EFFICIENT ESTIMATION FOR AN ACCELERATED TIME FAILURE MODEL WITH A CURE FRACTION

Wenbin Lu

North Carolina State University

Supplementary Material

Throughout the paper, $b^{(k)}(x)$ of a function $b(x)$ denotes its kth derivative. To established the asymptotic results given in Theorems 3.1. and 3.2., we need the following regularity conditions:

(C1) The covariate vectors Z and X have bounded support. The design matrices formed by the column vectors of Z and X are of full rank.

(C2) The true value θ_0 belongs to the interior of a known compact set $\mathcal{B}$ in $\mathcal{R}^{p+q}$, and $\lambda_0(x)$ is positive and thrice-continuously differentiable.

(C3) The censoring time C has a positive and twice-continuously differentiable density function.

(C4) With probability one, there exists a positive constant ζ_0 such that $P\,(Ce^{-\beta_0' Z} \geq e^{\epsilon} \geq \tau | Z,\, X) > \zeta_0$ for all possible values of Z and X, where τ is a finite positive number.

(C5) The kernel function $K(\cdot)$ is thrice-continuously differentiable, and $K^{(r)}(\cdot)$, $r = 0, 1, 2, 3$, have bounded variations in $\mathcal{R}$. In addition, the first $(m-1)$ moments of $K^{(m)}(\cdot)$ are 0 for some $m > 3$.

The conditions (C1), (C2), (C3) and (C5) are similar to those used by Zeng and Lin (2007) for establishing the asymptotic results of the nonparametric maximum likelihood estimates under the usual accelerated failure time model. Condition (C4) is needed for the identifiability of the accelerated failure time model mixture cure model on the interval $[0, \tau]$.

Proof of Theorem 3.1. Let $(\hat\beta_n, \hat\gamma_n, \hat\lambda_n(\cdot))$ denote the estimates of $(\beta_0, \gamma_0, \lambda_0(\cdot))$ obtained from the proposed EM algorithm. We have

$$\hat\Lambda_n(x) = \int_{-\infty}^{\log x} \frac{n^{-1}\sum_{j=1}^n \delta_j K_h\{R_j(\hat\beta_n) - s\}}{n^{-1}\sum_{j=1}^n \hat w_{n,j} \int_{-\infty}^{R_j(\hat\beta_n)-s} K_h(u)du} ds,\ x > 0.$$

where $\hat\Lambda_n(x) = \int_0^x \hat\lambda_n(u)du$, $\hat S_n(x) = \exp\{-\hat\Lambda_n(x)\}$ and

$$\hat w_{n,i} = \delta_i + (1-\delta_i)\frac{\pi(\hat\gamma_n' X_i)\hat S_n\{e^{R_i(\hat\beta_n)}\}}{1 - \pi(\hat\gamma_n' X_i) + \pi(\hat\gamma_n' X_i)\hat S_n\{e^{R_i(\hat\beta_n)}\}},\ i = 1, \cdots, n.$$

In addition, define

$$w_{i0} = \delta_i + (1-\delta_i)\frac{\pi(\gamma_0' X_i) S_0\{e^{R_i(\beta_0)}\}}{1-\pi(\gamma_0' X_i) + \pi(\gamma_0' X_i) S_0\{e^{R_i(\beta_0)}\}}, \ \ i = 1, \cdots, n,$$

$$\tilde{\Lambda}_n(x) = \int_{-\infty}^{\log x} \frac{n^{-1}\sum_{j=1}^n \delta_j K_h\{R_j(\beta_0) - s\}}{n^{-1}\sum_{j=1}^n w_{j0}\int_{-\infty}^{R_j(\beta_0)-s} K_h(u)du} ds, \ \ x > 0.$$

As discussed by Zeng and Lin (2007), following lemma 2.4 of Schuster (1969) and theorem 2.4.3 of van der Vaart and Wellner (1996), we can show that

$$\sup_{\theta\in\mathcal{B},s}\left|\frac{1}{n}\sum_{j=1}^n \delta_j K_h\{R_j(\beta)-s\} - \frac{dP\left(\delta=1, R(\beta)\le s\right)}{ds}\right| \to 0, \ \text{a.s.} \tag{S.1}$$

Note that

$$\frac{dP\left(\delta=1, R(\beta_0)\le s\right)}{ds} = \frac{dP\left(R^*(\beta_0)\le s\right)}{ds} E_{Z,X}\{\pi(\gamma_0' X) P\left(C(\beta_0)\ge s|Z,X\right)\}$$

where $R^*(\beta) = \log(T^*) - \beta' Z$, $C(\beta) = \log(C) - \beta' Z$ and the expectation $E_{Z,X}$ is taken with respect to the random variables Z and X. Moreover, we have

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^n w_{j0}\int_{-\infty}^{R_j(\beta_0)-s} K_h(u)du = \frac{1}{n}\sum_{j=1}^n \delta_j \int_{-\infty}^{R_j(\beta_0)-s} K_h(u)du \\ + \quad &\frac{1}{n}\sum_{j=1}^n (1-\delta_j)\frac{\pi(\gamma_0' X_j) S_0\{e^{R_j(\beta_0)}\}}{1-\pi(\gamma_0' X_j) + \pi(\gamma_0' X_j) S_0\{e^{R_j(\beta_0)}\}}\int_{-\infty}^{R_j(\beta_0)-s} K_h(u)du. \end{aligned}$$

Using the similar techniques, we can show that the first term on the right-hand side of the above equality converges uniformly in s to

$$E_{Z,X}\left[\pi(\gamma_0' X) P\left(C(\beta_0)\ge s|Z,X\right)\{P\left(R^*(\beta_0)\ge s\right) - P\left(R^*(\beta_0)\ge C(\beta_0)\right)\}\right]$$

and the second term converges uniformly in s to

$$E_{Z,X}\left[\pi(\gamma_0' X) S_0\{e^{C(\beta_0)}\} P\left(C(\beta_0)\ge s|Z,X\right)\right]$$

which is due to a fact that $E(1-\delta|C,Z,X) = 1-\pi(\gamma_0' X) + \pi(\gamma_0' X) S_0\{e^{C(\beta_0)}\}$.

Thus, $n^{-1}\sum_{j=1}^n w_{j0}\int_{-\infty}^{R_j(\beta_0)-s} K_h(u)du$ converges uniformly in s to

$$P\left(R^*(\beta_0)\ge s\right) E_{Z,X}[\pi(\gamma_0' X) P\left(C(\beta_0)\ge s|Z,X\right)]$$

since $S_0\{e^{C(\beta_0)}\} = P\left(R^*(\beta_0)\ge C(\beta_0)\right)$. It follows that

$$\sup_s \left|\frac{n^{-1}\sum_{j=1}^n \delta_j K_h\{R_j(\beta_0)-s\}}{n^{-1}\sum_{j=1}^n w_{j0}\int_{-\infty}^{R_j(\beta_0)-s} K_h(u)du} - \frac{dP\left(\epsilon\le s\right)/ds}{P\left(\epsilon\ge s\right)}\right| \to 0, \ \text{a.s.} \tag{S.2}$$

and $\tilde{\Lambda}_n(x) \to \Lambda_0(x)$, a.s. as $n\to\infty$. In addition, the pointwise convergence can be strengthened to the uniform convergence by applying the same monotonicity argument used in the proof of the Glivenko-Cantelli Theorem [Page 96 of Shorack and Wellner (1986)], i.e. $\sup_{x\in[0,\tau]}|\tilde{\Lambda}_n(x)\rightarrow\Lambda_0(x)|\rightarrow 0$, a.s..

Next we show that $(\hat{\Lambda}_n,\hat{\theta}_n)$ converges to (Λ_0,θ_0). Since every bounded sequence in R^{p+q} has a convergent subsequence, there exists a θ^* such that $\hat{\theta}_{m_k}\rightarrow\theta^*$. 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 $\hat{\Lambda}_{n_k}(x)\rightarrow\Lambda^*(x)$ for all $x\in[0,\tau]$ at which Λ^* is continuous. Therefore, $(\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})$ must converge to (Λ^*,θ^*). Recall that

$$\hat{\lambda}_{n_k}(x)=\frac{(nx)^{-1}\sum_{j=1}^n\delta_jK_h\{R_j(\hat{\beta}_{n_k})-\log x\}}{n^{-1}\sum_{j=1}^n\hat{w}_{n_k,j}\int_{-\infty}^{R_j(\hat{\beta}_{n_k})-\log x}K_h(u)du}.$$

Let $k\rightarrow\infty$ in the above equality. Using the similar techniques for proving (S.1) and (S.2), we can show that

$$\lambda^*(x)=\frac{E\{dN(x,\beta^*)\}}{Y(x,\beta^*)g(x,\theta^*,\Lambda^*)dx},$$

where $N(x,\beta)=\delta I(e^{R(\beta)}\le x)$, $Y(x,\beta)=I(e^{R(\beta)}\ge x)$ and

$$g(x,\theta,\Lambda)=\frac{\pi(\gamma'X)S(x)}{1-\pi(\gamma'X)+\pi(\gamma'X)S(x)}.$$

Thus, for any $x\in[0,\tau]$,

$$E\{dN(x,\beta^*)-Y(x,\beta^*)g(x,\theta^*,\Lambda^*)\lambda^*(x)dx\}=0. \tag{S.3}$$

Also note that

$$\begin{aligned}0&\le n_k^{-1}l_{n_k}^o(\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})-n_k^{-1}l_{n_k}^o(\tilde{\Lambda}_{n_k},\theta_0)\\&=\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^\tau\log\{\hat{\chi}_{n_k,i}(x)\}\{dN_i(x,\hat{\beta}_{n_k})-Y_i(x,\hat{\beta}_{n_k})g_i(x,\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})\hat{\lambda}_{n_k}(x)dx\}\\&\quad+\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^\tau\log\{\tilde{\chi}_{n_k,i}(x)\}\{dN_i(x,\beta_0)-Y_i(x,\beta_0)g_i(x,\tilde{\Lambda}_{n_k},\theta_0)\tilde{\lambda}_{n_k}(x)dx\}\\&\quad+\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^\tau[\log\{\chi_{n_k,i}(x)\}-\{\chi_{n_k,i}(x)-1\}]Y_i(x,\beta_0)g_i(x,\tilde{\Lambda}_{n_k},\theta_0)\tilde{\lambda}_{n_k}(x)dx,\end{aligned}$$

where $\hat{\chi}_{n_k,i}(x)=e^{-\hat{\beta}_{n_k}'Z_i}g_i(x,\hat{\Lambda}_{n_k},\hat{\theta}_{n_k})\hat{\lambda}_{n_k}(x)$, $\tilde{\chi}_{n_k,i}(x)=e^{-\beta_0'Z_i}g_i(x,\tilde{\Lambda}_{n_k},\theta_0)\tilde{\lambda}_{n_k}(x)$ and $\chi_{n_k,i}(x)=\hat{\chi}_{n_k,i}(x)/\tilde{\chi}_{n_k,i}(x)$. Based on (S.3) and the fact that

$$M(x)\equiv N(x,\beta_0)-\int_0^xY(u,\beta_0)g(u,\Lambda_0,\theta_0)\lambda_0(u)du$$

is a mean-zero martingale process, and applying the Glivenko-Cantelli Theorem, we have that the first two terms on the right-hand side of the above inequality converge to zero. The third term is less or equal to zero since for $x>0$, $\log(x)-(x-1)\le 0$, which converges to

$$\begin{aligned}&E\Big(\int_0^\tau\Big[\log\Big\{\frac{e^{-(\beta^*)'Z}g(x,\Lambda^*,\theta^*)\lambda^*(x)}{e^{-\beta_0'Z}g(x,\Lambda_0,\theta_0)\lambda_0(x)}\Big\}-\Big\{\frac{e^{-(\beta^*)'Z}g(x,\Lambda^*,\theta^*)\lambda^*(x)}{e^{-\beta_0'Z}g(x,\Lambda_0,\theta_0)\lambda_0(x)}-1\Big\}\Big]\\&\times Y(x,\beta_0)g(x,\Lambda_0,\theta_0)\lambda_0(x)dx\Big).\end{aligned}$$

Note that the above limit is the negative Kullback-Leibler information, $E[l^o(\Lambda^*, \theta^*)] - E[l^o(\Lambda_0, \theta_0)]$, where

$$l^o(\theta, \Lambda) = \int_0^\tau \log\{e^{-\beta' Z} g(x, \Lambda, \theta)\lambda(x)\} dN(x, \beta) - \int_0^\tau Y(x, \beta) g(x, \Lambda, \theta)\lambda(x) dx.$$

Thus, the Kullback-Leibler information must equal zero. Then, with probability one, we have

$$\begin{aligned}
& \int_0^\tau \log\{e^{-(\beta^*)' Z} g(x, \Lambda^*, \theta^*)\lambda^*(x)\} dN(x, \beta^*) - \int_0^\tau Y(x, \beta^*) g(x, \Lambda^*, \theta^*)\lambda^*(x) dx \\
= & \int_0^\tau \log\{e^{-\beta_0' Z} g(x, \Lambda_0, \theta_0)\lambda_0(x)\} dN(x, \beta_0) - \int_0^\tau Y(x, \beta_0) g(x, \Lambda_0, \theta_0)\lambda_0(x) dx.
\end{aligned}$$

This implies that $\Lambda^* = \Lambda_0$ and $\theta^* = \theta_0$ with probability one. Therefore, $(\hat{\Lambda}_{n_k}, \hat{\theta}_{n_k})$ must converge to (Λ_0, θ_0). By Helly's theorem, we know that $(\hat{\Lambda}_n, \hat{\theta}_n)$ must converge to (Λ_0, θ_0). Furthermore, the point-wise convergence can be strengthened to the uniform convergence by applying the Glivenko-Cantelli Theorem [Shorack and Wellner (1996)]. ♯

Proof of Theorem 3.2. Following Theorem 3.3.1. of van der Vaart and Wellner (1996), we first derive the score operators based on some submodels. To be specific, set $\Lambda_d(x) = \int_0^x \{1 + dh_1(u)\} d\hat{\Lambda}_n(u)$ and $\theta_d = dh_2 + \hat{\theta}_n$, where $h_1(u)$ is a function on $[0, \tau]$ 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 $U_n(\hat{\Lambda}_n, \hat{\theta}_n)(h_1, h_2)$ denote the derivative of $n^{-1} l_n^o(\theta_d, \Lambda_d)$ with respect to d and evaluated at $d = 0$. We have $U_n(\hat{\Lambda}_n, \hat{\theta}_n)(h_1, h_2) = U_{n1}(\hat{\Lambda}_n, \hat{\theta}_n)(h_1) + U_{n2}(\hat{\Lambda}_n, \hat{\theta}_n)(h_2)$, where

$$\begin{aligned}
U_{n1}(\hat{\Lambda}_n, \hat{\theta}_n)(h_1) &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left[h_1(x) - \{1 - g_i(x, \hat{\Lambda}_n, \hat{\theta}_n)\} \int_0^x h_1(u) d\hat{\Lambda}_n(u)\right] \\
&\quad \times \left\{dN_i(x, \hat{\beta}_n) - Y_i(x, \hat{\beta}_n) g_i(x, \hat{\Lambda}_n, \hat{\theta}_n) d\hat{\Lambda}_n(x)\right\},
\end{aligned}$$

$$U_{n2}(\hat{\Lambda}_n, \hat{\theta}_n)(h_2) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau h_2' W_i(x, \hat{\Lambda}_n, \hat{\theta}_n)\{dN_i(x, \hat{\beta}_n) - Y_i(x, \hat{\beta}_n) g_i(x, \hat{\Lambda}_n, \hat{\theta}_n) d\hat{\Lambda}_n(x)\},$$

and $W_i(x, \Lambda, \theta) = [-Z_i' w_i(x, \Lambda, \theta), X_i'\{1 - g_i(x, \Lambda, \theta)\}]'$ with

$$w_i(x, \Lambda, \theta) = x\frac{\lambda^{(1)}(x)}{\lambda(x)} + 1 - x\lambda(x)\{1 - g_i(x, \Lambda, \theta)\}.$$

Note that $U_n(\hat{\Lambda}_n, \hat{\theta}_n)(h_1, h_2) = 0$ for all (h_1, h_2) since $(\hat{\Lambda}_n, \hat{\theta}_n)$ maximizes l_n^o.

Let $BV[0, \tau]$ denote the space of bounded variation functions defined on $[0, \tau]$. We assume that the class of $h = (h_1, h_2)$ 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 $\langle \Lambda, \theta \rangle(h) = \int_0^\tau h_1(t) d\Lambda(t) + h_2'\theta$. 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 U_n is a random map from 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.

Define $U(\Lambda,\theta)(h) = E\{U_n(\Lambda,\theta)(h)\}$. Accordingly, write $U(\Lambda,\theta)(h) = U_1(\Lambda,\theta)(h_1) + U_2(\Lambda,\theta)(h_2)$. In addition, define

$$\begin{aligned} U_1^*(\Lambda,\theta)(h_1) &= \int_0^\tau \left[h_1(x) - \{1-g(x,\Lambda,\theta)\}\int_0^x h_1(u)d\Lambda(u)\right] \\ &\quad \times \{dN(x,\beta) - Y(x,\beta)g(x,\Lambda,\theta)d\Lambda(x)\}, \end{aligned}$$

$$U_2^*(\Lambda,\theta) = \int_0^\tau W(x,\Lambda,\theta)\{dN(x,\beta) - Y(x,\beta)g(x,\Lambda,\theta)d\Lambda(x)\}.$$

Conditions (C1)-(C3) imply that the class $\{U_1^*(\Lambda_0,\theta_0)(h_1) : h_1 \in BV[0,\tau], ||h_1||_v \le m\}$ can be written as the summation of bounded Donsker classes, and thus it is also a Donsker class [van der Vaart and Wellner (1996)]. Moreover, the class $\{h_2' U_2^*(\Lambda_0,\theta_0)(h_2) : h_2 \in R^{p+q},\ ||h_2||_1 \le m\}$ can also be shown to be Donsker since $U_2^*(\Lambda_0,\theta_0)$ is a bounded function. Thus, $n^{1/2}\{U_n(\Lambda_0,\theta_0)(h) - U(\Lambda_0,\theta_0)(h)\}$ converges weakly to a tight Gaussian process $G^*(h)$ on $l^\infty(H_m)$.

Next, we derive the information operator $I(\Lambda,\theta)(h)$. To do this, we write $U(\Lambda,\theta)$ linearly in $\Lambda-\Lambda_0$ and $\theta-\theta_0$. Define $\Upsilon = (\Lambda,\theta)$ and $\Upsilon_0 = (\Lambda_0,\theta_0)$. After some calculations, we have

$$\begin{aligned} U_1(\Upsilon)(h_1) &= -(\theta-\theta_0)'E\Big(\int_0^\tau W(t,\Upsilon_0)[h_1(x) - \{1-g(x,\Upsilon_0)\}\int_0^x h_1(u)d\Lambda_0(u)] \\ &\quad \times Y(x,\beta_0)g(x,\Upsilon_0)d\Lambda_0(t)\Big) \\ &\quad - E\Big(\int_0^\tau [h_1(x) - \{1-g(x,\Upsilon_0)\}\int_0^x h_1(u)d\Lambda_0(u)] \\ &\quad \times Y(x,\beta_0)g(x,\Upsilon_0)d\{\Lambda(x)-\Lambda_0(x)\}\Big) \\ &\quad + E\Big(\int_0^\tau [h_1(x) - \{1-g(x,\Upsilon_0)\}\int_0^x h_1(u)d\Lambda_0(u)] \\ &\quad \times \{\Lambda(x)-\Lambda_0(x)\}Y(x,\beta_0)g(x,\Upsilon_0)\{1-g(x,\Upsilon_0)\}d\Lambda_0(x)\Big) \\ &\quad + \mathrm{error}_1(\Upsilon)(h), \end{aligned}$$

$$\begin{aligned} U_2(\Upsilon)(h) &= -(\theta-\theta_0)'E\Big\{\int_0^\tau W(x,\Upsilon_0)W'(x,\Upsilon_0)h_2 d\Lambda_0(x)\Big\} \\ &\quad - E\Big[\int_0^\tau h_2'W(x,\Upsilon_0)Y(x,\beta_0)g(x,\Upsilon_0)d\{\Lambda(x)-\Lambda_0(x)\}\Big] \\ &\quad + E\Big[\int_0^\tau h_2'W(x,\Upsilon_0)\{\Lambda(x)-\Lambda_0(x)\} \\ &\quad \times Y(x,\beta_0)g(x,\Upsilon_0)\{1-g(x,\Upsilon_0)\}d\Lambda_0(x)\Big] + \ \mathrm{error}_2(\Upsilon)(h). \end{aligned}$$

In addition, since the functions involved in U_1 and U_2 are all bounded, it is easy to show that, as $||\theta-\theta_0||_1 \to 0$ and $||\Lambda-\Lambda_0|| \equiv \sup_{x\in[0,\tau]} |\Lambda(x)-\Lambda_0(x)| \to 0$,

$$\frac{\sup_{h\in H_m} |\mathrm{error}_i(\Upsilon)(h)|}{||\theta-\theta_0||_1 + ||\Lambda-\Lambda_0||} \to 0,\ i=1,2. \tag{S.4}$$

Thus, the information operator is given by

$$I(\Upsilon)(h) = \int_0^\tau \sigma_1(h)(x) d\Lambda(x) + \theta' \sigma_2(h), \tag{S.5}$$

where

$$\begin{aligned} \sigma_1(h)(x) &= E\{V(x,\Upsilon_0)(h) Y(x,\beta_0) g(x,\Upsilon_0)\} \\ &\quad - E[\int_x^\tau V(s,\Upsilon_0)(h) Y(s,\beta_0) g(s,\Upsilon_0)\{1-g(s,\Upsilon_0)\} d\Lambda_0(s)], \end{aligned}$$

$$\sigma_2(h) = E[\int_0^\tau W(x,\Upsilon_0) V(x,\Upsilon_0)(h) Y(x,\beta_0) g(x,\Upsilon_0) d\Lambda_0(x)],$$

and

$$V(x,\Upsilon) = h_1(x) - \{1-g(t,\Upsilon)\} \int_0^x h_1(u) d\Lambda(u) + h_2' W(t,\Upsilon).$$

In addition, based on (S.4), the uniform consistency of $\hat{\Upsilon} \equiv (\hat{\Lambda}, \hat{\theta})$, and the fact $U(\Upsilon_0) = 0$, we have $||U(\hat{\Upsilon})(h) - U(\Upsilon_0)(h) + I(\hat{\Upsilon} - \Upsilon_0)(h)||$ is $o_{\mathcal{P}^*}(||\hat{\theta} - \theta_0||_1 + ||\hat{\Lambda} - \Lambda_0||)$. By the kernel approximation and the zero-moments condition of $K(\cdot)$ as discussed in Zeng and Lin (2007), the above result can be further strengthened as $||\sqrt{n}\{U(\hat{\Upsilon})(h) - U(\Upsilon_0)(h) + I(\hat{\Upsilon} - \Upsilon_0)(h)\}|| = o_{\mathcal{P}^*}(1 + \sqrt{n}||\hat{\Upsilon} - \Upsilon_0||)$, where $||\hat{\Upsilon} - \Upsilon_0|| = ||\hat{\theta} - \theta_0||_1 + ||\hat{\Lambda} - \Lambda_0||$.

In the third step, we show that $I(\Upsilon)(h)$ is continuously invertible. Following Murphy, Rossini and Van der Vaart (1997) and Scharfsten, Tsiatis and Gilbert (1998), it suffices to show that $\sigma(h) = (\sigma_1(h), \sigma_2(h))$ is invertible almost everywhere $(d\Lambda_0)$. To do this, we first show that $\sigma(h)$ is one-to-one almost everywhere $(d\Lambda_0)$, i.e.

$$\int_0^\tau \sigma_1(h)(x) h_1(x) d\Lambda_0(x) + h_2' \sigma_2(h) = 0 \tag{S.6}$$

implies that $h_2 = 0$ and $h_1(x) = 0$ almost everywhere $(d\Lambda_0)$. Plug $\sigma_1(h)(x)$ and $\sigma_2(h)$ into (S.6), we have

$$\begin{aligned} 0 &= E\Big(\int_0^\tau h_1(x)\Big[V(x,\Upsilon_0) Y(x,\beta_0) g(x,\Upsilon_0) \\ &\quad - \int_x^\tau V(s,\Upsilon_0) Y(x,\beta_0) g(s,\Upsilon_0)\{1-g(s,\Upsilon_0)\} d\Lambda_0(s)\Big] d\Lambda_0(x)\Big) \\ &\quad + h_2' E \int_0^\tau \Big\{W(x,\Upsilon_0) V(x,\Upsilon_0) Y(x,\beta_0) g(x,\Upsilon_0) d\Lambda_0(x)\Big\} \\ &= E\Big[\int_0^\tau \{h_1(x) + h_2' W(x,\Upsilon_0)\} V(x,\Upsilon_0) Y(x,\beta_0) g(x,\Upsilon_0) d\Lambda_0(x)\Big] \\ &\quad - E\Big[\int_0^\tau \int_0^x h_1(u) d\Lambda_0(u) V(x,\Upsilon_0) Y(x,\beta_0) g(x,\Upsilon_0)\{1-g(x,\Upsilon_0)\} d\Lambda_0(x)\Big] \\ &= E\left[\int_0^\tau V^{\otimes 2}(x,\Upsilon) Y(x,\beta_0) g(x,\Upsilon_0) d\Lambda_0(x)\right]. \end{aligned}$$

where $a^{\otimes 2} = aa'$ for a real vector a. Since $Y(x,\beta_0) g(x,\Upsilon_0) > 0$ a.e. on $[0,\tau]$, it implies $V(x,\Upsilon_0) = 0$ a.e. $(d\Lambda_0)$. Therefore, with probability one $(d\Lambda_0)$,

$$h_1(x) + \{1 - g(x,\Upsilon_0)\}\left[h_{22}' X - \int_0^x \{h_{21}' Z + h_1(u)\} d\Lambda_0(u)\right] = -h_{21}' Z.$$

From this, h_{21} must be zero. It further implies that

$$\frac{h_1(x)}{1-g(x,\Upsilon_0)} - \int_0^x h_1(u)d\Lambda_0(u) = -h_{22}'X$$

a.e. $(d\Lambda_0)$. Similarly, h_{22} must be zero. With $h_2 = (h_{21}', h_{22}')' = 0$, we have $h_1(x) - \{1 - g(x,\Upsilon_0)\}\int_0^x h_1(u)d\Lambda_0(u) = 0$ 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. $(d\Lambda_0)$.

Then, we want to show that $\sigma(h)$ has a continuous inverse. To show that σ is invertible, since $\sigma(h)$ 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)]. This can be done following the similar techniques used in the proof of Theorem 3 of Lu (2008) and it is omitted here. Since $\sigma(h)$ is invertible, its inverse will be continuous [see page 149, Luenberger (1969)]. Thus, $I(\Upsilon)(h)$ is continuously invertible on its range.

In the fourth step, we establish the asymptotic distributions of $\hat{\Upsilon}$. To do this, we need to show that $||\sqrt{n}\{(U_n - U)(\hat{\Upsilon}_n) - (U_n - U)(\Upsilon_0)\}|| = o_{\mathcal{P}^*}(1 + \sqrt{n}||\hat{\Upsilon}_n - \Upsilon_0||)$. Based on Lemma 1 of van der Vaart (1995), it suffices to show that $\mathcal{F} \equiv \{\mathcal{U}^*(\Upsilon)(h) - \mathcal{U}^*(\Upsilon_0)(h) : h \in H_m, ||\Upsilon - \Upsilon_0|| < \epsilon\}$ is Donsker for some $\epsilon > 0$ and $\sup_{h\in H_m} E[\{\mathcal{U}^*(\Upsilon)(h) - \mathcal{U}^*(\Upsilon_0)(h)\}^2]$ converges to 0 as Υ converges to Υ_0, where $\mathcal{U}^*(\Upsilon)(h) = U_1^*(\Upsilon)(h_1) + h_2'U_2^*(\Upsilon)$. To prove this, we write $\mathcal{F}$ as the summation of two Donsker classes with uniformly bounded envelopes. Then based on 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)], $\mathcal{F}$ is Donsker. In addition, by the Dominant Convergence Theorem, it can be shown that $\sup_{h\in H_m} E[\{\mathcal{U}^*(\Upsilon)(h) - \mathcal{U}^*(\Upsilon_0)(h)\}^2]$ converges to 0 as Υ converges to Υ_0. The detailed proofs are similar to those in Murphy, Rossini and Van der Vaart (1997, Theorem 2.2) and in Lu (2008, Theorem 3) and are omitted here. Then, based on Theorem 3.3.1. of van der Vaart and Wellner (1996) and the facts that $U_n(\hat{\Upsilon}_n) = U(\Upsilon_0) = 0$, we have

$$\begin{aligned} -I(\sqrt{n}(\hat{\Upsilon}_n - \Upsilon_0))(h) &= \sqrt{n}\{U(\hat{\Upsilon}_n) - U(\Upsilon_0)\} + o_{\mathcal{P}^*}(1) \\ &= \sqrt{n}\{U_n(\hat{\Upsilon}_n) - U_n(\Upsilon_0)\} + o_{\mathcal{P}^*}(1) \\ &= -\sqrt{n}\{U_n(\Upsilon_0) - U(\Upsilon_0)\} + o_{\mathcal{P}^*}(1) \end{aligned}$$

Hence, $\sqrt{n}(\hat{\Upsilon}_n - \Upsilon_0)$ converges weakly to a tight Gaussian process $G = I^{-1}G^*$. In addition, the variance of G is given by

$$\mathrm{Var}\{G(h)\} = \int_0^\tau h_1(t)\sigma_{(1)}^{-1}(h)(t)d\Lambda_0(t) + h_2'\sigma_{(2)}^{-1}(h), \tag{S.7}$$

where $\sigma^{-1}(h) = (\sigma_{(1)}^{-1}, \sigma_{(2)}^{-1})(h)$ is the inverse of $\sigma(h)$.

Finally, we show that the variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ achieve the semiparametric efficiency bound [Bickel, Klaassen, Ritov, Y. and Wellner, J. A. (1993)]. We first calculate the efficiency bound by constructing the efficient score function using the projection method. To be specific, the efficient score U_{eff} for θ can be defined as $U_{eff} = U_\theta - \Pi[U_\theta|\Theta]$, where $\Pi[\cdot|\cdot]$ is the projection operator and Θ is the tangent set given by

$$\begin{aligned}\Theta &= \{\kappa : \kappa = \int_0^\tau [a(x) - \{1 - g(x, \Upsilon_0)\} \int_0^x a(u) d\Lambda_0(u)] dM(x), \text{ where} \\ &\quad a(x) \text{ is any } (p+q) - \text{dimensional function of } x,\ E[||\kappa||^2] < \infty\}.\end{aligned}$$

By some simple algebra, we can show that the vector $a(t)$ corresponding to the efficient score satisfies

$$a(t) - \int_0^\tau \Psi(t,s) a(s) d\Lambda_0(s) = \rho(t), \quad t \in [0, \tau], \tag{S.8}$$

where for $0 \le t,\ s \le \tau,\ s \vee t \equiv \max(s,t)$,

$$\begin{aligned}\Psi(t,s) &= \frac{E[g(s\vee t, \Upsilon_0)\{1 - g(s \vee t, \Upsilon_0)\} Y(s \vee t, \beta_0)]}{E\{Y(t,\beta_0) g(t, \Upsilon_0)\}} - \\ &\quad \frac{\int_{s\vee t}^\tau E[g(u,\Upsilon_0)\{1 - g(u, \Upsilon_0)\}^2 Y(u, \beta_0)] d\Lambda_0(u)}{E\{Y(t,\beta_0) g(t,\Upsilon_0)\}},\end{aligned}$$

$$\begin{aligned}\rho(t) &= \frac{E\{W(t,\Upsilon_0) Y(t,\beta_0) g(t,\Upsilon_0)\}}{E\{Y(t,\beta_0) g(t,\Upsilon_0)\}} - \\ &\quad \frac{\int_t^\tau E[W(s,\Upsilon_0) g(s,\Upsilon_0)\{1 - g(s,\Upsilon_0)\} Y(s,\Upsilon_0)] d\Lambda_0(s)}{E\{Y(t,\beta_0) g(t,\Upsilon_0)\}}.\end{aligned}$$

Note that (S.8) is a Fredholm integral equation of the second kind and it has a unique solution if $\sup_{t\in[0,\tau]} \int_0^\tau |\Psi(t,s)| d\Lambda_0(s) < \infty$ [Kress (1989)], which is true under our assumed conditions. Let $a_{eff}(t)$ denote the solution to (S.8). Then the efficient score for θ is

$$U_{eff} = \int_0^\tau [W(t,\Upsilon_0) - a_{eff}(t) + \{1 - g(t,\Upsilon_0)\} \int_0^t a_{eff}(s) d\Lambda_0(s)] dM(t).$$

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

$$\begin{aligned}\Phi^{-1} &= E\Big(\int_0^\tau W(t,\Upsilon_0)[W(t,\Upsilon_0) - a_{eff}(t) + \{1 - g(t,\Upsilon_0)\} \int_0^t a_{eff}(s) d\Lambda_0(s)]' \\ &\quad \times Y(t,\beta_0) g(t,\Upsilon_0) d\Lambda_0(t)\Big),\end{aligned}$$

Next, we show that the variance of $\sqrt{n}(\hat\theta_n - \theta_0)$ achieve the semiparametric efficiency bound Φ. By the Cramer-Wold device (see Serfling (1980)), it suffices to demonstrate that the asymptotic variance of $c'\sqrt{n}(\hat\theta_n - \theta_0)$ is equal to $c'\Phi c$, where c 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 $\sigma_2(h) = c$. Consider the solution, $h_2 = \Phi c$ and $h_1(t) = -a'_{eff}(t) A^{-1}(B\Phi - I)c$, where

$$B = E\left\{\int_0^\tau W^{\otimes 2}(t,\Upsilon_0) Y(t,\beta_0) g(t,\Upsilon_0) d\Lambda_0(t)\right\},$$

$$
\begin{aligned}
A = E\bigg(\int_0^\tau W(t,\Upsilon_0)[a_{eff}(t) - \{1-g(t,\Upsilon_0)\} \\
\times \int_0^t a_{eff}(s)d\Lambda_0(s)]'Y(t,\beta_0)g(t,\Upsilon_0)d\Lambda_0(t)\bigg).
\end{aligned}
$$

Note that with the h defined above

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

$$
\begin{aligned}
\sigma_1(h)(t) &= E\Big([-a_{eff}(t) + \{1-g(t,\Upsilon_0)\}\int_0^t a_{eff}(s)d\Lambda_0(s)] \\
&\qquad \times Y(t,\beta_0)g(t,\Upsilon_0)\Big)A^{-1}(B\Phi - I)c \\
&\quad + E\{W(t,\Upsilon_0)Y(t,\beta_0)g(t,\Upsilon_0)\}\Phi c \\
&\quad - E\Big(\int_t^\tau [-a_{eff}(s) + \{1-g(s,\Upsilon_0)\}\int_0^s a_{eff}(u)d\Lambda_0(u)] \\
&\qquad \times Y(s,\beta_0)g(s,\Upsilon_0)\{1-g(s,\Upsilon_0)\}d\Lambda_0(s)\Big)A^{-1}(B\Phi - I)c \\
&\quad - E\Big[\int_t^\tau W(s,\Upsilon_0)Y(s,\beta_0)g(s,\Upsilon_0)\{1-g(s,\Upsilon_0)\}d\Lambda_0(s)\Big]\Phi c \\
&= 0,
\end{aligned}
$$

since $A^{-1}(B\Phi - I) = \Phi$ and $a_{eff}(t)$ is the solution to (S.8). Therefore, we have proved that the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ achieve the semiparametric efficiency bound. ♯