Online Supplementary Material "Longitudinal Modeling of Rank-based Global Outcome"

1 Proofs for Theorems 1-4

In this section, we prove the main theorems for the MRC estimator. Theorem 1 describes the consistency and asymptotic normality of the proposed estimator for a single longitudinal outcome. Theorem 2 generalizes results to the longitudinal GPO setting. Theorems 3 and 4 summarize the asymptotic behavior with missing data for the single longitudinal outcome and the longitudinal GPO settings, respectively.

1.1 Proof of Theorem 1

Denote $\mathcal{L}(\boldsymbol{\theta}) = \sum_{u=0}^{M} \sum_{v=0}^{M} \mathbb{E}[I(Y_{iu} > Y_{jv}) I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}]$ and $\mu_{im}(\boldsymbol{\theta}) = \mu(\mathbf{X}_i, t_m, \boldsymbol{\theta})$.

1.1.1 Consistency

We assume the following regularity conditions for the consistency of $\widehat{\boldsymbol{\theta}}$:

(C1) $\boldsymbol{\theta}_0$ is an interior point of Θ, where Θ is a compact subset of $\mathbb{R}^q$.

(C2) There exists a unique maximizer for $\mathcal{L}(\boldsymbol{\theta})$ in the interior of Θ.

(C3) $\mathcal{L}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_0$.

Throughout our proof, we assume that the function class $\{\mu(\cdot, \cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ has finite Vapnik-Chervonenkis dimensions. Condition (C1) is a standard condition for the consistency of $\widehat{\boldsymbol{\theta}}$ (Cavanagh & Sherman 1998, Khan & Tamer 2007). Conditions (C2)-(C3) are easily satisfied under many realistic settings such as a setting with $\mu_{im}(\boldsymbol{\theta}) = \sum_{k=1}^{p} \beta_{mk} X_{ik} + \beta_t t_m$, where $\boldsymbol{\beta} = (1, \boldsymbol{\theta}^T)^T = (\beta_{01}, \ldots, \beta_{0p}, \beta_{11}, \ldots, \beta_{1p}, \ldots, \beta_{M1}, \ldots, \beta_{Mp}, \beta_t)$.

Identifiability of $\boldsymbol{\theta}_0$: Given $\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)$ and the monotonicity of ζ, we have

$$\mathbb{P}_{\boldsymbol{\varepsilon}}(Y_{iu} > Y_{jv}) \geq \mathbb{P}_{\boldsymbol{\varepsilon}}(Y_{iu} < Y_{jv}),$$

where $\mathbb{P}_{\boldsymbol{\varepsilon}}$ is the expectation with respect to $\boldsymbol{\varepsilon}$. Thus $\boldsymbol{\theta}_0$ maximizes the following function:

$$\mathbb{E}\left[I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}\mathbb{P}_{\boldsymbol{\varepsilon}}(Y_{iu} > Y_{jv}) + I\{\mu_{iu}(\boldsymbol{\theta}) < \mu_{jv}(\boldsymbol{\theta})\}\mathbb{P}_{\boldsymbol{\varepsilon}}(Y_{iu} < Y_{jv})\right].$$

It follows that $\boldsymbol{\theta}_0$ maximizes

$$\sum_{u=0}^{M}\sum_{v=0}^{M}\mathbb{E}\left[I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}\mathbb{P}_{\boldsymbol{\varepsilon}}(Y_{iu} > Y_{jv}) + I\{\mu_{iu}(\boldsymbol{\theta}) < \mu_{jv}(\boldsymbol{\theta})\}\mathbb{P}_{\boldsymbol{\varepsilon}}(Y_{iu} < Y_{jv})\right],$$

and therefore maximizes the objective function, $\mathcal{L}(\boldsymbol{\theta})$. By Condition (C2), we know that $\boldsymbol{\theta}_0$ is the unique maximizer of $\mathcal{L}(\boldsymbol{\theta})$.

Uniform convergence of $\mathcal{L}_n(\boldsymbol{\theta})$: A sufficient condition for the uniform convergence is that the kernel function class of the U-statistic $\mathcal{L}_n(\boldsymbol{\theta})$ satisfies the Euclidean property for some square integrable envelope. Consider the function class $\mathcal{F}_{u,v} = \{f_{u,v}(\cdot,\cdot,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}, 0 \leq u, v \leq M$, where $\mathbf{z}_1 = (\mathbf{x}_1, \mathbf{y}_1)$, $\mathbf{z}_2 = (\mathbf{x}_2, \mathbf{y}_2)$ and

$$f_{u,v}(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta}) = I(y_{1u} > y_{2v})I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) > \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\} + I(y_{1u} < y_{2v})I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) < \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\}.$$

Following Section 5 of Sherman (1993), $\mathcal{F}_{u,v}$ is Euclidean with a constant envelope of 1. By Lemma 2.14 of Pakes et al. (1989), $\mathcal{F}$ is also Euclidean with a constant envelope of $(M+1)^2$, where

$$\begin{aligned} f(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta}) =& \frac{1}{2}\sum_{u=0}^{M}\sum_{v=0}^{M}\left[I(y_{1u} > y_{2v})I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) > \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\}\right. \\ &\left. + I(y_{1u} < y_{2v})I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) < \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\}\right], \end{aligned}$$

and $\mathcal{F} = \{f(\cdot,\cdot,\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$. The uniform convergence of $\mathcal{L}_n(\boldsymbol{\theta})$ then follows from Conditions

(C1), (C3) and Corollary 7 of Sherman (1994). By Theorem 5.7 in Van der Vaart (2000), the uniform convergence implies that $\widehat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$.

1.1.2 Asymptotic normality

For each $\boldsymbol{\theta} \in \Theta$, denote $\Gamma(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta}_0)$ and $\Gamma_n(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}_n(\boldsymbol{\theta}_0)$. Note that $\Gamma_n(\boldsymbol{\theta}_0) = \Gamma(\boldsymbol{\theta}_0) = 0$, $\mathbb{E}\{\tau(\cdot, \boldsymbol{\theta}) - \tau(\cdot, \boldsymbol{\theta}_0)\} = 2\Gamma(\boldsymbol{\theta})$ and $\Gamma(\boldsymbol{\theta})$ is maximized at $\boldsymbol{\theta}_0$. We assume the following regularity conditions for the normality of $\widehat{\boldsymbol{\theta}}$:

The following additional condition is required for asymptotic normality.

(C4) There exists an integrable function $M(\mathbf{z})$ such that for all $\mathbf{z}$ and $\boldsymbol{\theta}$ in a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$, $\|\nabla_2\tau(\mathbf{z}, \boldsymbol{\theta}) - \nabla_2\tau(\mathbf{z}, \boldsymbol{\theta}_0)\| \leq M(\mathbf{z})\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$, where $\mathbb{E}\|\nabla_1\tau(\cdot, \boldsymbol{\theta}_0)\|^2 < \infty$, $\mathbb{E}|\nabla_2|\tau(\cdot, \boldsymbol{\theta}_0) < \infty$, and $\mathbb{E}\nabla_2\tau(\cdot, \boldsymbol{\theta}_0)$ is negative definite.

Condition (C4) assumes the smoothness and moment conditions of τ (Sherman 1993). For each $(\mathbf{z}_1, \mathbf{z}_2)$ and $\boldsymbol{\theta} \in \Theta$, define

$$f(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta}) = \sum_{u=0}^{M}\sum_{v=0}^{M} I(y_{1u} > y_{2v}) \left[I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) > \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\} - I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}_0) > \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta}_0)\}\right].$$

Using Hoeffding decomposition, we have $\Gamma_n(\boldsymbol{\theta}) = \Gamma(\boldsymbol{\theta}) + \mathbb{P}_n g(\cdot, \boldsymbol{\theta}) + U_n h(\cdot, \cdot, \boldsymbol{\theta})$, where $g(\mathbf{z}, \boldsymbol{\theta}) = \mathbb{E}f(\mathbf{z}, \cdot, \boldsymbol{\theta}) + \mathbb{E}f(\cdot, \mathbf{z}, \boldsymbol{\theta}) - 2\Gamma(\boldsymbol{\theta})$, and $h(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta}) = f(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta}) - \mathbb{E}f(\mathbf{z}_1, \cdot, \boldsymbol{\theta}) - \mathbb{E}f(\cdot, \mathbf{z}_2, \boldsymbol{\theta}) + \Gamma(\boldsymbol{\theta})$. By the same techniques as those in Sherman (1993), it can be shown that $\Gamma(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right)$, $\mathbb{P}_n g(\cdot, \boldsymbol{\theta}) = \frac{1}{\sqrt{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T W_n + o\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right)$, and $U_n h(\cdot, \cdot, \boldsymbol{\theta}) = o_P(1/n)$ uniformly over $o_P(1)$ neighbourhoods of $\boldsymbol{\theta}_0$, where $W_n = \sqrt{n}\mathbb{P}_n\nabla_1\tau(\cdot, \boldsymbol{\theta}_0) \xrightarrow{D} N(0, \Delta)$. Thus we have

$$\Gamma_n(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{\sqrt{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T W_n + o\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + o_P(1/n) \qquad (1)$$

uniformly over $o_P(1)$ neighbourhoods of $\boldsymbol{\theta}_0$. Combining Equation (1), Condition (C4) and Theorem 2 of Sherman (1993), we have that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to $N(0, V^{-1}\Delta V^{-1})$.

1.2 Proof of Theorem 2

1.2.1 Consistency

Denote $\mathcal{L}_{n,\mathcal{P}}(\boldsymbol{\theta}) = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \sum_{u=0}^{M} \sum_{v=0}^{M} I(\mathcal{P}_{iu} > \mathcal{P}_{jv}) I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}$, and $\mathcal{L}_{\mathcal{P}}(\boldsymbol{\theta}) = \sum_{u=0}^{M} \sum_{v=0}^{M} \mathbb{E}[I(\mathcal{P}_{iu} > \mathcal{P}_{jv}) I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}]$. Throughout our proof, we assume that the function class, $\{\mu(\cdot,\cdot,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_{\mathcal{P}}\}$, has finite Vapnik-Chervonenkis dimensions. We assume the following regularity conditions for the consistency of $\widehat{\boldsymbol{\theta}}_{\mathcal{P}}$:

(C5) $\boldsymbol{\theta}_{0,\mathcal{P}}$ is an interior point of $\Theta_{\mathcal{P}}$, where $\Theta_{\mathcal{P}}$ is a compact subset of $\mathbb{R}^q$.

(C6) There exists some $\boldsymbol{\theta}^*_{\mathcal{P}}$ in the interior of $\Theta_{\mathcal{P}}$ that uniquely maximizes $\mathcal{L}_{\mathcal{P}}(\boldsymbol{\theta})$.

(C7) $\mathcal{L}_{\mathcal{P}}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_{0,\mathcal{P}}$.

We first show that $\sup_{\boldsymbol{\theta}\in\Theta_{\mathcal{P}}} |\mathcal{L}_n^{GPO}(\boldsymbol{\theta}) - \mathcal{L}_{n,\mathcal{P}}(\boldsymbol{\theta})| = o_P(1)$. The difference between $\mathcal{L}_n^{GPO}(\boldsymbol{\theta})$ and $\mathcal{L}_{n,\mathcal{P}}(\boldsymbol{\theta})$ has the following upper bound:

$$\begin{aligned}
&|\mathcal{L}_n^{GPO}(\boldsymbol{\theta}) - \mathcal{L}_{n,\mathcal{P}}(\boldsymbol{\theta})| \\
=&\frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \sum_{u=0}^{M} \sum_{v=0}^{M} |I(\widehat{\mathcal{P}}_{iu} > \widehat{\mathcal{P}}_{jv}) - I(\mathcal{P}_{iu} > \mathcal{P}_{jv})| I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} \\
\le&\frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \sum_{u=0}^{M} \sum_{v=0}^{M} |I(\widehat{\mathcal{P}}_{iu} > \widehat{\mathcal{P}}_{jv}) - I(\mathcal{P}_{iu} > \mathcal{P}_{jv})|.
\end{aligned} \tag{2}$$

By the triangular inequality and Donsker's theorem, it can be shown that

$$\begin{aligned}
&\sup_{1\le i\ne j\le n, 0\le u,v\le M} \sqrt{n}|(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv}) - (\mathcal{P}_{iu} - \mathcal{P}_{jv})| \le \sup_{1\le i\le n, 0\le u\le M} 2\sqrt{n}|\widehat{\mathcal{P}}_{iu} - \mathcal{P}_{iu}| \\
\le&\frac{2}{K} \sum_{k=1}^{K} \sup_{t,u} \sqrt{n}|F_{ku}(t) - \widehat{F}_{ku}(t)| = O_P(1).
\end{aligned}$$

It follows that

$$\mathbb{P}\left\{\sup_{1\le i\ne j\le n, 0\le u,v\le M} |(\widehat{\mathcal{P}}_{iu}-\widehat{\mathcal{P}}_{jv})-(\mathcal{P}_{iu}-\mathcal{P}_{jv})|>n^{-1/3}\right\}=o(1). \tag{3}$$

On the other hand,

$$\sup_{0\le u,v\le M}\mathbb{P}(|\mathcal{P}_{iu}-\mathcal{P}_{jv}|\le 2n^{-1/3})\le \sup_{0\le u,v\le M}\sum_{k=1}^{K}\mathbb{P}\{|F_{ku}(Y_{iku})-F_{kv}(Y_{jkv})|\le 2n^{-1/3}\}=o(1). \tag{4}$$

Notice that the event $\left\{|(\widehat{\mathcal{P}}_{iu}-\widehat{\mathcal{P}}_{jv})-(\mathcal{P}_{iu}-\mathcal{P}_{jv})|\le n^{-1/3}\right\}\cap\left\{|\mathcal{P}_{iu}-\mathcal{P}_{jv}|>2n^{-1/3}\right\}$ is contained in $\left\{|I(\widehat{\mathcal{P}}_{iu}>\widehat{\mathcal{P}}_{jv})-I(\mathcal{P}_{iu}>\mathcal{P}_{jv})|=0\right\}$. By Equations (3) and(4), the supremum of $\left\{\mathbb{E}\{|I(\widehat{\mathcal{P}}_{iu}>\widehat{\mathcal{P}}_{jv})-I(\mathcal{P}_{iu}>\mathcal{P}_{jv})|\}:1\le i\ne j\le n, 0\le u,v\le M\right\}$ vanishes as n goes to infinity. Equation(2) implies $\mathbb{E}\left\{\sup_{\boldsymbol{\theta}\in\Theta}|\mathcal{L}_n^{GPO}(\boldsymbol{\theta})-\mathcal{L}_{n,\mathcal{P}}(\boldsymbol{\theta})|\right\}=o(1)$. Thus $\sup_{\boldsymbol{\theta}\in\Theta}|\mathcal{L}_n^{GPO}(\boldsymbol{\theta})-\mathcal{L}_{n,\mathcal{P}}(\boldsymbol{\theta})|=o_P(1)$. Then the consistency follows the same techniques as those used in Section 1.1.1

1.2.2 Asymptotic normality

The proof of asymptotic normality is challenged by the non-differentiable term $I(\widehat{\mathcal{P}}_{iu}>\widehat{\mathcal{P}}_{jv})$ in the objective function. We accordingly approximate it by some differentiable function in the proof below. Denote $K(\cdot)$ as a smooth kernel function such that $K(\cdot)\ge 0$, $K(x)=K(-x)$ and $\int K(x)dx=1$. Define $a_n=n^{-\nu}$ with $0<\nu<1/2$, $K_{a_n}(x)=a_n^{-1}K(x/a_n)$ and

$\widetilde{K}_{a_n}(t) = \int_{-\infty}^{t} K_{a_n}(u)du$. Let $\mathbf{z} = (y_{10}, \ldots, y_{KM}, \mathbf{x})$, and $\mathbf{Z}_i = (Y_{i10}, \ldots, Y_{iKM}, \mathbf{X}_i)$. Define

$$
\begin{aligned}
\xi_{iku} &= \frac{1}{K}\sum_{l=1}^{K}\left\{I(Y_{ilu} \geq Y_{klu}) - F_{lu}(Y_{ilu})\right\}, \\
g_{1n}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \boldsymbol{\theta}) &= \sum_{u=0}^{M}\sum_{v=0}^{M} K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv})(\xi_{iku} - \xi_{jkv}) I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\}, \\
\tau_{1n}(\mathbf{z}, \boldsymbol{\theta}) &= \mathbb{E}g_{1n}(\mathbf{z}, \cdot, \cdot, \boldsymbol{\theta}) + \mathbb{E}g_{1n}(\cdot, \mathbf{z}, \cdot, \boldsymbol{\theta}) + \mathbb{E}g_{1n}(\cdot, \cdot, \mathbf{z}, \boldsymbol{\theta}) = \mathbb{E}g_{1n}(\cdot, \cdot, \mathbf{z}, \boldsymbol{\theta}), \\
f_{1n}(\mathbf{z}, \boldsymbol{\theta}) &= \tau_{1n}(\mathbf{z}, \boldsymbol{\theta}) - \tau_{1n}(\mathbf{z}, \boldsymbol{\theta}_0), \\
E_{luv}(\mathbf{z}, \boldsymbol{\theta}) &= \mathbb{E}\left[\left\{I(Y_{ilu} \geq y_{lu}) - I(Y_{jlv} \geq y_{lv})\right\} I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} \middle| \mathcal{P}_{iu} = \mathcal{P}_{jv}\right], \\
\tau_{1}(\mathbf{z}, \boldsymbol{\theta}) &= \sum_{u,v}\int p_u(s)p_v(s)ds \cdot \frac{1}{K}\sum_{l=1}^{K} E_{luv}(\mathbf{z}, \boldsymbol{\theta}), \\
g_{2n}(\mathbf{Z}_i, \mathbf{Z}_j, \boldsymbol{\theta}) &= \sum_{u=0}^{M}\sum_{v=0}^{M} \widetilde{K}_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv}) I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\}, \\
\tau_{2n}(\mathbf{z}, \boldsymbol{\theta}) &= \mathbb{E}g_{2n}(\mathbf{z}, \cdot, \boldsymbol{\theta}) + \mathbb{E}g_{2n}(\cdot, \mathbf{z}, \boldsymbol{\theta}), \\
g_{2}(Z_i, Z_j, \boldsymbol{\theta}) &= \sum_{u=0}^{M}\sum_{v=0}^{M} I(\mathcal{P}_{iu} \geq \mathcal{P}_{jv}) I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\}, \\
\text{and } \tau_{2}(z, \boldsymbol{\theta}) &= \mathbb{E}g_{2}(z, \cdot, \boldsymbol{\theta}) + \mathbb{E}g_{2}(\cdot, z, \boldsymbol{\theta}),
\end{aligned}
$$

where $p_u(\cdot)$, $p_v(\cdot)$ are marginal densities for $\mathcal{P}_{iu}$, $\mathcal{P}_{jv}$, respectively. Note that $\mathbb{E}\tau_{1n}(\cdot, \boldsymbol{\theta}) = \mathbb{E}g_{1n}(Z_i, Z_j, Z_k, \boldsymbol{\theta}) = 0$. We assume the following regularity conditions for the normality of $\widehat{\boldsymbol{\theta}}_{\mathcal{P}}$.

(C8) There exists an integrable function $M_1(\mathbf{z})$ such that for all $\mathbf{z}$ and $\boldsymbol{\theta}$ in a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$, $\|\nabla_2\tau_{1n}(\mathbf{z}, \boldsymbol{\theta}) - \nabla_2\tau_{1n}(\mathbf{z}, \boldsymbol{\theta}_0)\| \leq M_1(\mathbf{z})\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$, where $\sup_n \mathbb{E}\|\nabla_1\tau_{1n}(\cdot, \boldsymbol{\theta}_0)\|^2 < \infty$ and $\sup_n \mathbb{E}|\nabla_2|\tau_{1n}(\cdot, \boldsymbol{\theta}_0) < \infty$.

(C9) $\mathbb{E}\|\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0)\|^2 < \infty$.

(C10) There exists an integrable function, $M_2(\mathbf{z})$, such that for all $\mathbf{z}$ and $\boldsymbol{\theta}$ in a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$, $\|\nabla_2\tau_{2n}(\mathbf{z}, \boldsymbol{\theta}) - \nabla_2\tau_{2n}(\mathbf{z}, \boldsymbol{\theta}_0)\| \leq M_2(\mathbf{z})\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$, where $\sup_n \mathbb{E}\|\nabla_1\tau_{2n}(\cdot, \boldsymbol{\theta}_0)\|^2 < \infty$ and $\sup_n \mathbb{E}|\nabla_2|\tau_{2n}(\cdot, \boldsymbol{\theta}_0) < \infty$.

(C11) $\mathbb{E}\|\nabla_1 \tau_2(\cdot, \boldsymbol{\theta}_0)\|^2 < \infty$ and the matrix $\nabla_2 \tau_2(\cdot, \boldsymbol{\theta}_0)$ is negative definite.

We first decompose the objective function into three terms:

$$\begin{aligned}
&\mathcal{L}_n^{GPO}(\boldsymbol{\theta}) - \mathcal{L}_n^{GPO}(\boldsymbol{\theta}_0) \\
=&\frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u,v} \left\{ I(\widehat{\mathcal{P}}_{iu} > \widehat{\mathcal{P}}_{jv}) - \widetilde{K}_{a_n}(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv}) \right\} \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\} \right] \\
&+ \frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u,v} \left\{ \widetilde{K}_{a_n}(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv}) - \widetilde{K}_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv}) \right\} \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\} \right] \\
&+ \frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u,v} \widetilde{K}_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv}) \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\} \right] \\
=&I + II + III.
\end{aligned} \tag{5}$$

The first term can be controlled uniformly over $o_P(1)$ neighbourhood of $\boldsymbol{\theta}_0$ by

$$\begin{aligned}
|I| &\leq \frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u=0}^{M} \sum_{v=0}^{M} \widetilde{K}_{a_n}(-|\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv}|) \left| I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\} \right| \\
&= o_P(1) \sum_{u,v} \frac{1}{n(n-1)} \sum_{i\neq j} \left| I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\} \right| \\
&= o_P(1) \sum_{u,v} \left\{ O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2) + \frac{1}{\sqrt{n}} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T O_P(1) + o_P(1/n) \right\} \\
&= o_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T o_P(1/\sqrt{n}) + o_P(1/n).
\end{aligned} \tag{6}$$

For the second term, we take the Taylor expansion at $\mathcal{P}_{iu} - \mathcal{P}_{jv}$,

$$\widetilde{K}_{a_n}(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv}) - \widetilde{K}_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv}) = K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv})(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv} - \mathcal{P}_{iu} + \mathcal{P}_{jv}) + O\left(\frac{\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv} - \mathcal{P}_{iu} + \mathcal{P}_{jv}}{a_n} \right).$$

It can be shown that $a_n^{-1}(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv} - \mathcal{P}_{iu} + \mathcal{P}_{jv}) = o_P(1)$ uniformly over $1 \le i \ne j \le n$. By these results, the second term can be further decompose into two part,

$$
\begin{aligned}
&II = II(a) + II(b) \\
=&\frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u,v} K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv})(\widehat{\mathcal{P}}_{iu} - \widehat{\mathcal{P}}_{jv} - \mathcal{P}_{iu} + \mathcal{P}_{jv}) \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right] \\
&+ o_P(1) \frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u,v} \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right].
\end{aligned}
$$

Part $II(a)$ can be controlled uniformly over $o_P(1)$ neighbourhood of $\boldsymbol{\theta}_0$ by

$$
\begin{aligned}
&\frac{1}{n(n-1)} \sum_{i\neq j} \sum_{u,v} K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv}) \frac{1}{n} \sum_{k=1}^{n} (\xi_{iku} - \xi_{jkv}) \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right] \\
=&\frac{1}{n^2(n-1)} \sum_{i\neq j\neq k} \sum_{u,v} K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv})(\xi_{iku} - \xi_{jkv}) \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right] \\
&+ \frac{1}{n^2(n-1)} \sum_{i\neq j} \sum_{u,v} K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv})(\xi_{iiu} - \xi_{jiv} + \xi_{iju} - \xi_{jjv}) \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right] \\
=&\frac{1}{n^2(n-1)} \sum_{i\neq j\neq k} \sum_{u,v} K_{a_n}(\mathcal{P}_{iu} - \mathcal{P}_{jv})(\xi_{iku} - \xi_{jkv}) \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right] + o_P(n^{-1}) \\
=&\frac{n-2}{n} \frac{1}{\sqrt{n}} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T W_{1n} + o_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + o_P(n^{-1}),
\end{aligned} \tag{7}
$$

where $W_{1n} = \sqrt{n}\mathbb{P}_n \nabla_1 \tau_{1n}(\cdot, \boldsymbol{\theta}_0)$. When the conditional density function of $(\mathcal{P}_{iu}, \mathcal{P}_{jv})|(\mathbf{X}_i, \mathbf{X}_j)$ is thrice continuously differentiable with probability 1, $\tau_{1n}(z, \boldsymbol{\theta}) \to \tau_1(z, \boldsymbol{\theta})$ and $W_{1n} \stackrel{P}{\to} \sqrt{n}\mathbb{P}_n \nabla_1 \tau_1(\cdot, \boldsymbol{\theta}_0)$ as $n \to \infty$. Therefore,

$$
II(b) = o_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2) + \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T o_P(1/\sqrt{n}) + o_P(1/n), \tag{8}
$$

uniformly over $o_P(1)$ neighbourhood of $\boldsymbol{\theta}_0$.

By some algebraic calculation, the third term converges uniformly over $o_P(1)$ neighbourhood of $\boldsymbol{\theta}_0$ to

$$
III = \frac{1}{2} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T V_n \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \frac{1}{\sqrt{n}} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T W_{2n} + o_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + o_P(1/n), \tag{9}
$$

where $2V_n = \mathbb{E}\nabla_2\tau_{2n}(\cdot, \boldsymbol{\theta}_0)$ and $W_{2n} = \sqrt{n}\mathbb{P}_n\nabla_1\tau_{2n}(\cdot, \boldsymbol{\theta}_0)$. Notice that $W_{2n} - \sqrt{n}\mathbb{P}_n\nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0) \xrightarrow{P} 0$ and $V_n \xrightarrow{P} V_{\mathcal{P}} = \mathbb{E}\nabla_2\tau_2(\cdot, \boldsymbol{\theta}_0)$. Plugging Equations (6) - (9) into Equation(5), we have

$$\mathcal{L}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}) - \mathcal{L}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}_0) = \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T V_{\mathcal{P}}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \frac{1}{\sqrt{n}}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \left(\frac{n-2}{n}W_{1n} + W_{2n}\right) + o_P(n^{-1}),$$

uniformly over $O_P(1/\sqrt{n})$ neighbourhoods of $\boldsymbol{\theta}_0$. Then by Theorem 2 of Sherman (1993), $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathcal{P}} - \boldsymbol{\theta}_{0,\mathcal{P}}) \xrightarrow{D} \mathcal{N}(0, V_{\mathcal{P}}^{-1}\Delta_{\mathcal{P}}V_{\mathcal{P}}^{-1})$, where $\Delta_{\mathcal{P}} = \mathbb{E}\left\{\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0) + \nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0)\right\}^{\otimes 2}$.

1.3 Proof of Theorem 3

We assume the following regularity conditions for the consistency and normality of $\widehat{\boldsymbol{\theta}}^W$ and $\widehat{\boldsymbol{\theta}}^W_{\mathcal{P}}$ in the presence of missing data:

(C12) There exists a constant $c > 0$ such that $\min_{i,m} w_{im}(\boldsymbol{\alpha}) > c > 0$ for all $\boldsymbol{\alpha} \in \mathbf{A}$, where $\mathbf{A}$ is a compact set.

(C13) $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = n^{-1}\sum_{i=1}^n \kappa_i + o_P(n^{-1/2})$, where κ_i is the influence function of $\widehat{\boldsymbol{\alpha}}$.

(C14) $\max_{i,m} |w_{im}(\widehat{\boldsymbol{\alpha}}) - w_{im}(\boldsymbol{\alpha}_0)| = o_P(1)$.

Condition (C12) is commonly used in missing data literature (Robins et al. 1995). Conditions (C13)-(C14) are satisfied by many regression models such as logistic regression models with bounded covariates.

1.3.1 Consistency

Denote $\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}\sum_{u=0}^M\sum_{v=0}^M \frac{\eta_{iu}\eta_{jv}}{w_{iu}(\boldsymbol{\alpha})w_{jv}(\boldsymbol{\alpha})} I(Y_{iu} > Y_{jv}) I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}$. We first prove $\sup_{\boldsymbol{\theta}\in\Theta} |\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}) - \widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \boldsymbol{\alpha}_0)| = o_P(1)$. It suffices to show that

$$\left|\frac{1}{w_{iu}(\widehat{\boldsymbol{\alpha}})w_{jv}(\widehat{\boldsymbol{\alpha}})} - \frac{1}{w_{iu}(\boldsymbol{\alpha}_0)w_{jv}(\boldsymbol{\alpha}_0)}\right| = \frac{|w_{iu}(\boldsymbol{\alpha}_0)w_{jv}(\boldsymbol{\alpha}_0) - w_{iu}(\widehat{\boldsymbol{\alpha}})w_{jv}(\widehat{\boldsymbol{\alpha}})|}{w_{iu}(\widehat{\boldsymbol{\alpha}})w_{jv}(\widehat{\boldsymbol{\alpha}})w_{iu}(\boldsymbol{\alpha}_0)w_{jv}(\boldsymbol{\alpha}_0)} = o_P(1)$$

uniformly over i, j, u, v, which is implied by Condition (C12). Notice that $\mathbb{E}\{\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \boldsymbol{\alpha}_0)\} = \mathcal{L}(\boldsymbol{\theta})$. By using similar techniques in Section 1.1.1, it can be shown that $\boldsymbol{\theta}_0$ uniquely maximizes $\mathbb{E}\{\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \boldsymbol{\alpha}_0)\}$. Consider the function class, $\mathcal{F} = \{f(\cdot, \cdot, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, where

$$
\begin{aligned}
f(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta}) = & \frac{1}{2} \sum_{u=0}^{M} \sum_{v=0}^{M} \frac{\eta_{1u}\eta_{2v}}{w(\mathbf{x}_1, \mathbf{y}_1, t_u, \boldsymbol{\alpha}_0) w(\mathbf{x}_2, \mathbf{y}_2, t_v, \boldsymbol{\alpha}_0)} \left[I(y_{1u} > y_{2v}) I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) > \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\} \right. \\
& \left. + I(y_{1u} < y_{2v}) I\{\mu(\mathbf{x}_1, t_u, \boldsymbol{\theta}) < \mu(\mathbf{x}_2, t_v, \boldsymbol{\theta})\} \right].
\end{aligned}
$$

It follows from Lemma 2.14 of Pakes et al. (1989) that $\mathcal{F}$ is Euclidean with a constant envelope of $(M+1)^2/c^2$. By Corollary 7 in Sherman (1994), we have $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \boldsymbol{\alpha}_0) - \mathcal{L}(\boldsymbol{\theta})| = o_P(1)$, which completes the proof.

1.3.2 Asymptotic normality

We assume the following regularity conditions to establish the asymptotic normality of $\widehat{\boldsymbol{\theta}}^W$:

(C15) There exists an integrable function $M(\mathbf{z})$ such that for all $\mathbf{z}$ and $\boldsymbol{\theta}$ in a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$,

$$
\|\nabla_2 \tau(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) - \nabla_2 \tau(\mathbf{z}, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)\| \leq M(\mathbf{z})(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|).
$$

Here $\nabla_1 \tau(\cdot, \boldsymbol{\theta}, \boldsymbol{\alpha})$ is continuous at $\boldsymbol{\alpha}_0$, $\mathbb{E}\|\nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})\|^2 < \infty$, and $\mathbb{E}|\nabla_2|\tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}) < \infty$. The matrix $V(\boldsymbol{\alpha}_0) = \mathbb{E}\nabla_2 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)$ is negative definite and $V(\boldsymbol{\alpha})$ is continuous at $\boldsymbol{\alpha}_0$.

Note that $\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}) = \left\{ \widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}) - \widetilde{\mathcal{L}}_n(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\alpha}}) \right\} + \widetilde{\mathcal{L}}_n(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\alpha}})$. Denote $\Gamma_n(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \boldsymbol{\alpha}) - \widetilde{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\alpha})$. The function $\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\alpha}})$ is independent of $\boldsymbol{\theta}$, and therefore maximizing $\widetilde{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}})$ is equivalent to maximizing $\Gamma_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}})$. By similar techniques used in Section 1.1.2, we have

$$
\Gamma_n(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V(\boldsymbol{\alpha})(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{\sqrt{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T W_n(\boldsymbol{\alpha}) + o_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + o_P(1/n),
$$

uniformly over $O_P(1/\sqrt{n})$ neighbourhood of $(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)$, where $W_n(\boldsymbol{\alpha}) = \sqrt{n}\mathbb{P}_n \nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}) + \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathbb{E}\nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})$. It implies that

$$\Gamma_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{\sqrt{n}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T W_n(\widehat{\boldsymbol{\alpha}}) + o_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + o_P(1/n). \quad (10)$$

Because of the continuity of $V(\cdot)$ at $\boldsymbol{\alpha}_0$, the first term of equation (10) equals to

$$\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T V(\boldsymbol{\alpha}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right). \quad (11)$$

By the continuity of $\mathbb{E}\nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})$ at $\boldsymbol{\alpha}_0$ and the fact that $\mathbb{E}\nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) = 0$, the second term of equation (10) equals to

$$W_n(\widehat{\boldsymbol{\alpha}}) = \sqrt{n}\mathbb{P}_n \nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\alpha}}) + o_P(1) = \sqrt{n}\mathbb{P}_n \nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) + \mathbb{E}\nabla_{\boldsymbol{\alpha}} \nabla_1 \tau(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o_P(1). \quad (12)$$

Combining Equations (10) - (12) and Theorem 2 of Sherman (1993), we have $\sqrt{n}(\widehat{\boldsymbol{\theta}}^W - \boldsymbol{\theta}_0)$ converges in distribution to $\mathcal{N}(0, V(\boldsymbol{\alpha}_0)^{-1}\Delta(\boldsymbol{\alpha}_0)V(\boldsymbol{\alpha}_0)^{-1})$.

1.4 Proof of Theorem 4

1.4.1 Consistency

Denote $\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})$ and $\widehat{F}_{ku}(\cdot, \boldsymbol{\alpha})$ as the estimated $\mathcal{P}_{iu}$ and $F_{ku}(\cdot)$ using the IPW method. Per results in Section 1.2.1 and Section 1.3.1, it is suffice to show that

$$\sup_{1 \le i \ne j \le n, 0 \le u, v \le M} \sqrt{n}|\{\widehat{\mathcal{P}}_{iu}(\widehat{\boldsymbol{\alpha}}) - \widehat{\mathcal{P}}_{jv}(\widehat{\boldsymbol{\alpha}})\} - (\mathcal{P}_{iu} - \mathcal{P}_{jv})| = O_P(1). \quad (13)$$

Notice that, for $1 \le k \le K$ and $0 \le u \le M$,

$$\sup_t \sqrt{n}\left|\widehat{F}_{ku}(t, \widehat{\boldsymbol{\alpha}}) - \widehat{F}_{ku}(t, \boldsymbol{\alpha}_0)\right| = \sup_t n^{-1}\sum_{i=1}^n \eta_{iu}\sqrt{n}\left|\frac{1}{w_{iu}(\widehat{\boldsymbol{\alpha}})} - \frac{1}{w_{iu}(\boldsymbol{\alpha}_0)}\right| I(Y_{iku} \le t) = O_P(1). \quad (14)$$

Combing Lemma 2.14 in Pakes et al. (1989), Corollary 4A in Sherman (1994) and the fact that $\mathbb{E}\widehat{F}_{ku}(t, \boldsymbol{\alpha}_0) = F_{ku}(t)$, we know that, for $1 \leq k \leq K$ and $0 \leq u \leq M$,

$$\sup_t \sqrt{n} \left| \widehat{F}_{ku}(t, \boldsymbol{\alpha}_0) - F_{ku}(t) \right| = O_P(1). \tag{15}$$

Equations (14) and (15)imply Equation(13), and thus we conclude that $\widehat{\boldsymbol{\theta}}_{\mathcal{P}}^{W}$ converges in probability to $\boldsymbol{\theta}_{0,\mathcal{P}}$.

1.4.2 Asymptotic normality

Denote the objective function as $\widetilde{\mathcal{L}}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}})$,

$$\widetilde{\mathcal{L}}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{u,v} \frac{\eta_{iu}\eta_{jv}}{w_{iu}(\boldsymbol{\alpha}) w_{jv}(\boldsymbol{\alpha})} I\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha}) > \widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})\} I\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\}.$$

Let $\mathbf{z} = (y_{10}, \ldots, y_{KM}, \eta_1, \ldots, \eta_M, \mathbf{x})$, and $\mathbf{Z}_i = (Y_{i10}, \ldots, Y_{iKM}, \eta_{i1}, \ldots, \eta_{iM}, \mathbf{X}_i)$. Define

$$\begin{aligned}
\xi_{iku}(\boldsymbol{\alpha}) &= \frac{1}{K}\sum_{l=1}^{K}\left\{\frac{\eta_{ku}}{w_{ku}(\boldsymbol{\alpha})}I(Y_{ilu} \geq Y_{klu}) - F_{lu}(Y_{ilu}, \boldsymbol{\alpha})\right\},\\
F_{lu}(y, \boldsymbol{\alpha}) &= \mathbb{E}\left\{\frac{\eta_{ku}}{w_{ku}(\boldsymbol{\alpha})}I(y \geq Y_{klu})\right\},\\
g_{1n}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \sum_{u=0}^{M}\sum_{v=0}^{M} K_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha}) - \mathcal{P}_{jv}(\boldsymbol{\alpha})\}\{\xi_{iku}(\boldsymbol{\alpha}) - \xi_{jkv}(\boldsymbol{\alpha})\}I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\},\\
\tau_{1n}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \mathbb{E}g_{1n}(\mathbf{z}, \cdot, \cdot, \boldsymbol{\theta}, \boldsymbol{\alpha}) + \mathbb{E}g_{1n}(\cdot, \mathbf{z}, \cdot, \boldsymbol{\theta}, \boldsymbol{\alpha}) + \mathbb{E}g_{1n}(\cdot, \cdot, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \mathbb{E}g_{1n}(\cdot, \cdot, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}),\\
f_{1n}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \tau_{1n}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) - \tau_{1n}(\mathbf{z}, \boldsymbol{\theta}_0, \boldsymbol{\alpha}),\\
E_{luv}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \mathbb{E}\left[\left\{\frac{\eta_u I(Y_{ilu} \geq y_{lu})}{w_u(\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha})} - \frac{\eta_v I(Y_{jlv} \geq y_{lv})}{w_v(\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha})}\right\} I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\}\middle| \mathcal{P}_{iu}(\boldsymbol{\alpha}) = \mathcal{P}_{jv}(\boldsymbol{\alpha})\right],\\
I_{uv}(\boldsymbol{\alpha}) &= \int p_u(s, \boldsymbol{\alpha})p_v(s, \boldsymbol{\alpha})ds,\\
\tau_1(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \sum_{u,v} I_{uv}(\boldsymbol{\alpha})\frac{1}{K}\sum_{l=1}^{K} E_{luv}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}),\\
g_{2n}(\mathbf{Z}_i, \mathbf{Z}_j, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \sum_{u=0}^{M}\sum_{v=0}^{M} \frac{\eta_{iu}\eta_{jv}}{w_{iu}(\boldsymbol{\alpha})w_{jv}(\boldsymbol{\alpha})}\widetilde{K}_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha}) - \mathcal{P}_{jv}(\boldsymbol{\alpha})\}I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\},\\
\tau_{2n}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \mathbb{E}g_{2n}(\mathbf{z}, \cdot, \boldsymbol{\theta}, \boldsymbol{\alpha}) + \mathbb{E}g_{2n}(\cdot, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}),\\
g_2(\mathbf{Z}_i, \mathbf{Z}_j, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \sum_{u=0}^{M}\sum_{v=0}^{M} I\{\mathcal{P}_{iu}(\boldsymbol{\alpha}) \geq \mathcal{P}_{jv}(\boldsymbol{\alpha})\}I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\},\\
\text{and}\tau_2(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) &= \mathbb{E}g_2(\mathbf{z}, \cdot, \boldsymbol{\theta}, \boldsymbol{\alpha}) + \mathbb{E}g_2(\cdot, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}),
\end{aligned}$$

where $p_u(\cdot, \boldsymbol{\alpha})$, $p_v(\cdot, \boldsymbol{\alpha})$ are marginal densities for $\mathcal{P}_{iu}(\boldsymbol{\alpha})$, $\mathcal{P}_{jv}(\boldsymbol{\alpha})$, respectively. We assume the following regularity conditions to establish the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\mathcal{P}}^{W}$:

(C16) There exists an integrable function $M_1(\mathbf{z})$ such that for all $\mathbf{z}$ and $\boldsymbol{\theta}$ in a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$, $\|\nabla_2\tau_{1n}(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) - \nabla_2\tau_{1n}(\mathbf{z}, \boldsymbol{\theta}_0, \boldsymbol{\alpha})\| \leq M_1(\mathbf{z})\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$, where $\sup_{n, \boldsymbol{\alpha} \in \mathbf{A}} \mathbb{E}\|\nabla_1\tau_{1n}(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})\|^2 < \infty$ and $\sup_{n, \boldsymbol{\alpha} \in \mathbf{A}} \mathbb{E}|\nabla_2|\tau_{1n}(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}) < \infty$.

(C17) $\mathbb{E}\|\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)\|^2 < \infty$.

(C18) There exists an integrable function $M_2(\mathbf{z})$ such that for all $\mathbf{z}$ and $\boldsymbol{\theta}$ in a neighbourhood

$\mathcal{N}$ of $\boldsymbol{\theta}_0$, $\|\nabla_2\tau_{2n}(\mathbf{z},\boldsymbol{\theta},\boldsymbol{\alpha})-\nabla_2\tau_{2n}(\mathbf{z},\boldsymbol{\theta}_0,\boldsymbol{\alpha})\| \le M_2(\mathbf{z})\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|$, where $\sup_{n,\boldsymbol{\alpha}\in\mathbf{A}} \mathbb{E}\|\nabla_1\tau_{2n}(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha})\|^2 < \infty$ and $\sup_{n,\boldsymbol{\alpha}\in\mathbf{A}} \mathbb{E}|\nabla_2|\tau_{2n}(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha}) < \infty$.

(C19) $\mathbb{E}\|\nabla_1\tau_2(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)\|^2 < \infty$, and $\nabla_1\tau_2(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha})$ is differentiable with respect to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}_0$. The matrix $\nabla_2\tau_2(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$ is negative definite, and $\nabla_2\tau_2(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha})$ is continuous at $\boldsymbol{\alpha}_0$.

Similarly, we decompose the objective function into three terms:

$$\widetilde{\mathcal{L}}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta},\boldsymbol{\alpha}) - \widetilde{\mathcal{L}}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}_0,\boldsymbol{\alpha}) = \widetilde{I} + \widetilde{II} + \widetilde{III}, \tag{16}$$

where

$$\begin{aligned}
\widetilde{I} =& \frac{1}{n(n-1)} \sum_{i\neq j}\sum_{u,v} \frac{\eta_{iu}\eta_{jv}}{w_{iu}(\boldsymbol{\alpha})w_{jv}(\boldsymbol{\alpha})} \left[I\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha}) > \widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})\} - \widetilde{K}_{a_n}\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha}) - \widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})\}\right] \\
& \times \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right], \\
\widetilde{II} =& \frac{1}{n(n-1)} \sum_{i\neq j}\sum_{u,v} \frac{\eta_{iu}\eta_{jv}}{w_{iu}(\boldsymbol{\alpha})w_{jv}(\boldsymbol{\alpha})} \left[\widetilde{K}_{a_n}\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha}) - \widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})\} - \widetilde{K}_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha}) - \mathcal{P}_{jv}(\boldsymbol{\alpha})\}\right] \\
& \times \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right], \\
\widetilde{III} =& \frac{1}{n(n-1)} \sum_{i\neq j}\sum_{u,v} \frac{\eta_{iu}\eta_{jv}}{w_{iu}(\boldsymbol{\alpha})w_{jv}(\boldsymbol{\alpha})} \widetilde{K}_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha}) - \mathcal{P}_{jv}(\boldsymbol{\alpha})\} \\
& \times \left[I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right],
\end{aligned}$$

and $\mathcal{P}_{iu}(\boldsymbol{\alpha}) = \mathbb{E}\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})$. The first term can be controlled uniformly over $o_P(1)$ neighbourhood of $(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$ by

$$\begin{aligned}
|\widetilde{I}| \le & \frac{c^{-2}}{n(n-1)} \sum_{i\neq j}\sum_{u=0}^{M}\sum_{v=0}^{M} \widetilde{K}_{a_n}\{-|\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha}) - \widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})|\}\eta_{iu}\eta_{jv} \left| I\left\{\mu_{iu}(\boldsymbol{\theta}) > \mu_{jv}(\boldsymbol{\theta})\right\} - I\left\{\mu_{iu}(\boldsymbol{\theta}_0) > \mu_{jv}(\boldsymbol{\theta}_0)\right\}\right| \\
= & \, o_P(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2) + (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T o_P(1/\sqrt{n}) + o_P(1/n).
\end{aligned} \tag{17}$$

We take the following Taylor expansion for the second term:

$$\begin{aligned}
&\widetilde{K}_{a_n}\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})-\widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})\}-\widetilde{K}_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha})-\mathcal{P}_{jv}(\boldsymbol{\alpha})\}\\
=&K_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha})-\mathcal{P}_{jv}(\boldsymbol{\alpha})\}\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})-\widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})-\mathcal{P}_{iu}(\boldsymbol{\alpha})+\mathcal{P}_{jv}(\boldsymbol{\alpha})\}\\
&+O\left\{\frac{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})-\widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})-\mathcal{P}_{iu}(\boldsymbol{\alpha})+\mathcal{P}_{jv}(\boldsymbol{\alpha})}{a_n}\right\}.
\end{aligned}$$

Note that $a_n^{-1}\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})-\widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})-\mathcal{P}_{iu}(\boldsymbol{\alpha})+\mathcal{P}_{jv}(\boldsymbol{\alpha})\}=o_P(1)$ uniformly over $1\le i\ne j\le n$ and $O(1)$ neighbourhoods of $\boldsymbol{\alpha}_0$. Thus, the second term, uniformly over $o_P(1)$ neighbourhood of $(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$, equals to

$$\begin{aligned}
\widetilde{II}=&\frac{1}{n(n-1)}\sum_{i\ne j}\sum_{u,v}K_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha})-\mathcal{P}_{jv}(\boldsymbol{\alpha})\}\{\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})-\widehat{\mathcal{P}}_{jv}(\boldsymbol{\alpha})-\mathcal{P}_{iu}(\boldsymbol{\alpha})+\mathcal{P}_{jv}(\boldsymbol{\alpha})\}\\
&\times\left[I\left\{\mu_{iu}(\boldsymbol{\theta})>\mu_{jv}(\boldsymbol{\theta})\right\}-I\left\{\mu_{iu}(\boldsymbol{\theta}_0)>\mu_{jv}(\boldsymbol{\theta}_0)\right\}\right]\\
&+o_P(1)\frac{1}{n(n-1)}\sum_{i\ne j}\sum_{u,v}\left[I\left\{\mu_{iu}(\boldsymbol{\theta})>\mu_{jv}(\boldsymbol{\theta})\right\}-I\left\{\mu_{iu}(\boldsymbol{\theta}_0)>\mu_{jv}(\boldsymbol{\theta}_0)\right\}\right]\\
=&\widetilde{II}(a)+o_P(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2)+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T o_P(1/\sqrt{n})+o_P(1/n).
\end{aligned}\tag{18}$$

Since $\widehat{\mathcal{P}}_{iu}(\boldsymbol{\alpha})-\mathcal{P}_{iu}(\boldsymbol{\alpha})=n^{-1}\sum_{k=1}^n\xi_{iku}(\boldsymbol{\alpha})$, it can be shown that

$$\begin{aligned}
\widetilde{II}(a)=&\frac{1}{n^2(n-1)}\sum_{i\ne j\ne k}\sum_{u,v}K_{a_n}\{\mathcal{P}_{iu}(\boldsymbol{\alpha})-\mathcal{P}_{jv}(\boldsymbol{\alpha})\}\{\xi_{iku}(\boldsymbol{\alpha})-\xi_{jkv}(\boldsymbol{\alpha})\}\\
&\times\left[I\left\{\mu_{iu}(\boldsymbol{\theta})>\mu_{jv}(\boldsymbol{\theta})\right\}-I\left\{\mu_{iu}(\boldsymbol{\theta}_0)>\mu_{jv}(\boldsymbol{\theta}_0)\right\}\right]+o_P(n^{-1})\\
=&\frac{n-2}{n}\frac{1}{\sqrt{n}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T W_{1n}(\boldsymbol{\alpha})+o_P\left(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2\right)+o_P(n^{-1})
\end{aligned}$$

uniformly over $o_P(1)$ neighbourhood of $(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0)$, where $W_{1n}(\boldsymbol{\alpha})=\sqrt{n}\mathbb{P}_n\nabla_1\tau_{1n}(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha})$. When the conditional density function of $\{\mathcal{P}_{iu}(\boldsymbol{\alpha}),\mathcal{P}_{jv}(\boldsymbol{\alpha})\}|(\mathbf{X}_i,\mathbf{X}_j)$ is thrice continuously differentiable with probability,$\tau_{1n}(z,\boldsymbol{\theta},\boldsymbol{\alpha})\to\tau_1(z,\boldsymbol{\theta},\boldsymbol{\alpha})$ and $W_{1n}-\sqrt{n}\mathbb{P}_n\nabla_1\tau_1(\cdot,\boldsymbol{\theta}_0,\boldsymbol{\alpha})\xrightarrow{P}0$ as $n\to\infty$. By some algebraic calculation, the third term converges uniformly over $o_P(1)$

neighbourhood of $(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)$ to

$$\widetilde{III} = \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T V_n(\boldsymbol{\alpha})\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \frac{1}{\sqrt{n}}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T W_{2n}(\boldsymbol{\alpha}) + o_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right) + o_P(1/n), \quad (19)$$

where $2V_n(\boldsymbol{\alpha}) = \mathbb{E}\nabla_2\tau_{2n}(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})$ and $W_{2n}(\boldsymbol{\alpha}) = \sqrt{n}\mathbb{P}_n\nabla_1\tau_{2n}(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})$. Notice that $\tau_{2n}(z, \boldsymbol{\theta}, \boldsymbol{\alpha}) \to \tau_2(z, \boldsymbol{\theta}, \boldsymbol{\alpha})$, $W_{2n}(\boldsymbol{\alpha}) - \sqrt{n}\mathbb{P}_n\nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}) \xrightarrow{P} 0$ and $V_n(\boldsymbol{\alpha}) \xrightarrow{P} V_{\mathcal{P}}(\boldsymbol{\alpha}) = \mathbb{E}\nabla_2\tau_2(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha})$. By Equations (16)-(19), we have

$$\begin{aligned}
&\widetilde{\mathcal{L}}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}) - \widetilde{\mathcal{L}}_{n,\widehat{\mathcal{P}}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\alpha}}) \\
=&\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T V_{\mathcal{P}}(\boldsymbol{\alpha}_0)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \frac{1}{\sqrt{n}}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \left\{\frac{n-2}{n}W_{1n}(\widehat{\boldsymbol{\alpha}}) + W_{2n}(\widehat{\boldsymbol{\alpha}})\right\} + o_P(n^{-1}),
\end{aligned}$$

uniformly over $O_P(1/\sqrt{n})$ neighbourhoods of $\boldsymbol{\theta}_0$. Combining the fact that

$$\begin{aligned}
W_{1n}(\widehat{\boldsymbol{\alpha}}) &= \sqrt{n}\mathbb{P}_n\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) + \mathbb{E}\nabla_{\boldsymbol{\alpha}}\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)n^{-1/2}\sum_{i=1}^n \kappa_i + o_P(1), \\
W_{2n}(\widehat{\boldsymbol{\alpha}}) &= \sqrt{n}\mathbb{P}_n\nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) + \mathbb{E}\nabla_{\boldsymbol{\alpha}}\nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)n^{-1/2}\sum_{i=1}^n \kappa_i + o_P(1),
\end{aligned}$$

and Theorem 2 in Sherman (1993), $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathcal{P}}^W - \boldsymbol{\theta}_{0,\mathcal{P}}) \xrightarrow{D} \mathcal{N}(0, V_{\mathcal{P}}(\boldsymbol{\alpha}_0)^{-1}\Delta_{\mathcal{P}}(\boldsymbol{\alpha}_0)V_{\mathcal{P}}(\boldsymbol{\alpha}_0)^{-1})$, where $\Delta_{\mathcal{P}}(\boldsymbol{\alpha}_0) = \mathbb{E}\left[\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) + \nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) + \left\{\mathbb{E}\nabla_{\boldsymbol{\alpha}}\nabla_1\tau_1(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) + \mathbb{E}\nabla_{\boldsymbol{\alpha}}\nabla_1\tau_2(\cdot, \boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)\right\}\kappa_i\right]^{\otimes 2}$.

2 Additional Simulation Results

Tables S1-S5 summarize the parameter estimation results of the proposed method (MRC) under Setups S1-S5, as well as its comparison with the GEE method under Setups S1-S3. Figures S1-S2 present the empirical coverage probabilities of MRC under Setups S1-S5.

Table S6 shows simulation results under a slightly modified Setup S4, where one of the continuous outcomes is dichotomized to binary. Table S7 examines the results when the empirical CDF and missing weights are set to their true counterparts in the estimation procedure. Table S8 compares the ASE and C95 when the perturbation sample size $B = 200$, 400, and 600. Table S9 reports sensitivity study when the missing scheme model was mis-specified using a probit regression model.

Figure S1: Empirical coverage probabilities of the MRC estimator for a single longitudinal outcome under Setups S1-S3. The points in the figures denote averages over 2000 repetitions.

Table S1: Parameter estimation results by MRC for a single longitudinal outcome ($n = 200$).

Setup	missing		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	β_t
S1	No	Bias	0.001	0.003	0.005	-0.004	0.001	0.011	-0.008	0.015	0.005
		ESD	0.077	0.134	0.043	0.071	0.094	0.067	0.088	0.179	0.047
		ASE	0.079	0.137	0.050	0.082	0.113	0.074	0.100	0.202	0.053
		ECP	0.931	0.929	0.939	0.938	0.950	0.947	0.943	0.944	0.949
	Yes	Bias	0.003	0.006	0.007	-0.006	0.006	0.009	-0.008	0.022	0.003
		ESD	0.077	0.133	0.053	0.083	0.121	0.081	0.111	0.222	0.063
		ASE	0.082	0.144	0.062	0.097	0.137	0.088	0.124	0.242	0.067
		ECP	0.934	0.935	0.932	0.939	0.933	0.933	0.937	0.933	0.925
S2	No	Bias	0.001	0.003	0.005	-0.004	0.002	0.010	-0.006	0.014	0.004
		ESD	0.069	0.119	0.038	0.063	0.084	0.060	0.079	0.159	0.042
		ASE	0.071	0.123	0.045	0.073	0.101	0.066	0.089	0.181	0.047
		ECP	0.934	0.935	0.947	0.943	0.951	0.944	0.943	0.953	0.944
	Yes	Bias	0.003	0.005	0.006	-0.005	0.005	0.012	-0.010	0.023	0.004
		ESD	0.069	0.120	0.045	0.075	0.108	0.076	0.103	0.206	0.056
		ASE	0.074	0.130	0.054	0.088	0.125	0.081	0.116	0.226	0.062
		ECP	0.933	0.941	0.939	0.941	0.929	0.939	0.941	0.943	0.943
S3	No	Bias	0.002	0.005	0.009	-0.007	0.003	0.022	-0.016	0.033	0.009
		ESD	0.129	0.223	0.073	0.118	0.157	0.117	0.149	0.302	0.079
		ASE	0.132	0.231	0.084	0.135	0.185	0.127	0.167	0.337	0.089
		ECP	0.938	0.940	0.944	0.940	0.951	0.946	0.946	0.946	0.944
	Yes	Bias	0.005	0.010	0.015	-0.013	0.012	0.039	-0.028	0.073	0.008
		ESD	0.131	0.226	0.090	0.142	0.206	0.158	0.204	0.421	0.109
		ASE	0.141	0.254	0.107	0.167	0.239	0.170	0.229	0.453	0.121
		ECP	0.947	0.952	0.950	0.943	0.955	0.928	0.945	0.942	0.946

Bias = empirical average of the bias for the parameter estimate, ESD = empirical standard deviation of the estimates over replications, ASE = average of estimated standard errors over replications, and ECP = empirical coverage probability of 95% confidence interval.

Table S2: Parameter estimation results by MRC for a single longitudinal outcome with and without missing data ($n = 400$).

Setup	Missing		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	β_t
S1	No	Bias	0.000	0.002	0.003	-0.001	0.000	0.004	-0.001	0.007	0.001
		ESD	0.052	0.091	0.029	0.046	0.067	0.047	0.060	0.127	0.031
		ASE	0.054	0.093	0.032	0.054	0.074	0.049	0.065	0.134	0.034
		ECP	0.946	0.950	0.936	0.950	0.937	0.945	0.947	0.937	0.952
	Yes	Bias	0.000	0.002	0.004	-0.001	0.002	0.003	-0.001	0.010	0.000
		ESD	0.052	0.094	0.035	0.054	0.080	0.055	0.073	0.149	0.039
		ASE	0.056	0.098	0.040	0.063	0.090	0.059	0.082	0.161	0.044
		ECP	0.942	0.935	0.934	0.949	0.943	0.933	0.943	0.941	0.947
S2	No	Bias	0.000	0.001	0.003	-0.001	0.001	0.004	-0.001	0.007	0.001
		ESD	0.046	0.081	0.025	0.042	0.060	0.042	0.053	0.113	0.028
		ASE	0.048	0.083	0.029	0.048	0.066	0.044	0.058	0.120	0.031
		ECP	0.949	0.943	0.944	0.955	0.943	0.942	0.948	0.942	0.949
	Yes	Bias	0.000	0.002	0.003	-0.001	0.000	0.004	-0.002	0.007	0.001
		ESD	0.047	0.084	0.030	0.049	0.072	0.051	0.069	0.140	0.037
		ASE	0.050	0.088	0.035	0.057	0.082	0.053	0.076	0.150	0.040
		ECP	0.942	0.936	0.946	0.948	0.938	0.937	0.940	0.940	0.940
S3	No	Bias	0.000	0.002	0.004	-0.002	0.002	0.009	-0.003	0.016	0.001
		ESD	0.085	0.150	0.046	0.076	0.109	0.078	0.099	0.209	0.051
		ASE	0.089	0.155	0.054	0.088	0.121	0.083	0.108	0.222	0.057
		ECP	0.951	0.948	0.959	0.949	0.945	0.951	0.952	0.953	0.950
	Yes	Bias	0.001	0.005	0.007	-0.004	0.002	0.014	-0.007	0.021	0.003
		ESD	0.087	0.157	0.057	0.091	0.137	0.107	0.134	0.283	0.071
		ASE	0.094	0.167	0.067	0.107	0.153	0.110	0.147	0.294	0.078
		ECP	0.951	0.945	0.952	0.955	0.938	0.923	0.948	0.945	0.940

Bias = empirical average of the bias for the parameter estimate, ESD = empirical standard deviation of the estimates over replications, ASE = average of estimated standard errors over replications, and ECP = empirical coverage probability of 95% confidence interval.

Table S3: Comparison of parameter estimation results by MRC and GEE on complete data for a single longitudinal outcome ($n = 400$).

Setup	Method		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	β_t
S1	MRC	Bias	0.000	0.002	0.003	-0.001	0.000	0.004	-0.001	0.007	0.001
		ESD	0.052	0.091	0.029	0.046	0.067	0.047	0.060	0.127	0.031
		ASE	0.054	0.093	0.032	0.054	0.074	0.049	0.065	0.134	0.034
		ECP	0.946	0.950	0.936	0.950	0.937	0.945	0.947	0.937	0.952
	GEE	Bias	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	-0.001	0.000
		ESD	0.044	0.079	0.017	0.036	0.049	0.017	0.038	0.069	0.024
		ASE	0.044	0.077	0.017	0.036	0.049	0.017	0.039	0.068	0.023
		ECP	0.951	0.942	0.945	0.950	0.944	0.950	0.953	0.947	0.945
S2	MRC	Bias	0.000	0.001	0.003	-0.001	0.001	0.004	-0.001	0.007	0.001
		ESD	0.046	0.081	0.025	0.042	0.060	0.042	0.053	0.113	0.028
		ASE	0.048	0.083	0.029	0.048	0.066	0.044	0.058	0.120	0.031
		ECP	0.949	0.943	0.944	0.955	0.943	0.942	0.948	0.942	0.949
	GEE	Bias	-0.003	-0.008	-0.081	0.057	-0.075	-0.329	0.207	-0.669	0.011
		ESD	0.050	0.092	0.022	0.043	0.064	0.028	0.052	0.094	0.033
		ASE	0.051	0.089	0.022	0.044	0.064	0.028	0.053	0.092	0.032
		ECP	0.957	0.940	0.040	0.731	0.781	0.000	0.030	0.000	0.931
S3	MRC	Bias	0.000	0.002	0.004	-0.002	0.002	0.009	-0.003	0.016	0.001
		ESD	0.085	0.150	0.046	0.076	0.109	0.078	0.099	0.209	0.051
		ASE	0.089	0.155	0.054	0.088	0.121	0.083	0.108	0.222	0.057
		ECP	0.951	0.948	0.959	0.949	0.945	0.951	0.952	0.953	0.950
	GEE	Bias	0.008	0.020	0.180	-0.135	0.171	0.533	-0.331	1.050	0.006
		ESD	0.086	0.144	0.044	0.073	0.116	0.060	0.108	0.205	0.061
		ASE	0.082	0.141	0.043	0.072	0.115	0.061	0.106	0.206	0.058
		ECP	0.939	0.942	0.016	0.555	0.691	0.000	0.101	0.000	0.939

Bias = empirical average of the bias for the parameter estimate, ESD = empirical standard deviation of the estimates over replications, ASE = average of estimated standard errors over replications, and ECP = empirical coverage probability of 95% confidence interval.

Table S4: Parameter estimation results by MRC for the longitudinal GPO ($n = 200$).

Setup	missing		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	β_t
S4	No	Bias	-0.008	0.009	0.009	0.001	0.000	0.012	-0.001	0.000	0.001
		ESD	0.078	0.109	0.082	0.044	0.062	0.096	0.058	0.066	0.023
		ASE	0.085	0.113	0.096	0.054	0.080	0.112	0.073	0.084	0.027
		ECP	0.950	0.948	0.943	0.933	0.952	0.944	0.936	0.956	0.941
	Yes	Bias	-0.012	0.013	0.015	0.003	-0.004	0.019	0.002	-0.009	0.004
		ESD	0.084	0.108	0.093	0.081	0.078	0.118	0.078	0.086	0.032
		ASE	0.095	0.123	0.119	0.090	0.100	0.142	0.096	0.108	0.038
		ECP	0.948	0.950	0.939	0.952	0.948	0.942	0.951	0.959	0.960
S5	No	Bias	-0.008	0.009	0.011	-0.001	0.000	0.013	-0.003	0.000	0.001
		ESD	0.074	0.100	0.069	0.039	0.061	0.085	0.057	0.066	0.021
		ASE	0.084	0.111	0.090	0.051	0.077	0.105	0.071	0.082	0.025
		C95	0.941	0.940	0.945	0.943	0.941	0.940	0.945	0.947	0.940
	Yes	Bias	-0.013	0.014	0.017	0.004	-0.002	0.018	0.001	-0.007	0.003
		ESD	0.082	0.104	0.087	0.077	0.076	0.110	0.078	0.088	0.031
		ASE	0.093	0.119	0.112	0.090	0.097	0.133	0.093	0.106	0.037
		C95	0.944	0.934	0.948	0.962	0.951	0.933	0.937	0.954	0.956

Bias = empirical average of the bias for the parameter estimate, ESD = empirical standard deviation of the estimates over replications, ASE = average of estimated standard errors over replications, and ECP = empirical coverage probability of 95% confidence interval.

Table S5: Parameter estimation results by MRC for the longitudinal GPO ($n = 400$).

Setup	Missing		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	β_t
S4	No	Bias	-0.003	0.006	0.005	-0.001	0.000	0.008	-0.002	0.000	0.000
		ESD	0.049	0.066	0.047	0.027	0.040	0.059	0.036	0.042	0.015
		ASE	0.054	0.072	0.059	0.034	0.049	0.069	0.045	0.052	0.017
		ECP	0.959	0.947	0.943	0.948	0.953	0.937	0.953	0.949	0.950
	Yes	Bias	-0.003	0.006	0.008	-0.001	0.000	0.009	-0.001	-0.003	0.002
		ESD	0.053	0.069	0.058	0.053	0.051	0.074	0.050	0.056	0.020
		ASE	0.060	0.078	0.073	0.059	0.061	0.087	0.059	0.067	0.024
		ECP	0.949	0.942	0.948	0.951	0.940	0.947	0.947	0.951	0.953
S5	No	Bias	-0.003	0.007	0.006	-0.001	-0.001	0.007	-0.002	-0.001	0.001
		ESD	0.051	0.068	0.047	0.024	0.040	0.057	0.036	0.043	0.014
		ASE	0.055	0.073	0.057	0.032	0.049	0.067	0.045	0.052	0.016
		ECP	0.946	0.938	0.938	0.958	0.945	0.934	0.951	0.949	0.949
	Yes	Bias	-0.004	0.007	0.007	-0.002	-0.002	0.008	-0.002	-0.003	0.002
		ESD	0.055	0.069	0.057	0.054	0.051	0.071	0.049	0.058	0.019
		ASE	0.061	0.077	0.070	0.059	0.061	0.084	0.059	0.066	0.023
		ECP	0.939	0.941	0.948	0.951	0.946	0.946	0.953	0.937	0.961

Bias = empirical average of the bias for the parameter estimate, ESD = empirical standard deviation of the estimates over replications, ASE = average of estimated standard errors over replications, and ECP = empirical coverage probability of 95% confidence interval.

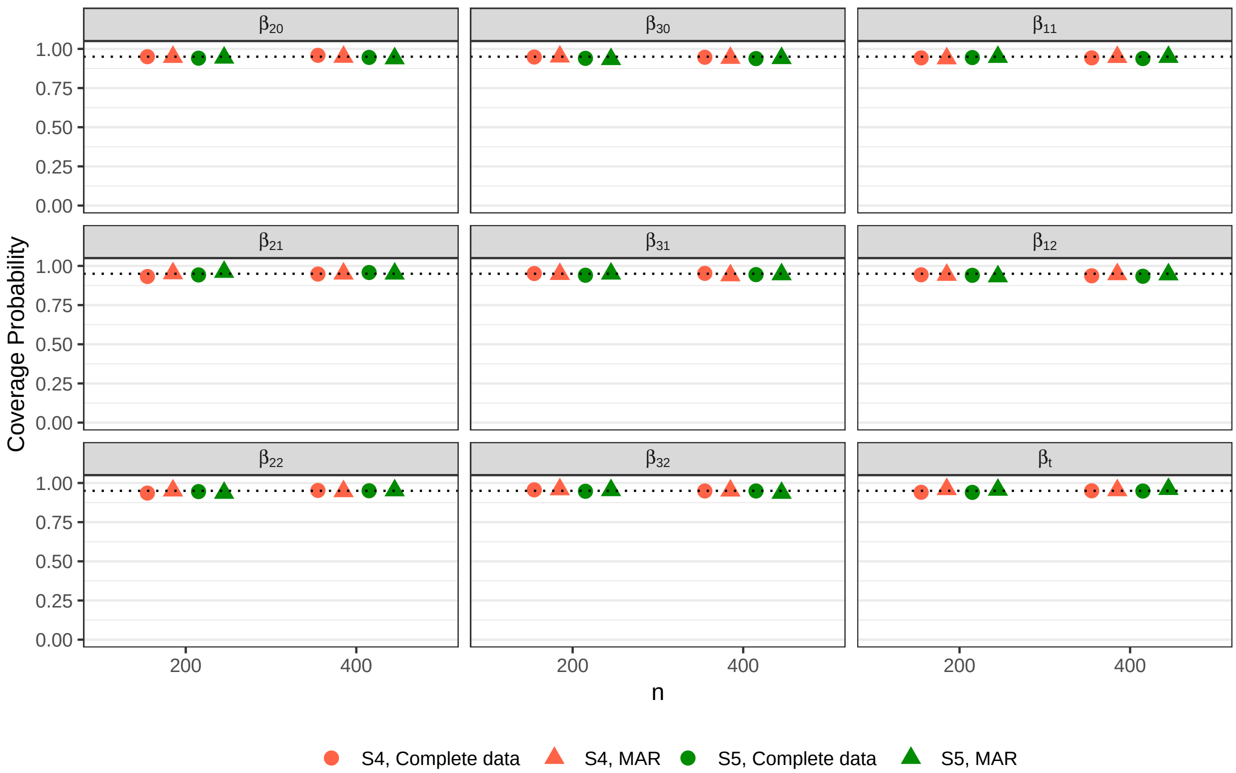

Figure S2: Empirical coverage probabilities of the MRC estimator for the longitudinal GPO under Setups S4-S5. The points in the figures denote averages over 2000 repetitions.

Table S6: Simulation results when one of the outcomes is binary.

n		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	t
200	Bias	-0.011	0.012	0.013	-0.001	0.001	0.015	-0.002	-0.000	0.001
	ESD	0.093	0.126	0.099	0.052	0.080	0.113	0.075	0.085	0.026
	ASE	0.099	0.133	0.119	0.067	0.101	0.139	0.092	0.105	0.030
	C95	0.935	0.934	0.940	0.944	0.939	0.945	0.932	0.944	0.922
400	Bias	-0.003	0.008	0.008	-0.002	-0.002	0.010	-0.003	-0.000	0.001
	ESD	0.060	0.081	0.061	0.034	0.054	0.074	0.048	0.057	0.017
	ASE	0.066	0.088	0.078	0.044	0.066	0.090	0.060	0.068	0.020
	C95	0.934	0.937	0.948	0.943	0.934	0.952	0.947	0.941	0.931

Table S7: Performance of the proposed method when marginal CDF and censoring weights are estimated from data or set to true.

n			β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	t
200	Estimated	Bias	-0.012	0.013	0.015	0.003	-0.004	0.019	0.002	-0.009	0.004
		ESD	0.084	0.108	0.093	0.081	0.078	0.118	0.078	0.086	0.032
	True	Bias	-0.010	0.011	0.016	0.001	-0.002	0.021	-0.003	-0.005	0.005
		ESD	0.082	0.104	0.091	0.082	0.074	0.115	0.078	0.086	0.032
400	Estimated	Bias	-0.003	0.006	0.008	-0.001	0.000	0.009	-0.001	-0.003	0.002
		ESD	0.053	0.069	0.058	0.053	0.051	0.074	0.050	0.056	0.020
	True	Bias	-0.003	0.007	0.005	0.002	-0.003	0.004	0.002	-0.006	0.001
		ESD	0.056	0.071	0.060	0.054	0.049	0.072	0.051	0.057	0.020

Table S8: Comparison of simulation results when $B = 200$, 400, and 600 under Setup S4.

n		B	β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	t
200	Bias		-0.010	0.011	0.013	-0.002	0.002	0.014	-0.002	-0.002	0.001
	ESD		0.077	0.102	0.075	0.043	0.060	0.089	0.059	0.065	0.023
	ASE	200	0.085	0.113	0.096	0.054	0.080	0.111	0.073	0.084	0.027
	ASE	400	0.085	0.113	0.096	0.054	0.080	0.112	0.073	0.084	0.027
	ASE	600	0.085	0.113	0.096	0.054	0.080	0.112	0.073	0.084	0.027
	C95	200	0.951	0.947	0.940	0.936	0.947	0.948	0.934	0.951	0.932
	C95	400	0.950	0.948	0.943	0.933	0.952	0.944	0.936	0.956	0.941
	C95	600	0.951	0.956	0.956	0.948	0.956	0.955	0.943	0.961	0.944
400	Bias		-0.002	0.006	0.005	-0.001	0.000	0.008	-0.002	0.001	0.000
	ESD		0.049	0.066	0.047	0.027	0.041	0.057	0.036	0.044	0.015
	ASE	200	0.054	0.072	0.059	0.034	0.049	0.069	0.045	0.052	0.017
	ASE	400	0.054	0.072	0.059	0.034	0.049	0.069	0.045	0.052	0.017
	ASE	600	0.054	0.072	0.059	0.034	0.049	0.069	0.045	0.052	0.017
	C95	200	0.941	0.938	0.943	0.938	0.938	0.938	0.949	0.939	0.943
	C95	400	0.959	0.947	0.943	0.948	0.953	0.937	0.953	0.949	0.950
	C95	600	0.955	0.947	0.952	0.938	0.948	0.954	0.948	0.946	0.940

Table S9: Performance of the proposed method when the missing weights were estimated using a probit regression model under Setup S4.

n		β_{20}	β_{30}	β_{11}	β_{21}	β_{31}	β_{12}	β_{22}	β_{32}	t
200	Bias	-0.012	0.010	0.015	0.004	-0.003	0.019	0.002	-0.007	0.005
	ESD	0.085	0.106	0.098	0.078	0.075	0.113	0.076	0.088	0.033
	ASE	0.094	0.123	0.119	0.090	0.099	0.142	0.095	0.107	0.038
	C95	0.949	0.963	0.953	0.967	0.959	0.960	0.956	0.954	0.953
400	Bias	-0.001	0.003	0.006	-0.002	0.001	0.007	-0.002	-0.002	0.001
	ESD	0.053	0.071	0.060	0.051	0.051	0.071	0.049	0.055	0.020
	ASE	0.060	0.078	0.073	0.059	0.061	0.088	0.059	0.067	0.024
	C95	0.958	0.942	0.944	0.971	0.929	0.942	0.960	0.948	0.963

References

Cavanagh, C. & Sherman, R. P. (1998), 'Rank estimators for monotonic index models', *Journal of Econometrics* **84**(2), 351–382.

Khan, S. & Tamer, E. (2007), 'Partial rank estimation of duration models with general forms of censoring', *Journal of Econometrics* **136**(1), 251–280.

Pakes, A., Pollard, D. et al. (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica* **57**(5), 1027–1057.

Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1995), 'Analysis of semiparametric regression models for repeated outcomes in the presence of missing data', *Journal of the American Statistical Association* **90**(429), 106–121.

Sherman, R. P. (1993), 'The limiting distribution of the maximum rank correlation estimator', *Econometrica* **61**(1), 123–37.

Sherman, R. P. (1994), 'Maximal inequalities for degenerate u-processes with applications to optimization estimators', *The Annals of Statistics* **22**(1), 439–459.

Van der Vaart, A. W. (2000), *Asymptotic Statistics*, Vol. 3, Cambridge university press.