\}Y(t)g(t, \Upsilon_0)\{1 - g(t, \Upsilon_0)\}\exp(2\beta_0'Z)\mathrm{d}\Lambda_0(t)\Bigg)$$

$$+ \mathrm{error}_1(\Upsilon)(h),$$

$$S_2(\Upsilon)(h) = (\theta - \theta_0)' E\Bigg\{\int_0^\tau W(t, \Upsilon_0)h_2'W(t, \Upsilon_0)\mathrm{d}\Lambda_0(t)\Bigg\}$$

$$+ E\Bigg[\int_0^\tau h_2'W(t, \Upsilon_0)Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\{\Lambda(t) - \Lambda_0(t)\}\Bigg]$$

$$- E\Bigg[\int_0^\tau h_2'W(t, \Upsilon_0)\{\Lambda(t) - \Lambda_0(t)\}$$

$$\times Y(t)g(t, \Upsilon_0)\{1 - g(t, \Upsilon_0)\}\exp(2\beta_0'Z)\mathrm{d}\Lambda_0(t)\Bigg] + \mathrm{error}_2(\Upsilon)(h).$$

And the error terms can be very easily shown to satisfy

$$\sup_{h \in H_m} \frac{||\mathrm{error}_i(\Upsilon)(h)||}{||\theta - \theta_0||_1 \vee ||\Lambda - \Lambda_0||_\infty} \to 0$$

as $||\theta - \theta_0||_1 \vee ||\Lambda - \Lambda_0||_\infty \to 0$, where $i = 1, 2$. This follows from the boundedness of Z, X, N, Y, θ and Λ. Furthermore,

$$\frac{||S(\Upsilon) - S(\Upsilon_0) + \dot{S}(\Upsilon_0)(\Upsilon - \Upsilon_0)||}{||\Upsilon - \Upsilon_0||} \leq \frac{||\mathrm{error}_1(\Upsilon)(h)|| + ||\mathrm{error}_2(\Upsilon)(h)||}{m||\theta - \theta_0||_1 \vee m||\Lambda - \Lambda_0||_\infty}$$

Since $m||\theta - \theta_0||_1 \vee m||\Lambda - \Lambda_0||_\infty \to 0$ as $||\Upsilon - \Upsilon_0|| \to 0$, we conclude that $||S(\Upsilon) - S(\Upsilon_0) + \dot{S}(\Upsilon_0)(\Upsilon - \Upsilon_0)||$ is $o(||\Upsilon - \Upsilon_0||)$ as $||\Upsilon - \Upsilon_0||$ goes to 0. Note that $S(\Upsilon_0) = 0$, combining S_1 and S_2 we have

$$\dot{S}(\Upsilon_0)(\hat{\Upsilon}_n - \Upsilon_0)(h) = \int_0^\tau \sigma_1(h)(t)\mathrm{d}(\hat{\Lambda}_n - \Lambda_0)(t) + (\hat{\theta}_n - \theta_0)'\sigma_2(h). \quad (19)$$

Step 3. To check condition 3, we need to show that $\dot{S}(\Upsilon_0)$ is continuously invertible. Following Scharfsten et al. (1998), we need to show that $\sigma(h)$ is invertible

everywhere due to the discreteness of $\hat{\Lambda}_n$. To do this, we first show that $\sigma(h)$ is one-to-one map on $L_2(d\Lambda_0) \bigotimes R^{p+q}$, i.e.

$$\int_0^\tau \sigma_1(h)(t)h_1(t)d\Lambda_0(t) + h_2'\sigma_2(h) = 0 \tag{20}$$

implies that $h_2 = 0$ and $h_1(t) = 0$ almost everywhere $(d\Lambda_0)$. Plug $\sigma_1(h)(t)$ and $\sigma_2(h)$ into the above equation, we have

$$
\begin{aligned}
0 = {} & E\bigg(\int_0^\tau h_1(t)\bigg[V(t,\Upsilon_0)Y(t)g(t,\Upsilon_0)\exp(\beta_0'Z) \\
& - \int_t^\tau V(s,\Upsilon_0)Y(t)g(s,\Upsilon_0)\{1 - g(s,\Upsilon_0)\}\exp(2\beta_0'Z)d\Lambda_0(s)\bigg]d\Lambda_0(t)\bigg) \\
& + h_2'E\int_0^\tau \Big\{W(t,\Upsilon_0)V(t,\Upsilon_0)Y(t)g(t,\Upsilon_0)\exp(\beta_0'Z)d\Lambda_0(t)\Big\} \\
= {} & E\bigg[\int_0^\tau \{h_1(t) + h_2'W(t,\Upsilon_0)\}V(t,\Upsilon_0)Y(t)g(t,\Upsilon_0)\exp(\beta_0'Z)d\Lambda_0(t)\bigg] \\
& - E\bigg[\int_0^\tau \int_0^t h_1(s)d\Lambda_0(s)V(t,\Upsilon_0)Y(t)g(t,\Upsilon_0)\{1-g(t,\Upsilon_0)\}\exp(2\beta_0'Z)d\Lambda_0(t)\bigg] \\
= {} & E\bigg[\int_0^\tau V^{\otimes 2}(t,\Upsilon)Y(t)g(t,\Upsilon_0)\exp(\beta_0'Z)d\Lambda_0(t)\bigg].
\end{aligned}
$$

Since $Y(t)g(t,\Upsilon_0)\exp(\beta_0'Z) > 0$ a.e. on $[0,\tau]$, it implies $V(t,\Upsilon_0) = 0$ a.e. $(d\Lambda_0)$. Therefore, for almost all $\omega \in \Omega$,

$$
\begin{aligned}
& h_1(t) + \{1 - g(t,\Upsilon_0)\}[h_{22}'X(\omega) \\
& \quad - \exp\{\beta_0'Z(\omega)\}\int_0^t \{h_{21}'Z(\omega) + h_1(s)\}d\Lambda_0(s)] = -h_{21}'Z(\omega)
\end{aligned}
$$

a.e. $(d\Lambda_0)$. From this, h_{21} must be zero. It further implies that

$$\frac{h_1(t)}{1 - g(t,\Upsilon_0)} - \exp\{\beta_0'Z(\omega)\}\int_0^t h_1(s)d\Lambda_0(s) = -h_{22}'X(\omega)$$

a.e. $(d\Lambda_0)$. Similarly, h_{22} must be zero. With $h_2 = (h_{21}', h_{22}')' = 0$, we have $h_1(t) - \{1 - g(t,\Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)d\Lambda_0(s) = 0$ a.e. $(d\Lambda_0)$. This is a Volterra integral equation of the first kind. It is easy to show that the solution, $h_1(\cdot)$, to this equation must be zero a.e. $(d\Lambda_0)$. Now based on the fact (20), we want to show that σ is one-to-one everywhere, i.e. set $\sigma = 0$ and show that $h_2 = 0$ and $h_1(t) = 0$ for all t on $[0,\tau]$. If $\sigma_1(h)(t) = 0$ for all t and $\sigma_2(h) = 0$, from (20) we have $h_2 = 0$ and $h_1(t) = 0$ a.e. $(d\Lambda_0)$. Then

$$0 = \sigma_1(h)(t) = E\{V(t, \Upsilon_0)Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\}$$
$$-E\left[\int_t^\tau V(s, \Upsilon_0)g(s, \Upsilon_0)\{1 - g(s, \Upsilon_0)\}\exp(2\beta_0'Z)d\Lambda_0(s)\right].$$

Based on the fact $h_1(t)d\Lambda_0(t) = 0$ for all t on $[0, \tau]$, we have

$$\sigma_1(h)(t) = h_1(t)E\{Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\} = 0.$$

Since $E\{Y(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\} > 0$ for all t on $[0, \tau]$, we conclude that $h_1(t) = 0$ for all t. Therefore, σ is one-to-one everywhere.

Then, we want to show that σ, as a continuous linear operator from H_∞ to H_∞, has a continuous inverse. Note that H_∞ is a Banach space, if σ is invertible, then the inverse will be continuous (see p. 149, Luenberger, 1969). To show that σ is invertible, since σ is one-to-one, we only need to show that it can be written as the difference of a bounded, linear operator with a bounded inverse and a compact, linear operator (see Corollary 3.8 and Theorem 3.4 of Kress, 1989). To do this, we define the following linear operator

$$\Sigma(h)(t) = \left(h_1(t)E\{Y(t)\exp(\beta_0'Z)g(t, \Upsilon_0)\},\right.$$
$$\left.h_2E\left\{\int_0^\tau W^{\otimes 2}(t, \Upsilon_0)Y(t)\exp(\beta_0'Z)g(t, \Upsilon_0)d\Lambda_0(t)\right\}\right)$$

This is a bounded linear operator due to the boundedness of $Z, X, \theta_0, Y(t)$ and $\Lambda_0(t)$. In addition, under conditions 1–4, $E\{Y(t)\exp(\beta_0'Z)g(t, \Upsilon_0)\} > \epsilon$ on $[0, \tau]$ for some $\epsilon > 0$ and $W^{\otimes 2}(t, \Upsilon_0)$ is positive definite a.e. on $[0, \tau]$, which implies

$$E\left\{\int_0^\tau W^{\otimes 2}(t, \Upsilon_0)Y(t)\exp(\beta_0'Z)g(t, \Upsilon_0)d\Lambda_0(t)\right\}$$

is a positive definite matrix. Hence, the inverse of $\Sigma(h)(t)$ exists, which is given by

$$\Sigma^{-1}(h)(t) = \left(h_1(t)E\{Y(t)\exp(\beta_0'Z)g(t, \Upsilon_0)\}^{-1},\right.$$
$$\left.h_2E\left\{\int_0^\tau W^{\otimes 2}(t, \Upsilon_0)Y(t)\exp(\beta_0'Z)g(t, \Upsilon_0)d\Lambda_0(t)\right\}^{-1}\right)$$

and it is also a bounded linear operator.

Now we want to show that $\Sigma(h) - \sigma(h)$ is compact. Let $\{h_n\}$ be a sequence in H_m. Then it only needs to show that there exists a convergent subsequence of $\Sigma(h_n) - \sigma(h_n)$. Since h_{1n} is of bounded variation, we can write h_{1n} as the difference of increasing functions (see Lemma 2.3.3 of Ash 1972). In addition, both of these increasing functions are bounded in absolute value by $2m$. This

implies that there exists a pointwise convergent subsequence according to Hel-
ly's theorem. Let $\{h_{n_k}\}$ be the convergent subsequence with limit h^*. We must
prove that $\Sigma(h_{n_k}) - \sigma(h_{n_k})$ converges to $\Sigma(h^*) - \sigma(h^*)$ in $||h||_H$ norm. To do
this, note that $\Sigma(h) - \sigma(h)$ can be written as

$$
\left(E\left(\left[\left\{ 1 - g(t, \Upsilon_0) \right\} \exp(\beta_0' Z) \int_0^t h_1(s) \mathrm{d}\Lambda_0(s) - h_2' W(t, \Upsilon_0) \right] Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) \right) \right.
$$
$$
+ E\left[\int_t^\tau V(s, \Upsilon_0) Y(s) g(s, \Upsilon_0) \{ 1 - g(s, \Upsilon_0) \} \exp(2\beta_0' Z) \mathrm{d}\Lambda_0(s) \right],
$$
$$
- E\left(\int_0^\tau W(t, \Upsilon_0) \left[h_1(t) - \{ 1 - g(t, \Upsilon_0) \} \exp(\beta_0' Z) \int_0^t h_1(s) \mathrm{d}\Lambda_0(s) \right] \right.
$$
$$
\left. \left. \times Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) \mathrm{d}\Lambda_0(t) \right) \right)
$$

Therefore,

$$
||\Sigma(h_{n_k}) - \sigma(h_{n_k}) - \Sigma(h^*) + \sigma(h^*)||_H
$$
$$
\leq \left|\left| E\left([\{ 1 - g(t, \Upsilon_0) \} \exp(\beta_0' Z) \int_0^t (h_{1n_k} - h_1^*)(s) \mathrm{d}\Lambda_0(s) \right.\right.
$$
$$
- (h_{2n_k} - h_2^*)' W(t, \Upsilon_0)] Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) \Big)
$$
$$
+ E\left(\int_0^\tau [(h_{1n_k} - h_1^*)(s) - \{ 1 - g(s, \Upsilon_0) \} \exp(\beta_0' Z) \int_0^s (h_{1n_k} - h_1^*)(u) \mathrm{d}\Lambda_0(u) \right.
$$
$$
\left. \left. + (h_{2n_k} - h_2^*)' W(s, \Upsilon_0)] Y(s) g(s, \Upsilon_0) \{ 1 - g(s, \Upsilon_0) \} \exp(2\beta_0' Z) \mathrm{d}\Lambda_0(s) \right) \right|\right|_H
$$
$$
+ \left|\left| E\left(\int_0^\tau W(t, \Upsilon_0) [(h_{1n_k} - h_1^*)(t) - \{ 1 - g(t, \Upsilon_0) \} \exp(\beta_0' Z) \right.\right.
$$
$$
\left. \left. \times \int_0^t (h_{1n_k} - h_1^*)(s) \mathrm{d}\Lambda_0(s)] Y(t) g(t, \Upsilon_0) \exp(\beta_0' Z) d\Lambda_0(t) \right) \right|\right|_v
$$

Due to the boundedness of $Z, X, \theta_0, Y(t)$ and $\Lambda_0(t)$ and the fact that $0 < g(t, \Upsilon_0) < 1$ on $[0, \tau]$, the right-hand side of the above inequality is bounded
above by

$$
M_1 ||h_{2n_k} - h_2^*||_1 + \int_0^\tau \{ M_2 |(h_{1n_k} - h_1^*)(t)| + M_3 \int_0^t |(h_{1n_k} - h_1^*)(s)| \mathrm{d}\Lambda_0(s) \} \mathrm{d}\Lambda_0(t)
$$

where M_1, M_2 and M_3 are some positive constants. Then applying the dom-
inated convergence theorem, we can show that this sum converges to zero.
Hence, $\Sigma(h) - \sigma(h)$ is a compact linear operator from H_m onto its range for
all finite m. Now following the similar steps (14) and (15) of Scharfsten et al.
(1998) , we can show that $\dot{S}(\Upsilon_0)$ is continuously invertible on its range.

Step 4. To check condition 4, it suffices to show (by Lemma 1 of van der Vaart 1995) that $\mathcal{F} = \{\mathcal{S}^*(\Upsilon)(h) - \mathcal{S}^*(\Upsilon_0)(h) : h \in H_m, ||\Upsilon - \Upsilon_0|| < \epsilon\}$ is Donsker for some $\epsilon > 0$ and $\sup_{h \in H_m} E[\{\mathcal{S}^*(\Upsilon)(h) - \mathcal{S}^*(\Upsilon_0)(h)\}^2]$ converges to 0 as Υ converges to Υ_0. We can write $\mathcal{F} = \mathcal{F}_3 + \mathcal{F}_4$, where

$$\mathcal{F}_3 = \left\{ h_2' \left\{ \int_0^\tau W(t, \Upsilon) \mathrm{d}M(t, \Upsilon) - \int_0^\tau W(t, \Upsilon_0) \mathrm{d}M(t, \Upsilon_0) \right\} : h_2 \in R^{p+q}, ||h_2||_1 \leq m, \right.$$

$$\left. \theta \in [-\epsilon, \epsilon]^{p+q}, \Lambda \text{ nonnegative, increasing with } \Lambda(\tau) \leq 2\Lambda_0(\tau) \right\}$$

$$\mathcal{F}_4 = \left\{ \{\delta + (1 - \delta)g(\tilde{T}, \Upsilon)\} \exp(\beta' Z) \int_0^{\tilde{T}} h_1(t) \mathrm{d}\Lambda(t) \right.$$

$$- \{\delta + (1 - \delta)g(\tilde{T}, \Upsilon)\} \exp(\beta' Z) \int_0^{\tilde{T}} h_1(t) \mathrm{d}\Lambda(t) : h_1 \in BV[0, \tau], ||h_1||_v \leq m,$$

$$\left. \theta \in [-\epsilon, \epsilon]^{p+q}, \Lambda \text{ nonnegative, increasing with } \Lambda(\tau) \leq 2\Lambda_0(\tau) \right\}$$

We want to show that $\mathcal{F}_3$ and $\mathcal{F}_4$ are Donsker with uniformly bounded envelopes. Using the result from empirical process theory, that classes of smooth functions are Donsker (see Theorem 2.7.1 of van der Vaart and Wellner 1996), we know that the following two classes

$$\left\{ \delta W(t, \Upsilon) : \theta \in [-\epsilon, \epsilon]^{p+q}, \Lambda \text{ nonnegative, increasing with } \Lambda(\tau) \leq 2\Lambda_0(\tau) \right\}$$

and

$$\left\{ \int_0^\tau W(t, \Upsilon) Y(t) g(t, \Upsilon) \exp(\beta' Z) \mathrm{d}\Lambda(t) : \theta \in [-\epsilon, \epsilon]^{p+q}, \right.$$

$$\left. \Lambda \text{ nonnegative, increasing with } \Lambda(\tau) \leq 2\Lambda_0(\tau) \right\}$$

are Donsker. In addition, both classes are uniformly bounded due to the boundedness of $Z, X, \theta, Y(t)$ and $\Lambda(t)$. Hence, according to the result, that classes of Lipschitz transformations of Donsker classes with integrable envelope functions are Donsker (see Theorem 2.10.6 of van der Vaart and Wellner (1996), $\mathcal{F}_3$ is Donsker with uniformly bounded envelope.

Theorem 2.10.6 of van der Vaart and Wellner (1996) also implies that the class BV_M of all real-valued functions on $[0, \tau]$, which are uniformly bounded by a constant M and are of variation bounded by M, is Donsker. Since

$$\left\| \int_0^{\tilde{T}} h_1(t) \mathrm{d}\Lambda(t) \right\|_v \leq ||h_1||_\infty \times ||\Lambda||_v < \infty,$$

the class

$$\left\{ \int_0^{\tilde{T}} h_1(t)\mathrm{d}\Lambda(t) : h_1 \in BV[0,\tau], ||h_1||_v \le m, \right.$$

$$\left. \Lambda \text{ nonnegative, increasing with } \Lambda(\tau) \le 2\Lambda_0(\tau) \right\}$$

is Donsker and uniformly bounded. In addition, since $\{\delta + (1 - \delta)g(\tilde{T}, \Upsilon)\}$ $\exp(\beta'Z)$ is uniformly bounded, $\mathcal{F}_4$ is also Donsker with uniformly bounded envelope (see Example 2.10.8 of van der Vaart and Wellner 1996). Thus, we have shown that all the four conditions in Theorem 4 hold. Then, we know that for all finite m

$$-\dot{S}(\Upsilon_0)\sqrt{n}(\hat{\Upsilon}_n - \Upsilon_0)(h) = -\int_0^\tau \sigma_1(h)(t)\mathrm{d}(\sqrt{n}\hat{\Lambda}_n - \Lambda_0)(t) - \sqrt{n}(\hat{\theta}_n - \theta_0)'\sigma_2(h)$$

$$= \sqrt{n}\{S(\hat{\Upsilon}_n) - S(\Upsilon_0)\} + o_{\mathcal{P}*}(1)$$

$$= \sqrt{n}\{S_n(\hat{\Upsilon}_n) - S_n(\Upsilon_0)\} + o_{\mathcal{P}*}(1)$$

$$= -\sqrt{n}\{S_n(\Upsilon_0) - S(\Upsilon_0)\} + o_{\mathcal{P}*}(1)$$

uniformly in $h \in H_m$, where the last equality in the above equation holds since $S_n(\hat{\Upsilon}_n) = S(\Upsilon_0) = 0$. Hence, $\sqrt{n}(\hat{\Upsilon}_n - \Upsilon_0) \Rightarrow \dot{S}(\Upsilon_0)^{-1}G^*$. Following the similar steps of Scharfsten et al. (1998), we can show that $\dot{S}(\Upsilon_0)^{-1}G^* = G$, where G is defined in Theorem 3.   □

Proof of Theorem 5 Since σ is continuously invertible on its range, $\hat{\sigma}$ is continuously invertible on a set of probability going to 1. According to Scharfsten et al. (1998), it suffices to show that $\sup_{||h||_H=1} ||\hat{\sigma}(h) - \sigma(h)||_H$ converges in probability to 0.

The most difficult part in the proof is to deal with the terms in $\hat{\sigma}(h) - \sigma(h)$ involving h_1. For example, one such term can be written as

$$\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[h_1(t) - \{1 - g_i(t, \hat{\Upsilon}_n)\}\exp(\hat{\beta}_n'Z_i)\int_0^t h_1(s)\mathrm{d}\hat{\Lambda}_n(s) \right]$$

$$\times W_i(t, \hat{\Upsilon}_n)Y_i(t)g_i(t, \hat{\Upsilon}_n)\exp(\hat{\beta}_n'Z_i)\mathrm{d}\hat{\Lambda}_n(t)$$

$$-E\left(\int_0^\tau \left[h_1(t) - \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)\mathrm{d}\Lambda_0(s) \right] \right.$$

$$\left. \times W(t, \Upsilon_0)Y_i(t)g(t, \Upsilon_0)\exp(\beta_0'Z)d\Lambda_0(t) \right)$$

The variation of this term can be bounded by the variation of

$$
\frac{1}{n}\sum_{i=1}^{n}\left(\int_0^\tau \left[h_1(t) - \{1 - g_i(t, \hat{\Upsilon}_n)\}\exp(\hat{\beta}_n'Z_i)\int_0^t h_1(s)\mathrm{d}\hat{\Lambda}_n(s)\right]\right.
$$
$$
\times W_i(t, \hat{\Upsilon}_n)Y_i(t)g_i(t, \hat{\Upsilon}_n)\exp(\hat{\beta}_n'Z_i)\mathrm{d}\hat{\Lambda}_n(t)
$$
$$
\left.-\int_0^\tau \left[h_1(t) - \{1 - g_i(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)\mathrm{d}\Lambda_0(s)\right]\right.
$$
$$
\left.\times W_i(t, \Upsilon_0)Y_i(t)g_i(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\Lambda_0(t)\right)
$$

plus the variation of

$$
\frac{1}{n}\sum_{i=1}^{n}\left\{\int_0^\tau \left[h_1(t) - \{1 - g_i(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)\mathrm{d}\Lambda_0(s)\right]\right.
$$
$$
\times W_i(t, \Upsilon_0)Y_i(t)g_i(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\Lambda_0(t)
$$
$$
-E\left(\int_0^\tau \left[h_1(t) - \{1 - g(t, \Upsilon_0)\}\exp(\beta_0'Z)\int_0^t h_1(s)\mathrm{d}\Lambda_0(s)\right]\right.
$$
$$
\left.\left.\times W(t, \Upsilon_0)Y_i(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\mathrm{d}\Lambda_0(t)\right)\right\}
$$

Uniform consistency of $\hat{\Upsilon}$ implies the first term converges to zero. The variation of the second term is bounded by

$$
\int_0^\tau \left\|\frac{1}{n}\sum_{i=1}^{n}W_i(t, \Upsilon_0)Y_i(t)g_i(t, \Upsilon_0)\exp(\beta_0'Z)\right.
$$
$$
\left.-E\{W(t, \Upsilon_0)Y_i(t)g(t, \Upsilon_0)\exp(\beta_0'Z)\}\right\|_v \mathrm{d}\Lambda_0(t) \times \|h_1\|_v
$$
$$
+\int_0^\tau \left\|\frac{1}{n}\sum_{i=1}^{n}W_i(t, \Upsilon_0)Y_i(t)g_i(t, \Upsilon_0)\{1 - g_i(t, \Upsilon_0)\}\exp(2\beta_0'Z)\right.
$$
$$
\left.-E[W(t, \Upsilon_0)Y_i(t)g(t, \Upsilon_0)\{1 - g(t, \Upsilon_0)\}\exp(2\beta_0'Z)]\right\|_v \mathrm{d}\Lambda_0(t) \times \|h_1\|_v\Lambda_0(\tau)
$$

The integrand converges to zero by the strong law of large numbers. Thus the term converges to zero by the dominated convergence theorem, since $\|h_1\|_v \le 1$ and $\Lambda_0(\tau)$ is finite. All other terms in $\hat{\sigma}(h) - \sigma(h)$ can be handled similarly. The details of the proof is omitted here. $\qquad\square$

References

Andersen, P. K., Borgan, Ø., Gill, R., Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer

Ash, R.B. (1972). *Real analysis and probability*. San Diego: Academic

Berkson, J., Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47, 501–515.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore: Johns Hopkins University Press.

Bilias, Y., Gu, M., Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Annals of Statistics*, 25, 662–682.

Chen, K., Jin, Z., Ying, Z. (2002). Semiparametric of transformation models with censored data. *Biometrika*, 89, 659–668.

Cox, D. R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society Series B*, 34, 187–220.

Fang, H. B., Li, G., Sun, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics*, 32, 59–75.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38, 1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *Canadian Journal of Statistics*, 14, 257–262.

Fleming, T. R., Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

Kress, R. (1989). *Linear integral equations*. New York: Springer.

Kuk, A. Y. C., Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79, 531–541.

Loeve, M. (1963). *Probability theory*. Princeton: D. Van Nordstrand Company.

Lu, W., Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, 91, 331–343.

Luenberger, D. G. (1969). *Optimization by vector space methods*. New York: Wiley.

Maller, R. A., Zhou, S. (1996). *Survival analysis with long term survivors*. New York: Wiley.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics*, 22, 721–731.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Annals of Statistics*, 23, 182–198.

Peng, Y., Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56, 237–243.

Scharfsten, D. O., Tsiatis, A. A., Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*, 4, 355–391.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Shorack, G. R., Wellner, J. A. (1986). *Empirical processes with applications to statistics*. New York: Wiley.

Sy, J. P., Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56, 227–236.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, 9, 93–108.

Tsodikov, A. D. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, 54, 138–146.

Tsodikov, A. D. (2001). Estimation of survival based on proportional hazards when cure is a possibility. *Mathematical and Computer Modelling*, 33, 1227–1236.

Tsodikov, A. D., Ibrahim, J. G., Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 98, 1063–1078.

van der Vaart, A. W. (1995). Efficiency of infinite dimensional M-estimators. *Statistica Neerlandica*, 49, 9–30.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical Processes*. New York: Springer.

Zeng, D., Ying, G., Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101, 670–684.