Supplementary Materials for: "Analysis of Proportional Odds Models with Censoring and Errors-in-Covariates"

Samiran SINHA and Yanyuan MA

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.

emails: sinha@stat.tamu.edu and ma@stat.tamu.edu

These supplementary materials contain a detailed derivation of $\widehat{\gamma}_2$, regularity conditions, necessary lemmas, and the proof of the theorems and Corollary 1.

S1 Derivation of $\widehat{\gamma}_2$

Note that

$$\begin{aligned}
\gamma_2 &= E\{U_i \exp(\beta_2 U_i)\} \\
&= \frac{\partial}{\partial \beta_2} E\{\exp(\beta_2 U_i)\} \\
&= \frac{\partial}{\partial \beta_2} \Big\{\mathcal{M}(\beta_2/m)\Big\}^m \\
&= m\Big\{\mathcal{M}(\beta_2/m)\Big\}^{m-1} \frac{\partial}{\partial \beta_2}\Big\{\mathcal{M}(\beta_2/m)\Big\}.
\end{aligned}$$

Since a consistent estimator of $\mathcal{M}(\beta_2/m)$ is $(\widehat{\gamma}_1)^{1/m}$, $\widehat{\gamma}_2$ can be consistently estimated by

$$\widehat{\gamma}_2 = m\,(\widehat{\gamma}_1)^{(m-1)/m} \frac{\partial}{\partial \beta_2} (\widehat{\gamma}_1)^{1/m}. \tag{S.1}$$

Note that

$$\widehat{\gamma}_1 = \left[\frac{2}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} \exp\{(W^*_{ij} - W^*_{ik})\beta_2/m\}\right]^{m/2}.$$

So,

$$\begin{aligned}
\frac{\partial}{\partial \beta_2} (\widehat{\gamma}_1)^{1/m} &= \frac{\partial}{\partial \beta_2}\left[\frac{2}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} \exp\{(W^*_{ij} - W^*_{ik})\beta_2/m\}\right]^{1/2} \\
&= \frac{1}{2}\left[\frac{2}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} \exp\{(W^*_{ij} - W^*_{ik})\beta_2/m\}\right]^{(-1/2)}
\end{aligned}$$

$$\times \left[\frac{2}{nm^2(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} (W_{ij}^* - W_{ik}^*) \exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\} \right]$$

$$= (\widehat{\gamma}_1)^{(-1/m)} \left[\frac{1}{nm^2(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} (W_{ij}^* - W_{ik}^*) \exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\} \right].$$

Now plugging in the above expression in (S.1) we get

$$\begin{aligned} \widehat{\gamma}_2 &= m\,(\widehat{\gamma}_1)^{(m-1)/m}\,(\widehat{\gamma}_1)^{(-1/m)} \left[\frac{1}{nm^2(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} (W_{ij}^* - W_{ik}^*) \exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\} \right] \\ &= (\widehat{\gamma}_1)^{(m-2)/m} \left[\frac{1}{nm(m-1)} \sum_{j,k=1,j<k}^{m} \sum_{i=1}^{n} (W_{ij}^* - W_{ik}^*) \exp\{(W_{ij}^* - W_{ik}^*)\beta_2/m\} \right], \end{aligned}$$

and this last expression is given in Equation (8) of the main document.

S2 Regularity conditions

Define a class of functions $\mathcal{F} \equiv \{\Lambda : [0,\infty) \to [0,\infty), \Lambda \text{ is monotonically non-decreasing}, \Lambda(0) = 0\}$, and let $\mathcal{B}$ be a compact subset of of the Euclidean space $\mathcal{R}^{p+1}$. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \beta_2)$ and $\theta = (\boldsymbol{\beta}, \Lambda)$. Thus the parameter space of θ is $\Theta = \mathcal{B} \times \mathcal{F}$. Define a metric d on Θ as

$$d(\theta, \theta^*) = \{(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^{\mathrm{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}^*) + \sup_{t\in[0,\tau]} |\Lambda(t) - \Lambda^*(t)|^2\}^{1/2}.$$

To derive the asymptotic properties of the proposed error corrected estimator, we assume the following regularity conditions to hold.

C1. $f(\Lambda, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ is a continuous function of Λ such that $E(\mathbf{S}_\beta)$ does not vanish except at the true $\boldsymbol{\beta}$ value, where $\mathbf{S}_\beta = (\mathbf{S}_{\beta_1}^T, S_{\beta_2})^T$. In addition, the matrix $E(\partial \mathbf{S}_\beta / \partial \boldsymbol{\beta}^{\mathrm{T}})$ is a continuous function of $\boldsymbol{\beta}$ and at the true $\boldsymbol{\beta}$ value it has eigenvalues bounded away from zero and infinity. The matrix Σ_H defined in (S.3) is nonsingular.

C2. The true $\boldsymbol{\beta}$ lies in the interior of $\mathcal{B}$.

C3. $g_1(W, \mathbf{Z}, \boldsymbol{\beta}), \mathbf{Z}g_1(W, \mathbf{Z}, \boldsymbol{\beta}), Wg_2(W, \mathbf{Z}, \boldsymbol{\beta}), g_2(W, \mathbf{Z}, \boldsymbol{\beta})$ are integrable functions of $(W, \mathbf{Z})$ for all $\boldsymbol{\beta} \in \mathcal{B}$.

C4. The true baseline cumulative hazard and hazard functions $\Lambda(u)$ and $\lambda(u)$ are bounded for $u \in [0, \tau]$.

C5. The estimated $\widehat{\boldsymbol{\alpha}}$ satisfies $\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = O_p(1)$ when $n \to \infty$ for all $\boldsymbol{\alpha}$ in a compact set.

S3 Proof of Theorem 1

We first inspect the situation where an arbitrary fixed $\boldsymbol{\alpha}$ is used in the construction. Define $\boldsymbol{\psi}_{n,x,1} = n^{-1}\mathbf{S}_{\beta_1}$, $\psi_{n,x,2} = n^{-1}S_{\beta_2}$, $\psi_{n,x,3} = n^{-1}S_{\Lambda}$, $\boldsymbol{\psi}_{n,x} = (\boldsymbol{\psi}_{n,x,1}^{\mathrm{T}}, \psi_{n,x,2}, \psi_{n,x,2})^{\mathrm{T}}$, where the subindex $_x$ indicates that these are equations associated with the unobservable covariate X. Define $\boldsymbol{\psi}_{n,1} = n^{-1}\mathbf{S}_{\beta_1}^{\mathrm{me}}$, $\psi_{n,2} = n^{-1}S_{\beta_2}^{\mathrm{me}}$, $\psi_{n,3} = n^{-1}S_{\Lambda}^{\mathrm{me}}$, $\boldsymbol{\psi}_1 = E\{\boldsymbol{\psi}_1^*(\theta,u)\}$, $\psi_2 = E\{\psi_2^*(\theta,u)\}$, and $\psi_3 = E\{\psi_3^*(\theta,u)\}$, where

$$
\begin{aligned}
\boldsymbol{\psi}_1^*(\theta,u) &= \Delta\mathbf{Z}\{1+\Lambda(V)g_1(W,\mathbf{Z},\boldsymbol{\beta})\}f\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} \\
&\quad -\mathbf{Z}g_1(W,\mathbf{Z},\boldsymbol{\beta})\left[F\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} - F(0,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha})\right], \\
\psi_2^*(\theta,u) &= \Delta\{W+\Lambda(V)g_2(W,\mathbf{Z},\boldsymbol{\beta})\}f\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} \\
&\quad -g_2(W,\mathbf{Z},\boldsymbol{\beta})\left[F\{\Lambda(V),\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha}\} - F(0,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\alpha})\right], \\
\psi_3^*(\theta,u) &= \{1+\Lambda(u)g_1(W,\mathbf{Z},\boldsymbol{\beta})\}dN(u) - Y(u)\lambda(u)g_1(W,\mathbf{Z},\boldsymbol{\beta}).
\end{aligned}
$$

Let $\boldsymbol{\psi}_n(\theta,u) = (\boldsymbol{\psi}_{n,1}^{\mathrm{T}}, \psi_{n,2}, \psi_{n,3})^{\mathrm{T}}$, $\boldsymbol{\psi}^*(\theta,u) = (\boldsymbol{\psi}_1^{*\mathrm{T}}, \psi_2^*, \psi_3^*)^{\mathrm{T}}$ and $\boldsymbol{\psi}(\theta,u) = (\boldsymbol{\psi}_1^{\mathrm{T}}, \psi_2, \psi_3)^{\mathrm{T}}$. For every $u \in [0,\tau]$, $E(\boldsymbol{\psi}_n) = \boldsymbol{\psi}$. Obviously $\boldsymbol{\psi}_n : \Theta \mapsto L$ where L is a normed space equipped with the supreme norm $||\cdot||_L$. Following Theorem 2.10 of Kosorok (2008), to prove $d(\widehat{\theta}_n,\theta) \xrightarrow{P} 0$ for $||\psi_n(\widehat{\theta}_n)||_L \xrightarrow{P} 0$ we need to show i) (Identifiability) Let $\psi(\theta) = 0$ for some $\theta \in \Theta$, if for a sequence $\theta_n \in \Theta$, $||\psi(\theta_n)||_L \to 0$ then $d(\theta,\theta_n) \to 0$; and ii) (Uniform convergence) $\sup_{\theta\in\Theta} ||\psi_n(\theta) - \psi(\theta)||_L \xrightarrow{p} 0$.

To show i), we only need to show that $\boldsymbol{\psi}(\theta,u) = \mathbf{0}$ has a unique solution θ. $\boldsymbol{\psi}(\theta,u) = \mathbf{0}$ implies $\mathbf{0} = E[E\{\boldsymbol{\psi}^*(\theta,u) \mid \mathbf{Z},X,V,\Delta\}] = E\{\boldsymbol{\psi}_{n,x}(\theta,u)\}$. Because $\boldsymbol{\psi}_{n,x}(\theta,u) = \mathbf{0}$ leads to a consistent estimator of θ (Chen, Jin and Ying, 2002), hence $E\{\boldsymbol{\psi}_{n,x}(\theta,u)\} = \mathbf{0}$ has a unique root θ in the neighborhood of the true parameter. To show ii) we need to show that the class of functions $\{\psi_1^*(\theta,u), \psi_2^*(\theta,u), \psi_3^*(\theta,u), \theta \in \Theta, u \in [0,\tau]\}$ is Glivenko-Cantelli which requires us to show that $\sup_{u\in[0,\tau]} |\psi_1^*(\theta,u)|$, $\sup_{u\in[0,\tau]} |\psi_2^*(\theta,u)|$, and $\sup_{u\in[0,\tau]} |\psi_3^*(\theta,u)|$ are dominated by integrable functions (Lemma 6.1 of Wellner (2003)).

Under the above regularity conditions $\sup_{u\in[0,\tau]} \psi_1^*(\theta,u)$ and $\sup_{u\in[0,\tau]} \psi_2^*(\theta,u)$ are obviously dominated by integrable functions. For $\psi_3^*(\theta,u)$,

$$
\begin{aligned}
& \sup_{u\in[0,\tau]} |\{1+\Lambda(u)g_1(W,\mathbf{Z},\boldsymbol{\beta})\}dN(u) - Y(u)\lambda(u)g_1(W,\mathbf{Z},\boldsymbol{\beta})| \\
\leq\ & \sup_{u\in[0,\tau]} dN(u) + \sup_{u\in[0,\tau]} dN(u)\Lambda(u)g_1(W,\mathbf{Z},\boldsymbol{\beta}) + g_1(W,\mathbf{Z},\boldsymbol{\beta}) \sup_{u\in[0,\tau]} Y(u)\lambda(u).
\end{aligned}
$$

Under the regularity conditions $\sup_{u\in[0,\tau]} \psi_3^*(\theta,u)$ is also dominated by an integrable function.

Having established the local consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}$ under a fixed $\boldsymbol{\alpha}$, we can now easily extend the results to the situation where $\widehat{\boldsymbol{\alpha}}$ is used. Assume $\widehat{\boldsymbol{\alpha}} \to \boldsymbol{\alpha}$ in probability when $n \to \infty$. Write the estimators under $\boldsymbol{\alpha}$ as $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and $\widehat{\Lambda}(\boldsymbol{\alpha})$, and the ones under an estimated $\widehat{\boldsymbol{\alpha}}$ as $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$ and $\widehat{\Lambda}(\widehat{\boldsymbol{\alpha}})$. Then $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and $\widehat{\Lambda}(\widehat{\boldsymbol{\alpha}}) - \widehat{\Lambda}(\boldsymbol{\alpha})$ go to zero in probability when $n \to \infty$, hence $\widehat{\Lambda}(\boldsymbol{\alpha})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$ are also consistent. □

S4 Necessary Lemmas for Theorem 2

Result 1. (Polyanin and Manzhirov, 2008) If $y(t) = \int_0^t a(u)y(u)du + b(t)$ and $y(0) = 0$ then

$$y(t) = \exp\{\int_0^t a(u)du\} \int_0^t \exp\{-\int_0^s a(u)du\} b'(s)ds.$$

Lemma 2. *The cumulative hazard function estimator has the martingale representation*

$$\sqrt{n}[\widehat{\Lambda}\{t, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(t)] = \frac{\gamma_1(\boldsymbol{\beta})}{\sqrt{n} D_1(t)} \sum_{i=1}^n \int_0^t \frac{D_1(s)}{C_1(s)} \left\{1 + \Lambda(s) \frac{\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\} dM_i(s) + o_p(1)$$

for all $\boldsymbol{\beta}$ in the interior of $\mathcal{B}$.

Proof: From the estimating equation $S_\Lambda^{\text{me}} = 0$ we can write

$$\begin{aligned}
\widehat{\Lambda}\{t, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} &= \int_0^t \frac{\sum_{i=1}^n [1 + \widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} \eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})/\gamma_1(\boldsymbol{\beta})] dN_i(s)}{\sum_{i=1}^n Y_i(s) \eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})/\gamma_1(\boldsymbol{\beta})} \\
&= \int_0^t \frac{\sum_{i=1}^n \{1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})/\gamma_1(\boldsymbol{\beta})\} dN_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})/\gamma_1(\boldsymbol{\beta})} \\
&\quad + \int_0^t \frac{\sum_{i=1}^n \eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})} [\widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(s)] dN_i(s).
\end{aligned}$$

Using $dN_i(s) = Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})ds/\{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} + dM_i(s)$ and using the strong law of large numbers, we obtain

$$\begin{aligned}
\widehat{\Lambda}\{t, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} &= \int_0^t \lambda(s)ds + \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})/\gamma_1(\boldsymbol{\beta})}{C_1(s)/\gamma_1(\boldsymbol{\beta})} dM_i(s) \\
&\quad + \int_0^t \frac{C_2(s)}{C_1(s)} [\widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(s)] ds \\
&\quad + \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}{C_1(s)} [\widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(s)] dM_i(s). \qquad \text{(S.2)}
\end{aligned}$$

The fourth term on the right hand side of (S.2) is of order $o_p[\int_0^t |\widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(s)|ds]$, hence it is negligible. Therefore,

$$\widehat{\Lambda}\{t, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})/\gamma_1(\boldsymbol{\beta})}{C_1(s)/\gamma_1(\boldsymbol{\beta})} dM_i(s)$$
$$+ \int_0^t \frac{C_2(s)}{C_1(s)}[\widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(s)]ds + o_p[\int_0^t |\widehat{\Lambda}\{s, \boldsymbol{\beta}, \gamma_1(\boldsymbol{\beta})\} - \Lambda(s)|ds].$$

To solve the above integral equation in the leading order we use Result 1, and we get the desired result. □

<u>Lemma</u> 3. *For large n, $\widehat{\gamma}(\boldsymbol{\beta})$ satisfies*

$$\sqrt{n}\{\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}) - \boldsymbol{\gamma}(\boldsymbol{\beta})\} = n^{-1/2}\sum_{i=1}^{n}\mathbf{f}_\gamma(\mathbf{W}_i^*, \boldsymbol{\beta}) + o_p(1)$$

for all $\boldsymbol{\beta}$ in the interior of $\mathcal{B}$.

Proof: By definition

$$\left\{\mathcal{M}\left(\frac{\beta_2}{m}\right)\right\}^{m-2} = \left[\frac{2}{m(m-1)}\int\sum_{j,k=1,j<k}^{m}\exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\}dP\right]^{m/2-1}.$$

Let P_n be the empirical distribution of based on $W_i^* = (W_{i1}^*, \cdots, W_{im}^*)$ and δ_i be the Dirac measure at the i^{th} observation. Let P be the population version of P_n. Then $\widehat{\gamma}_1(\boldsymbol{\beta})$ can be written as

$$\widehat{\gamma}_1(\boldsymbol{\beta}) = \Phi(P_n) = \left[\frac{2}{m(m-1)}\int\sum_{j,k=1,j<k}^{m}\exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\}dP_n\right]^{m/2}.$$

Now using von Mises expansion (van der Vaart, 1998; p. 292), we can write

$$\begin{aligned}
&\sqrt{n}\{\widehat{\gamma}_1(\boldsymbol{\beta}) - \gamma_1(\boldsymbol{\beta})\} \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial}{\partial t}\Phi\{(1-t)P + t\delta_i\}\right]_{t=0} + o_p(1) \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathcal{M}^{(m-2)}(\beta_2/m)}{m-1}\sum_{j,k=1,j<k}^{m}\left[\exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\} - E\exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\}\right] + o_p(1) \\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathcal{M}^{(m-2)}(\beta_2/m)}{m-1}\sum_{j,k=1,j<k}^{m}\left[\exp\left\{(W_{ij}^* - W_{ik}^*)\frac{\beta_2}{m}\right\} - \mathcal{M}^2\left(\frac{\beta_2}{m}\right)\right] + o_p(1).
\end{aligned}$$

Now consider $\widehat{\gamma}_2(\boldsymbol{\beta})$, and we can write

$$\begin{aligned}\gamma_2(\boldsymbol{\beta}) &= \left[\frac{2}{m(m-1)}\int \sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP\right]^{m/2-1} \\ &\times \left[\frac{1}{m(m-1)}\int \sum_{j,k=1,j<k}^{m} (W_j^* - W_k^*) \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP\right].\end{aligned}$$

Using the von Mises expansion we write

$$\begin{aligned}&\sqrt{n}\{\widehat{\gamma}_2(\boldsymbol{\beta}) - \gamma_2(\boldsymbol{\beta})\} \\ =\ & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m}{2}-1\right)\left[\frac{2}{m(m-1)}\int \sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP\right]^{m/2-2} \\ & \frac{2}{m(m-1)}\left[-\int \sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP + \sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\}\right] \\ & \times \left[\frac{1}{m(m-1)}\int \sum_{j,k=1,j<k}^{m} (W_j^* - W_k^*) \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP\right] \\ & + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{2}{m(m-1)}\int \sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP\right]^{m/2-1} \\ & \times \left[-\frac{1}{m(m-1)}\int \sum_{j,k=1,j<k}^{m} (W_j^* - W_k^*) \exp\left\{\frac{(W_j^* - W_k^*)\beta_2}{m}\right\} dP\right. \\ & \left. + \frac{1}{m(m-1)} \sum_{j,k=1,j<k}^{m} (W_{ij}^* - W_{ik}^*) \exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\}\right] + o_p(1) \\ =\ & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m}{2}-1\right)\mathcal{M}^{(m-4)}\left(\frac{\beta_2}{m}\right)\left[\frac{2}{m(m-1)}\sum_{j,k=1,j<k}^{m} \exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\} - \mathcal{M}^2\left(\frac{\beta_2}{m}\right)\right] \\ & \times \frac{m}{2}\frac{\partial \mathcal{M}^2(\beta_2/m)}{\partial \beta_2} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{M}^{(m-2)}\left(\frac{\beta_2}{m}\right) \\ & \left[\frac{1}{m(m-1)}\sum_{j,k=1,j<k}^{m} (W_{ij}^* - W_{ik}^*) \exp\left\{\frac{(W_{ij}^* - W_{ik}^*)\beta_2}{m}\right\} - \frac{m}{2}\frac{\partial \mathcal{M}^2(\beta_2/m)}{\partial \beta_2}\right] + o_p(1).\end{aligned}$$

□

Lemma 4. *At any $t \in (0, \tau]$,*

$$i)\ \widehat{\Lambda}'_{\gamma_1}(t, \boldsymbol{\beta}, \gamma_1) = D_2(t) + o_p(1), \quad ii)\ \widehat{\Lambda}'_{\beta}(t, \boldsymbol{\beta}, \gamma_1) = \mathbf{D}_3^{\mathrm{T}}(t) + o_p(1).$$

Proof of part i): Since at any $\boldsymbol{\beta}, \boldsymbol{\gamma}$, $\widehat{\Lambda}(t, \boldsymbol{\beta}, \boldsymbol{\gamma})$ satisfies $S_{\Lambda}^{\mathrm{me}}(t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = 0$, we have

$$\widehat{\Lambda}(t, \boldsymbol{\beta}, \gamma_1) = \int_0^t \frac{\sum_{i=1}^n \{\gamma_1 + \widehat{\Lambda}(s, \boldsymbol{\beta}, \gamma_1)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}.$$

Taking partial derivative with respect to γ_1 on both sides, we have

$$\begin{aligned}
&\widehat{\Lambda}'_{\gamma_1}(t, \boldsymbol{\beta}, \gamma_1) \\
=& \int_0^t \frac{\sum_{i=1}^n \{1 + \widehat{\Lambda}'_{\gamma_1}(s, \boldsymbol{\beta}, \gamma_1)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})} \\
=& \int_0^t \frac{\sum_{i=1}^n \{1 + \widehat{\Lambda}'_{\gamma_1}(s, \boldsymbol{\beta}, \gamma_1)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dM_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})} \\
&+ \int_0^t \frac{\sum_{i=1}^n \{1 + \widehat{\Lambda}'_{\gamma_1}(s, \boldsymbol{\beta}, \gamma_1)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})/\{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\} ds}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})} \\
=& \int_0^t \frac{C_2(s)}{C_1(s)} \widehat{\Lambda}'_{\gamma_1}(s, \boldsymbol{\beta}, \gamma_1) ds + \int_0^t \frac{C_3(s)}{C_1(s)} ds + o_p(1).
\end{aligned}$$

To solve the above integral equation in the leading order we use Result 1. Thus the solution of the integral equation is

$$\begin{aligned}
\widehat{\Lambda}'_{\gamma_1}(t, \boldsymbol{\beta}, \gamma_1) &= \exp\{\int_0^t \frac{C_2(u)}{C_1(u)} du\} \int_0^t \exp\{-\int_0^s \frac{C_2(u)}{C_1(u)} du\} \frac{C_3(s)}{C_1(s)} ds + o_p(1) \\
&= \frac{1}{D_1(t, \boldsymbol{\beta}, \Lambda)} \int_0^t \frac{D_1(s)C_3(s)}{C_1(s)} ds + o_p(1).
\end{aligned}$$

Proof of part ii): Since at any $\boldsymbol{\beta}, \boldsymbol{\gamma}$, $\widehat{\Lambda}(t, \boldsymbol{\beta}, \boldsymbol{\gamma})$ satisfies $S_{\Lambda}^{\mathrm{me}}(t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = 0$, we have

$$\widehat{\Lambda}(t, \boldsymbol{\beta}, \gamma_1) = \int_0^t \frac{\sum_{i=1}^n \{\gamma_1 + \widehat{\Lambda}(s, \boldsymbol{\beta}, \gamma_1)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} dN_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}.$$

Taking partial derivative with respect to $\boldsymbol{\beta}$ on both sides, we have

$$\begin{aligned}
&\widehat{\Lambda}'_{\beta}(t, \boldsymbol{\beta}, \gamma_1) \\
=& \int_0^t \frac{\sum_{i=1}^n \{\widehat{\Lambda}'_{\beta}(s, \boldsymbol{\beta}, \gamma_1) + \widehat{\Lambda}(s, \boldsymbol{\beta}, \gamma_1)(\mathbf{Z}_i^{\mathrm{T}}, W_i)^{\mathrm{T}}\}\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) dN_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})} \\
&- \int_0^t \frac{\sum_{i=1}^n \{\gamma_1 + \widehat{\Lambda}(s, \boldsymbol{\beta}, \gamma_1)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\}\{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})(\mathbf{Z}_i^{\mathrm{T}}, W_i)^{\mathrm{T}}\} dN_i(s)}{\{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\}^2} \\
=& \int_0^t \frac{\sum_{i=1}^n \{\widehat{\Lambda}'_{\beta}(s, \boldsymbol{\beta}, \gamma_1) + \widehat{\Lambda}(s, \boldsymbol{\beta}, \gamma_1)(\mathbf{Z}_i^{\mathrm{T}}, W_i)^{\mathrm{T}}\}\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) dM_i(s)}{\sum_{i=1}^n Y_i(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})}
\end{aligned}$$

$$
\begin{aligned}
& -\int_0^t \frac{\sum_{i=1}^n \{\gamma_1 + \widehat{\Lambda}(s,\boldsymbol{\beta},\gamma_1)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}\mathbf{C}_4(s)dM_i(s)}{\{\sum_{i=1}^n Y_i(s)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}C_1(s)} \\
& +\int_0^t \frac{\widehat{\Lambda}'_\beta(s,\boldsymbol{\beta},\gamma_1)C_2(s) + \widehat{\Lambda}(s,\boldsymbol{\beta},\gamma_1)\mathbf{C}_5(s)}{C_1(s)}ds - \int_0^t \frac{\{\gamma_1 C_3(s) + \widehat{\Lambda}(s,\boldsymbol{\beta},\gamma_1)C_2(s)\}\mathbf{C}_4(s)}{C_1^2(s)}ds \\
& +o_p(1) \\
= & \int_0^t \frac{C_2(s)}{C_1(s)}\widehat{\Lambda}'_\beta(s,\boldsymbol{\beta},\gamma_1)ds + \int_0^t \frac{\widehat{\Lambda}(s,\boldsymbol{\beta},\gamma_1)\{C_1(s)\mathbf{C}_5(s) - C_2(s)\mathbf{C}_4(s)\} - \gamma_1 C_3(s)\mathbf{C}_4(s)}{C_1^2(s)}ds \\
& +o_p(1).
\end{aligned}
$$

To solve the above integral equation in the leading order we use Result 1. Thus we obtain

$$
\widehat{\Lambda}'_\beta(t,\boldsymbol{\beta},\gamma_1) = \frac{1}{D_1(t)}\int_0^t D_1(s)\frac{\Lambda(s,\boldsymbol{\beta},\gamma_1)\{C_1(s)\mathbf{C}_5(s) - C_2(s)\mathbf{C}_4(s)\} - \gamma_1 C_3(s)\mathbf{C}_4(s)}{C_1^2(s)}ds + o_p(1).
$$

□

S5 Proof of Theorem 2

Proof of part i): We first prove the results under a fixed $\boldsymbol{\alpha}$. Later we show that even when $\boldsymbol{\alpha}$ is replaced by $\widehat{\boldsymbol{\alpha}}$ the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ remained unchanged. From the estimation procedure, we know that $\widehat{\boldsymbol{\beta}}$ satisfies

$$
\mathbf{0} = n^{-1/2}\sum_{i=1}^n \boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\},\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\beta}}),\boldsymbol{\alpha}] = \sum_{k=1}^8 \mathbf{A}_k,
$$

where

$$
\begin{aligned}
\mathbf{A}_1 &= n^{-1/2}\sum_{i=1}^n \boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\}, \\
\mathbf{A}_2 &= n^{-1/2}\sum_{i=1}^n \left(\boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\},\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}] - \boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\}\right), \\
\mathbf{A}_3 &= n^{-1/2}\sum_{i=1}^n \left(\boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}] - \boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\},\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}]\right), \\
\mathbf{A}_4 &= n^{-1/2}\sum_{i=1}^n \left(\boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}] - \boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}]\right), \\
\mathbf{A}_5 &= n^{-1/2}\sum_{i=1}^n \left(\boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}] - \boldsymbol{\phi}[\mathbf{O}_i;\boldsymbol{\beta},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}]\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A}_6 &= n^{-1/2}\sum_{i=1}^{n}\left(\boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}] - \boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}]\right),\\
\mathbf{A}_7 &= n^{-1/2}\sum_{i=1}^{n}\left(\boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}] - \boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\boldsymbol{\beta})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}]\right),\\
\mathbf{A}_8 &= n^{-1/2}\sum_{i=1}^{n}\left(\boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\},\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\beta}}),\boldsymbol{\alpha}] - \boldsymbol{\phi}[\mathbf{O}_i;\widehat{\boldsymbol{\beta}},\widehat{\Lambda}\{V_i;\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\},\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}),\boldsymbol{\alpha}]\right).
\end{aligned}
$$

Following Lemma 3, and using the definitions of $\boldsymbol{\phi}_\gamma$, $\boldsymbol{\phi}_\beta$ and $\boldsymbol{\gamma}_\beta$, we have

$$
\begin{aligned}
\mathbf{A}_3 &= \left\{\boldsymbol{\phi}_\gamma + o_p(1)\right\}\sqrt{n}\{\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}) - \boldsymbol{\gamma}(\boldsymbol{\beta})\} = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta}) + o_p(1),\\
\mathbf{A}_5 &= \boldsymbol{\phi}_\beta\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1),\\
\mathbf{A}_8 &= \boldsymbol{\phi}_\gamma\boldsymbol{\gamma}_\beta\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1).
\end{aligned}
$$

Using Lemma 4, we have

$$
\begin{aligned}
\mathbf{A}_4 &= E\{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\Lambda'_{\gamma_1}(V_i,\boldsymbol{\beta},\gamma_1)\}\sqrt{n}(\widehat{\gamma}_1 - \gamma_1) + o_p(1)\\
&= n^{-1/2}\sum_{i=1}^{n}E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)D_2(V_i)\right\}f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta}) + o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{A}_7 &= E[\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\{\Lambda'_{\gamma_1}(V_i,\boldsymbol{\beta},\gamma_1),0\}]\boldsymbol{\gamma}_\beta\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1)\\
&= \left[\mathbf{0}_{(p+1)\times p},\gamma_2 E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)D_2(V_i)\right\}\right]\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1),
\end{aligned}
$$

Similarly, using Lemma 4, we have

$$
\mathbf{A}_6 = E\{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\Lambda'_\beta(V_i,\boldsymbol{\beta},\gamma_1)\}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1) = E\{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\mathbf{D}_3^{\mathrm{T}}(V_i)\}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1).
$$

Using Lemma 2, we have

$$
\begin{aligned}
\mathbf{A}_2 &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)[\widehat{\Lambda}\{V_i;\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\} - \Lambda(V_i)] + o_p(1)\\
&= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\gamma_1(\boldsymbol{\beta})}{D_1(V_i)}\int_0^{V_i}\frac{D_1(s)}{C_1(s)}\left\{1 + \Lambda(s)\frac{\eta(W_j,\mathbf{Z}_j,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_j(s) + o_p(1)\\
&= n^{-1/2}\sum_{i=1}^{n}\frac{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\gamma_1(\boldsymbol{\beta})}{D_1(V_i)}E\left[\int_0^{V_i}\frac{D_1(s)}{C_1(s)}\left\{1 + \Lambda(s)\frac{\eta(W_j,\mathbf{Z}_j,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_j(s)\mid \mathbf{O}_i\right]
\end{aligned}
$$

$$
\begin{aligned}
& +n^{-1/2}\sum_{j=1}^{n} E\left[\frac{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\gamma_1(\boldsymbol{\beta})}{D_1(V_i)}\int_0^{V_i}\frac{D_1(s)}{C_1(s)}\left\{1+\Lambda(s)\frac{\eta(W_j,\mathbf{Z}_j,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_j(s)\mid \mathbf{O}_j\right] \\
& -n^{-1/2}E\left[\frac{\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\gamma_1(\boldsymbol{\beta})}{D_1(V_i)}\int_0^{V_i}\frac{D_1(s)}{C_1(s)}\left\{1+\Lambda(s)\frac{\eta(W_j,\mathbf{Z}_j,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_j(s)\right]+o_p(1) \\
= \;& n^{-1/2}\sum_{j=1}^{n}\gamma_1(\boldsymbol{\beta})\int_0^\infty E\left\{\frac{Y_i(s)\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)}{D_1(V_i)}\right\}\frac{D_1(s)}{C_1(s)}\left\{1+\Lambda(s)\frac{\eta(W_j,\mathbf{Z}_j,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_j(s)+o_p(1) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\gamma_1(\boldsymbol{\beta})\int_0^\infty \mathbf{D}_4(s)\left\{1+\Lambda(s)\frac{\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_i(s)+o_p(1) \\
= \;& n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \mathbf{g}(s,W_i,\mathbf{Z}_i)dM_i(s)+o_p(1),
\end{aligned}
$$

where we used the U-statistic property to obtain the above third equality.

Combining the above results, we have

$$
\begin{aligned}
\mathbf{0} \;=\; & n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\}+n^{-1/2}\sum_{i=1}^{n}\int_0^\infty \mathbf{g}(s,W_i,\mathbf{Z}_i)]dM_i(s) \\
& +n^{-1/2}\sum_{i=1}^{n}\left[\boldsymbol{\phi}_\gamma\mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta})+E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\}f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\right] \\
& +\Sigma_H\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+o_p(1),
\end{aligned}
$$

where

$$
\Sigma_H=\boldsymbol{\phi}_\beta+\boldsymbol{\phi}_\gamma\boldsymbol{\gamma}_\beta+E\left[\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)\mathbf{D}_3(V_i)+\left\{\mathbf{0}_{(p+1)\times p},\gamma_2\boldsymbol{\phi}_\Lambda(\mathbf{O}_i)D_2(V_i)\right\}\right]. \tag{S.3}
$$

Hence

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \;=\; & -\frac{\Sigma_H^{-1}}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\}+\int_0^\infty \mathbf{g}(s,W_i,\mathbf{Z}_i)dM_i(s) \\
& +\boldsymbol{\phi}_\gamma\mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta})+E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\}f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\Bigg]+o_p(1).
\end{aligned}
$$

The first term of the summand can be written as

$$
\begin{aligned}
& \boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\} \\
& =\int_0^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1+\Lambda(s)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\} \\ \{W_i\gamma_1^2+\Lambda(s)(\gamma_1 W_i-\gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\end{bmatrix} dM_i(s)
\end{aligned} \tag{S.4}
$$

$$+\mathbf{h}(U_i, W_i, \mathbf{Z}_i)\int_0^\infty \frac{f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}{1+\Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}ds, \tag{S.5}$$

where $\mathbf{h}(U_i, W_i, \mathbf{Z}_i) = [\mathbf{Z}_i^{\mathrm{T}}\{\gamma_1 - \exp(\beta_2 U_i)\}, \{W_i\gamma_1^2 - \gamma_1 W_i \exp(\beta_2 U_i) + \gamma_2 \exp(\beta_2 U_i)\}]^{\mathrm{T}}$. Expression given in (S.4) has mean zero as it is a stochastic integral with respect to a martingale where the integrand is a predictable process. The expression given in (S.5) has mean zero as $E\{\mathbf{h}(U_i, W_i, \mathbf{Z}_i)\} = 0$ and $\int_0^\infty f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\{1+\Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})\}^{-1}ds$ and $\mathbf{h}$ are conditionally independent. The mean of the second term of the summand is zero as it is also a stochastic integral with respect to a martingale where the integrand is a predictable process. The mean of the third and fourth terms of the summand are zero as $E\{\mathbf{f}_\gamma(\mathbf{W}_i^*, \boldsymbol{\beta})\} = 0$, $E\{f_{\gamma,1}(\mathbf{W}_i^*, \boldsymbol{\beta})\} = 0$. Therefore, using the central limit theorem we obtain asymptotic normality of the estimator. Consequently $n\mathrm{var}(\widehat{\boldsymbol{\beta}}) \to \Sigma_H^{-1}\Sigma_M\Sigma_H^{-T}$, where

$$\begin{aligned}\Sigma_M =& E\Bigg[\boldsymbol{\phi}\{\mathbf{O}; \boldsymbol{\beta}, \Lambda(V_i), \boldsymbol{\gamma}, \boldsymbol{\alpha}\} + \int_0^\tau \mathbf{g}(s, W, \mathbf{Z})dM(s) \\ &+ E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\} f_{\gamma,1}(\mathbf{W}^*, \boldsymbol{\beta}) + \boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}^*, \boldsymbol{\beta})\Bigg]^{\otimes 2}.\end{aligned}$$

We now consider the estimation under $\widehat{\boldsymbol{\alpha}}$. We have $n\mathrm{var}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})\} = E[n\mathrm{var}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}) \mid \widehat{\boldsymbol{\alpha}}\}] + n\mathrm{var}[E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}) \mid \widehat{\boldsymbol{\alpha}}\}]$. Thus, $n\mathrm{var}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})\} = \Sigma_H^{-1}\Sigma_M\Sigma_H^{-T} + E\{\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^{\mathrm{T}}\}n\mathrm{var}(\widehat{\boldsymbol{\alpha}})E\{\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})^{\mathrm{T}}/\partial\boldsymbol{\alpha}\} + o(1)$, where $E\{\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^{\mathrm{T}}\} = n^{-1/2}\Sigma_H^{-1}\Sigma_\alpha$ with

$$\begin{aligned}\Sigma_\alpha \;=\; & E\Bigg[\frac{\partial\boldsymbol{\phi}\{\mathbf{O}_i; \boldsymbol{\beta}, \Lambda(V_i), \boldsymbol{\gamma}(\boldsymbol{\beta}), \boldsymbol{\alpha}\}}{\partial\boldsymbol{\alpha}^{\mathrm{T}}} + \int_0^\tau \frac{\partial\mathbf{g}(s, t, W_i, \mathbf{Z}_i)}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}dM_i(s) \\ &+ \frac{\partial\boldsymbol{\phi}_\gamma}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}\mathbf{f}_\gamma(\mathbf{W}_i^*, \boldsymbol{\beta}) + E\left\{\frac{\partial\boldsymbol{\varphi}_\Lambda(\mathbf{O})}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}D_2(V)\right\} f_{\gamma,1}(\mathbf{W}_i^*, \boldsymbol{\beta})\Bigg] = 0.\end{aligned}$$

The first term of Σ_α is

$$\begin{aligned}& E\left(\frac{\partial}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}\boldsymbol{\phi}\{\mathbf{O}_i; \boldsymbol{\beta}, \Lambda(V_i), \boldsymbol{\gamma}(\boldsymbol{\beta}), \boldsymbol{\alpha}\}\right) \\ =& E\Bigg(\int_0^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} \\ \{W_i\gamma_1^2 + \Lambda(s)(\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\}\end{bmatrix} \frac{f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}E\{dM_i(s)|\mathcal{F}(s)\} \\ &+ E\{\mathbf{h}(U_i, W_i, \mathbf{Z}_i)|X_i, Y_i, \mathbf{Z}_i\}\int_0^\infty \frac{[\partial f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}/\partial\boldsymbol{\alpha}^{\mathrm{T}}]Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}{1+\Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}ds\Bigg) \\ =& \mathbf{0}.\end{aligned}$$

The second term of Σ_α is zero due to again the martingale property with the integrand $\partial\mathbf{g}(s,t,W_i,\mathbf{Z}_i)/\partial\boldsymbol{\alpha}^{\mathrm{T}}$ that is a predictable process. The third and fourth terms are zero due to $E\{f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\} = E\{f_{\gamma,2}(\mathbf{W}_i^*,\boldsymbol{\beta})\} = 0$. Hence, $n\mathrm{var}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})\} = \Sigma_H^{-1}\Sigma_M\Sigma_H^{-T} + o_p(1)$. □

Proof of part ii): We first prove the result for fixed $\boldsymbol{\alpha}$. We can write

$$\sqrt{n}[\widehat{\Lambda}\{t,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \Lambda(t)] = \sum_{k=1}^{4} B_k,$$

where $B_1 = \sqrt{n}[\widehat{\Lambda}\{t,\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\} - \Lambda(t)]$, $B_2 = \sqrt{n}[\widehat{\Lambda}\{t,\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\} - \widehat{\Lambda}\{t,\boldsymbol{\beta},\gamma_1(\boldsymbol{\beta})\}]$, $B_3 = \sqrt{n}[\widehat{\Lambda}\{t,\boldsymbol{\beta},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \widehat{\Lambda}\{t,\boldsymbol{\beta},\widehat{\gamma}_1(\boldsymbol{\beta})\}]$, $B_4 = \sqrt{n}[\widehat{\Lambda}\{t,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \widehat{\Lambda}\{t,\boldsymbol{\beta},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\}]$. Observe that using Lemmas 3 and 4 we can write $B_2 = D_2(t)n^{-1/2}\sum_{i=1}^n f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta}) + o_p(1)$, $B_3 = D_2(t)(\mathbf{0}^{\mathrm{T}},\gamma_2)\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1)$, and $B_4 = \mathbf{D}_3^{\mathrm{T}}(t)\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1)$. Adding all four terms and using the expression for $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ from the proof of the previous theorem we can write

$$\begin{aligned}
&\sqrt{n}[\widehat{\Lambda}\{t,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \Lambda(t)]\\
=&\frac{\gamma_1(\boldsymbol{\beta})}{\sqrt{n}D_1(t)}\sum_{i=1}^n\int_0^t \frac{D_1(s)}{C_1(s)}\left\{1+\Lambda(s)\frac{\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}dM_i(s) + \frac{D_2(t)}{\sqrt{n}}\sum_{i=1}^n f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\\
&-\left\{D_2(t)(\mathbf{0}^{\mathrm{T}},\gamma_2) + \mathbf{D}_3^{\mathrm{T}}(t)\right\}\Sigma_H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\bigg[\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\} + \int_0^\infty \mathbf{g}(s,W_i,\mathbf{Z}_i)dM_i(s)\\
&+\boldsymbol{\phi}_\gamma\mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta}) + E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\}f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\bigg] + o_p(1)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\int_0^\infty \psi_1(s,t,W_i,\mathbf{Z}_i)dM_i(s) + \psi_2(t,X_i,U_i,W_i^*,\mathbf{Z}_i,Y_i)\right\} + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\psi_1(s,t,W_i,\mathbf{Z}_i) =& \frac{\gamma_1(\boldsymbol{\beta})}{D_1(t)}I(0\le s\le t)\frac{D_1(s)}{C_1(s)}\left\{1+\Lambda(s)\frac{\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})}{\gamma_1(\boldsymbol{\beta})}\right\}\\
&-\left\{D_2(t)(\mathbf{0}^{\mathrm{T}},\gamma_2) + \mathbf{D}_3^{\mathrm{T}}(t)\right\}\Sigma_H^{-1}\mathbf{g}(s,W_i,\mathbf{Z}_i),\\
\psi_2(t,X_i,U_i,W_i^*,\mathbf{Z}_i,Y_i) =& D_2(t)f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta}) - \left\{D_2(t)(\mathbf{0},\gamma_2) + \mathbf{D}_3^{\mathrm{T}}(t)\right\}\Sigma_H^{-1}\left[\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma},\boldsymbol{\alpha}\}\right.\\
&\left.+\boldsymbol{\phi}_\gamma\mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta}) + E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\}f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\right].
\end{aligned}$$

Therefore, for any $0 < t \le t' < \tau$, the covariance kernel of this process is

$$\Omega(t,t') \;=\; n\mathrm{cov}\Big([\widehat{\Lambda}\{t,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \Lambda(t)],[\widehat{\Lambda}\{t',\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\} - \Lambda(t')]\Big)$$

$$= E\left\{\int_0^\tau \psi_1(s,t,W,\mathbf{Z})dM(s) + \psi_2(t,X,U,W^*,\mathbf{Z},Y)\right\}^{\otimes 2}.$$

Now consider the case where $\boldsymbol{\alpha}$ is replaced by $\widehat{\boldsymbol{\alpha}}$. To emphasize the dependence of $\widehat{\boldsymbol{\beta}}$ on $\widehat{\boldsymbol{\alpha}}$, we use $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$ to denote the estimator and use $\widehat{\boldsymbol{\beta}}$ to denote $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$. Writing $B_5 = \sqrt{n}(\widehat{\Lambda}[t,\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}),\widehat{\gamma}_1\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})\}] - \widehat{\Lambda}\{t,\widehat{\boldsymbol{\beta}},\widehat{\gamma}_1(\widehat{\boldsymbol{\beta}})\})$, and using the similar derivation as in part i) of Theorem 2, we show that $B_5 = o_p(1)$. Therefore, the covariance kernel of this process remained unchanged even if we replace $\widehat{\boldsymbol{\beta}}$ by $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$. □

S6 Derivation of an estimator of Σ_M

Based on the derivation of Appendix S5 we can write $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\Sigma_H^{-1}(\mathbf{A}_1 + \mathbf{A}_2^* + \mathbf{A}_3^*) + o_p(1)$, where

$$\begin{aligned}
\mathbf{A}_1 &= \frac{1}{n^{1/2}}\sum_{i=1}^n \boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\} \\
&= \frac{1}{n^{1/2}}\sum_{i=1}^n \int_0^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(s)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}dN_i(s) \\ -Y_i(s)\mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)ds \\ \\ \{W_i\gamma_1^2 + \Lambda(s)(\gamma_1 W_i - \gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})\}f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}dN_i(s) \\ -Y_i(s)(\gamma_1 W_i - \gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta})f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)ds \end{bmatrix}, \\
\mathbf{A}_2^* &= \frac{1}{n^{1/2}}\sum_{i=1}^n \int_0^\infty \mathbf{g}(s,W_i,\mathbf{Z}_i)dM_i(s), \\
\mathbf{A}_3^* &= \frac{1}{n^{1/2}}\sum_{i=1}^n \left[\boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}_i^*,\boldsymbol{\beta}) + E\left\{\boldsymbol{\phi}_\Lambda(\mathbf{O})D_2(V)\right\} f_{\gamma,1}(\mathbf{W}_i^*,\boldsymbol{\beta})\right],
\end{aligned}$$

and each term $\mathbf{A}_1$, $\mathbf{A}_2^*$ and $\mathbf{A}_3^*$ has mean zero. Thus, for calculating var$(\widehat{\boldsymbol{\beta}})$ we need to calculate var$(\mathbf{A}_1 + \mathbf{A}_2^* + \mathbf{A}_3^*) = \text{var}(\mathbf{A}_1) + \text{var}(\mathbf{A}_2^*) + \text{var}(\mathbf{A}_3^*) + \{\text{cov}(\mathbf{A}_1,\mathbf{A}_2^*) + \text{cov}(\mathbf{A}_1,\mathbf{A}_3^*)\} + \{\text{cov}(\mathbf{A}_1,\mathbf{A}_2^*) + \text{cov}(\mathbf{A}_1,\mathbf{A}_3^*)\}^T$. Note that cov$(\mathbf{A}_2^*,\mathbf{A}_3^*) = \mathbf{0}$. Now

$$\mathbf{G}^{(1)} = \text{var}(\mathbf{A}_1) = E[\boldsymbol{\phi}^{\otimes 2}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\}].$$

Next consider

$$\mathbf{G}^{(2)} = \text{var}(\mathbf{A}_2^*) = E\int_0^\infty \mathbf{g}^{\otimes 2}(s,W,\mathbf{Z})\frac{Y(s)\eta(W,\mathbf{Z},\boldsymbol{\beta})\lambda(s)ds}{1+\Lambda(s)\eta(W,\mathbf{Z},\boldsymbol{\beta})}.$$

Since $\mathbf{g}$ is a predictable function $E\int_0^\infty \mathbf{g}^{\otimes 2}(s,W,\mathbf{Z})dM(s) = 0$, so we can write

$$\mathbf{G}^{(2)} = \text{var}(\mathbf{A}_2^*) = E\int_0^\infty \mathbf{g}^{\otimes 2}(s,W,\mathbf{Z})dN(s).$$

Next

$$\mathbf{G}^{(3)} = \mathrm{var}(\mathbf{A}_3^*) = E\left(\left[\boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}_i^*, \boldsymbol{\beta}) + E\left\{ \boldsymbol{\phi}_\Lambda(\mathbf{O}) D_2(V) \right\} f_{\gamma,1}(\mathbf{W}_i^*, \boldsymbol{\beta}) \right]^{\otimes 2} \right)$$

and

$$\begin{aligned} \mathbf{G}^{(4)} &= \mathrm{cov}(\mathbf{A}_1, \mathbf{A}_3^*) \\ &= E\left(\boldsymbol{\phi}\{\mathbf{O}_i; \boldsymbol{\beta}, \Lambda(V_i), \boldsymbol{\gamma}(\boldsymbol{\beta}), \boldsymbol{\alpha}\} \left[\boldsymbol{\phi}_\gamma \mathbf{f}_\gamma(\mathbf{W}_i^*, \boldsymbol{\beta}) + E\left\{ \boldsymbol{\phi}_\Lambda(\mathbf{O}) D_2(V) \right\} f_{\gamma,1}(\mathbf{W}_i^*, \boldsymbol{\beta}) \right]^{\mathrm{T}} \right). \end{aligned}$$

Note that all the above described terms are expectations with respect to observable variables, not involving X or U. Therefore, they are consistently estimated by the respective empirical averages. Finally, we consider $\mathrm{cov}(\mathbf{A}_1, \mathbf{A}_2^*)$:

$$\begin{aligned} \mathbf{G}^{(5)} &= \mathrm{cov}(\mathbf{A}_1, \mathbf{A}_2^*) \\ &= E\left(\Delta_i \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(V_i)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\ -\mathbf{Z}_i \eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) \int_0^\infty Y_i(s) f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}\lambda(s) ds \\ \\ \{W_i\gamma_1^2 + \Lambda(V_i)(\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(V_i), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\ -(\gamma_1 W_i - \gamma_1^2)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) \int_0^\infty Y_i(s) f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}\lambda(s) ds \end{bmatrix} \mathbf{g}^T(V_i, W_i, \mathbf{Z}_i) \right) \\ &\quad - E\left(\int_0^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\ \\ \{W_i\gamma_1^2 + \Lambda(s)(\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \end{bmatrix} dN_i(s) \right. \\ &\quad \left. \times \int_0^\infty \mathbf{g}^T(s, W_i, \mathbf{Z}_i) \frac{Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})} ds \right) \\ &\quad + E\left(\begin{bmatrix} \mathbf{Z}_i \eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) \int_0^\infty Y_i(s) f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}\lambda(s) ds \\ \\ (\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta}) \int_0^\infty Y_i(s) f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\}\lambda(s) ds \end{bmatrix} \right. \\ &\quad \left. \times \int_0^\infty \mathbf{g}^T(s, W_i, \mathbf{Z}_i) \frac{Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})} ds \right). \end{aligned}$$

Among the three expectations the first term is $E[\Delta_i \boldsymbol{\phi}\{\mathbf{O}_i; \boldsymbol{\beta},\ \Lambda(V_i), \boldsymbol{\gamma}(\boldsymbol{\beta}), \boldsymbol{\alpha}\} \mathbf{g}^T(V_i, W_i, \mathbf{Z}_i)]$. The second expectation is

$$\begin{aligned} &E\left(\int_0^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(u)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \\ \\ \{W_i\gamma_1^2 + \Lambda(u)(\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}\} \end{bmatrix} dN_i(u) \right. \\ &\quad \left. \times \int_0^\infty \mathbf{D}_4^T(s)\{\gamma_1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol{\beta})\} \frac{Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol{\beta})} ds \right) \end{aligned}$$

$$
\begin{aligned}
= & \; E\left(\int_0^\infty \int_u^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(u)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta)\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\} \\ \\ \{W_i\gamma_1^2 + \Lambda(u)(\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta)\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\} \end{bmatrix}\right. \\
& \left. \times \mathbf{D}_4^T(s)\{\gamma_1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta)\} \frac{\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)} Y_i(s) dN_i(u) ds\right) \qquad \text{(S.6)} \\
& + E\left(\int_0^\infty \int_s^\infty \begin{bmatrix} \mathbf{Z}_i\{\gamma_1 + \Lambda(u)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta)\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\} \\ \\ \{W_i\gamma_1^2 + \Lambda(u)(\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta)\} f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\} \end{bmatrix}\right. \\
& \left. \times \mathbf{D}_4^T(s)\{\gamma_1 + \Lambda(s)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta)\} \frac{\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)} Y_i(s) dN_i(u) ds\right). \qquad \text{(S.7)}
\end{aligned}
$$

Note that the expression given in (S.6) is zero as $E\{Y_i(s)dN_i(u)I(s > u)\} = 0$. Now, (S.7) becomes

$$
\begin{aligned}
& E\left(\int_0^\infty \int_s^\infty \begin{bmatrix} \mathbf{Z}_i[\gamma_1^2 + \gamma_1^2\{\Lambda(s) + \Lambda(u)\}\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta) + \Lambda(s)\Lambda(u)\eta^2(X_i, \mathbf{Z}_i, \boldsymbol\beta)E\{\exp(2\beta_2 U)\}] \\ \\ \gamma_1^3 X_i + \gamma_1^3 X_i \eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)\Lambda(u) \\ +\gamma_1^2\eta(X_i, Z_i, \boldsymbol\beta)\Lambda(s)[\gamma_1 X_i + E\{U\exp(\beta_2 U)\}] + \eta^2(X_i, Z_i, \boldsymbol\beta)\Lambda(s)\Lambda(u) \\ \times[\gamma_1 X_i E\{\exp(2\beta_2 U)\} + \gamma_1 E\{U\exp(2\beta_2 U)\} - \gamma_2 E\{\exp(2\beta_2 U)\}] \end{bmatrix}\right. \\
& \left. \times f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\}\mathbf{D}_4^T(s) \times \frac{Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)} dN_i(u) ds\right) \\
= & \; E\left(\int_0^\infty \int_s^\infty \begin{bmatrix} \mathbf{Z}_i[\gamma_1^2 + \gamma_1^2\{\Lambda(s) + \Lambda(u)\}\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta) + \Lambda(s)\Lambda(u)\eta^2(X_i, \mathbf{Z}_i, \boldsymbol\beta)\kappa_1] \\ \\ \gamma_1^3 X_i + \gamma_1^3 X_i\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)\Lambda(u) + \gamma_1^2\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)\Lambda(s)(\gamma_1 X_i + \gamma_2) \\ +\eta^2(X_i, \mathbf{Z}_i, \boldsymbol\beta)\Lambda(s)\Lambda(u)(\gamma_1\kappa_1 X_i + \gamma_1\kappa_2 - \gamma_2\kappa_1) \end{bmatrix}\right. \\
& \left. \times f\{\Lambda(u), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\}\mathbf{D}_4^T(s) \frac{Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)} dN_i(u) ds\right).
\end{aligned}
$$

Now the last term of $\text{cov}(\mathbf{A}_1, \mathbf{A}_2^*)$ is

$$
\begin{aligned}
& E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta) \\ \\ (\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta) \end{bmatrix}\right. \\
& \left. \quad \times \int_0^\infty Y_i(s) f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\}\lambda(s) ds \int_0^\infty \mathbf{g}^T(s, W_i, \mathbf{Z}_i) \frac{Y_i(s)\lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)}{1 + \Lambda(s)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)} ds\right) \\
= & \; E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta) \\ \\ (\gamma_1 W_i - \gamma_2)\eta(W_i, \mathbf{Z}_i, \boldsymbol\beta) \end{bmatrix}\right. \\
& \left. \quad \times \int_0^\infty \int_s^\infty f\{\Lambda(s), \mathbf{Z}_i, \boldsymbol\beta, \boldsymbol\alpha\}\lambda(s)\mathbf{g}^T(u, W_i, \mathbf{Z}_i) \frac{\lambda(u)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)}{1 + \Lambda(u)\eta(X_i, \mathbf{Z}_i, \boldsymbol\beta)} Y_i(s) Y_i(u) ds du\right)
\end{aligned}
$$

$$
\begin{aligned}
&+E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \\ \\ (\gamma_1 W_i-\gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \end{bmatrix}\right. \\
&\quad\left.\times\int_0^\infty\int_u^\infty f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{g}^T(u,W_i,\mathbf{Z}_i)\frac{\lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}Y_i(s)Y_i(u)dsdu\right) \\
= {}& E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \\ \\ (\gamma_1 W_i-\gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \end{bmatrix}\right. \\
&\quad\left.\times\int_0^\infty\int_s^\infty f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{g}^T(u,W_i,\mathbf{Z}_i)\frac{\lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}Y_i(u)dsdu\right) \\
&+E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \\ \\ (\gamma_1 W_i-\gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \end{bmatrix}\right. \\
&\quad\left.\times\int_0^\infty\int_u^\infty f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{g}^T(u,W_i,\mathbf{Z}_i)\frac{\lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}Y_i(s)dsdu\right) \\
= {}& E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \\ \\ (\gamma_1 W_i-\gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \end{bmatrix}\int_0^\infty\int_s^\infty f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{g}^T(u,W_i,\mathbf{Z}_i)dN_i(u)ds\right) \\
&+E\left(\int_0^\infty\int_u^\infty\begin{bmatrix} \mathbf{Z}_i\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})[\gamma_1^2+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})E\{\exp(2\beta_2 U)\}] \\ \\ \gamma_1^3\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})X_i+\Lambda(u)\eta^2(X_i,\mathbf{Z}_i,\boldsymbol{\beta})[\gamma_1 X_i E\{\exp(2\beta_2 U)\} \\ +\gamma_1 E\{U\exp(2\beta_2 U)\}-\gamma_2 E\{\exp(2\beta_2 U)\}] \end{bmatrix}\right. \\
&\quad\left.\times f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{D}_4^T(u)\frac{\lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}Y_i(s)dsdu\right) \qquad \text{(S.8)}
\end{aligned}
$$

Expression (S.8) can be written as

$$
\begin{aligned}
&E\left(\int_0^\infty\int_u^\infty\begin{bmatrix} \mathbf{Z}_i\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\{\gamma_1^2+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\kappa_1\} \\ \\ \gamma_1^3\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})X_i+\Lambda(u)\eta^2(X_i,\mathbf{Z}_i,\boldsymbol{\beta})(\gamma_1 X_i\kappa_1+\gamma_1\kappa_2-\gamma_2\kappa_1) \end{bmatrix}\right. \\
&\quad\left.\times f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{D}_4^T(u)\frac{\lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}Y_i(s)dsdu\right).
\end{aligned}
$$

Combining the above derivation we can write

$$
\mathbf{G}^{(5)} = E[\Delta_i\boldsymbol{\phi}\{\mathbf{O}_i;\boldsymbol{\beta},\Lambda(V_i),\boldsymbol{\gamma}(\boldsymbol{\beta}),\boldsymbol{\alpha}\}\mathbf{g}^T(V_i,W_i,\mathbf{Z}_i)]
$$

$$+E\left(\begin{bmatrix} \mathbf{Z}_i\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \\ (\gamma_1 W_i-\gamma_2)\eta(W_i,\mathbf{Z}_i,\boldsymbol{\beta}) \end{bmatrix}\int_0^\infty\int_s^\infty f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{g}^T(u,W_i,\mathbf{Z}_i)dN_i(u)ds\right)$$
$$+E\left(\int_0^\infty\int_u^\infty \begin{bmatrix} \mathbf{Z}_i\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\{\gamma_1^2+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\kappa_1\} \\ \gamma_1^3\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})X_i+\Lambda(u)\eta^2(X_i,\mathbf{Z}_i,\boldsymbol{\beta})(\gamma_1X_i\kappa_1+\gamma_1\kappa_2-\gamma_2\kappa_1) \end{bmatrix}\right.$$
$$\left.\times f\{\Lambda(s),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\lambda(s)\mathbf{D}_4^T(u)\frac{\lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(u)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}Y_i(s)dsdu\right)$$
$$-E\left(\int_0^\infty\int_s^\infty \begin{bmatrix} \mathbf{Z}_i[\gamma_1^2+\gamma_1^2\{\Lambda(s)+\Lambda(u)\}\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})+\Lambda(s)\Lambda(u)\eta^2(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\kappa_1] \\ \gamma_1^3X_i+\gamma_1^3X_i\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\Lambda(u)+\gamma_1^2\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\Lambda(s)(\gamma_1X_i+\gamma_2) \\ +\eta^2(X_i,\mathbf{Z}_i,\boldsymbol{\beta})\Lambda(s)\Lambda(u)(\gamma_1\kappa_1X_i+\gamma_1\kappa_2-\gamma_2\kappa_1) \end{bmatrix}\right.$$
$$\left.\times f\{\Lambda(u),\mathbf{Z}_i,\boldsymbol{\beta},\boldsymbol{\alpha}\}\mathbf{D}_4^T(s)\frac{Y_i(s)\lambda(s)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}{1+\Lambda(s)\eta(X_i,\mathbf{Z}_i,\boldsymbol{\beta})}dN_i(u)ds\right).$$

The first two expectations of $\mathbf{G}^{(5)}$ are estimated by the corresponding empirical averages while the last two expectations involve with unobserved X and their estimation is described in the main paper.

References

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

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer: New York.

Polyanin, A. D. and Manzhirov, A. V. (2008). *Handbook of Integral Equations, Second Edition.* Chapman & Hall/CRC.

Wellner, J. A. (2003). *Empirical Process: Theory and Applications.* https://www.stat.washington.edu/people/jaw/RESEARCH/TALKS/BocconiSS/emp-prc-bk-big2.pdf

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.